APA Newsletters

ASIAN AND ASIAN-AMERICAN PHILOSOPHERS AND PHILOSOPHIES
FEMINISM AND PHILOSOPHY
HISPANIC/LATINO ISSUES IN PHILOSOPHY
NATIVE AMERICAN AND INDIGENOUS PHILOSOPHY
PHILOSOPHY AND THE BLACK EXPERIENCE
PHILOSOPHY AND COMPUTERS
TEACHING PHILOSOPHY
# Table of Contents

## Asian and Asian-American Philosophers and Philosophies

- The Timeliness of Translating Chinese Philosophy: An Introduction to the APA Newsletter Special Issue on Translating Chinese Philosophy ........................................... 1
- Preparing a New Sourcebook in Classical Confucian Philosophy ............................................................................ 4
- The Impossibility of Literal Translation of Chinese Philosophical Texts into English ......................................... 9
- Translating Today’s Chinese Masters ................................................. 11
- Three Thoughts on Translating Classical Chinese Philosophical Texts ................................................................. 13
- Introducing Premodern Text Translation: A New Field at the Crossroads of Sinology and Translation Studies ......................................................... 25

## Feminism and Philosophy

- The "Daddy Dividend:" The Gender Division of Labor and Regression Towards Patriarchy ........ 42
- Gender, Disability, and the Violent Undercurrents of Parenting Inspiration Porn .................. 44
- (Philosophizing about) Gender-Open Children .......................................................... 45
- Taking Children’s Autonomy Seriously as a Parent .......................................................... 50
- Parenting in Trauma ......................................................................................................... 53
- On Muddling Through ........................................................................................................ 60
- Parenting, Feminism, and Academic Life: My Happy Story ........................................ 61
- Bridging the Divide: Thoughts on Parenting as a Grad Student ........................................ 64
- Anthropologists from Mars ............................................................................................... 66
- Raised in Philosophy ........................................................................................................... 66
- Children, Parenting, and the Nature of Work ........................................................................ 68
- Learning from My Daughter: The Value and Care of Disabled Minds .......................... 69
- Overcoming Epistemic Injustice: Social and Psychological Perspectives ......................... 71

## Hispanic/Latino Issues in

- The Virtues of Mestizaje: Lessons from Las Casas on Aztec Human Sacrifice ......................... 82
- Chicano/a Philosophy: Rupturing Gringo Anti-Chicano/a Paradigms and Philosophies .......... 88
- Coloniality of the U-S///Mexico Border: Power, Violence, and the Decolonial Imperative .... 97
- Lessons in Exile ................................................................................................................ 98

## Native American and Indigenous Philosophy

- Empowering Relations: An Indigenous Understanding of Allyship ........................................ 103
- Indigenizing Philosophy Through the Land: A Trickster Methodology for Decolonizing Environmental Ethics and Indigenous Futures ..... 109
- The Phenomenology of a Performative Knowledge System Dancing with Native American Epistemology ................................................................. 116

## Philosophy and the Black Experience

- Joyce Mitchell Cook (1933–2014) .......................................................... 120
- Textual Mysticism: Reading the Sublime in Philosophical Mysticism ........................................... 121
- Spectacle Lynching, Sovereignty, and Genocide: A Dialog with Al Frankowski .................. 124
- Purdue University and President Mitch Daniels: Confession of a Rare Creature .................. 127
Table of Contents, continued

Philosophy and Computers ........................................... 129
   The Shrinking Difference Between Artifacts and Natural Objects* ........................................... 129
   Artifacts and Mind-Independence: Comments on Lynne Rudder Baker’s “The Shrinking Difference between Artifacts and Natural Objects” ........................................... 132
   Explaining an Explanatory Gap .................................... 134
   Formulating the Explanatory Gap ..................................... 136
   Logic as a Theory of Computability .............................. 137
   Robots Need Conscious Perception: A Reply to Aleksander and Haikonen ........................................ 142
   Flawed Workspaces? ...................................................... 144
   Unity from Multiplicity: A Reply to Haikonen .............. 146
   Leibniz, Complexity, and Incompleteness ...................... 147
   Architecture-Based Motivation vs. Reward-Based Motivation ......................................................... 150
   Consciousness, Engineering, and Anthropomorphism ................................................................. 154
   Sleep, Boredom, and Distraction: What Are the Computational Benefits for Cognition? ................. 160
   DABUS in a Nutshell ..................................................... 163
   The Real Moral of the Chinese Room: Understanding Requires Understanding Phenomenology ....................... 166
   A Refutation of Searle on Bostrom (re: Malicious Machines) and Floridi (re: Information) ................. 172
   Understanding Information Ethics ................................ 175
   Too Much Information: Questioning Information Ethics ............................................................. 185
   Ethics of Entropy ......................................................... 189
   Taking the Intentional Stance Toward Robot Ethics ................................................................. 194
   Measuring a Distance: Humans, Cyborgs, Robots ................................................................. 197
   Remediation Revisited: Replies to Gaut, Matravers, and Tavinor ............................................... 202
   My Philosophy in AI: A Very Short Set of Notes Towards my Barwise Prize Acceptance Talk ........... 207
   The Association for Philosophy and Computing .......... 208

Teaching Philosophy .................................................... 209
   Devoting a Course to the Exploration of a Book: Journeying Intellectually with the Students ........... 210
   Aristotle, De Anima ..................................................... 214
   To Teachers Who Hope to Inspire Their Students ................................................................. 216
   To Those Who Think the Unexamined Life Is Not Worth Living .................................................. 216
   To Cynthia Ozick* ..................................................... 216
   In Praise of Campus Culture Wars .............................. 216
   Philosophy Rap ......................................................... 217
If there is one theme running through all of the articles in this issue, it is that of timeliness. There is an urgency in reevaluating how we, the Western audience, approach Chinese philosophy: how we study it, how we explain it to others, and how we translate it. There is more than one reason this reevaluation is timely.

There are cycles in academia. Old projects are revived with new data and new perspectives. In the sixteenth century, toward the end of China’s Ming Dynasty, Catholic missionaries from Europe began to arrive in China, wave after continuous wave. Their goal was to convert the locals, and in their efforts they studied and translated Chinese philosophy in earnest. As this tide faded away, a new wave of Protestant missionaries took its place, with their own translation work. Enter the twentieth century and new generations of sinologists, though with less firsthand experience in China than the missionaries, attempted to tackle Chinese philosophy in a more academic, less religious way. We do not need to come to a consensus on whether they succeeded; we only need to recognize that through these cycles, progress was being made on how ancient Chinese texts were being rendered into Western languages. We are now two full decades into the twenty-first century, and the tide has come back again.

China is rising to power in an unprecedented way. It is neither cliché nor irrelevant to mention in this philosophy newsletter China’s growing economic, political, and military power. These are not the reasons why people study the beautiful world of Chinese philosophy. They are the reasons why so many people are studying it. China is poised to become a, if not the, superpower of this century. It is completely expected that so many institutions and individuals around the world are trying to understand it. Is not Chinese philosophy a good place to start?

Since most of us reading this newsletter have at least a vague idea of what Western philosophy is, we must understand that to then learn Chinese philosophy is truly to reinvent the wheel. It is necessary to start from the most basic notions of what philosophy is to be able to understand what Chinese philosophy is.

In the West, religion is religion and philosophy is philosophy. In China, this line does not exist. For China and its close East Asian neighbors, Confucianism has guided the social and spiritual lives of people for thousands of years in the same way the Judeo-Christian tradition has guided people in the West. It is a feeling we have in our bones that there exists a persistent tradition that connects me and my people to our past and our future. In this sense Confucianism is analogous to religion in the West. And yet in Confucianism there is no god to whom we can address all our ultimate questions. Confucianism has historically been recognized by the government as the ideology that binds people to each other and to the state itself, and yet it has never been institutionalized the way Judeo-Christianity has: there is no church, no pope. Confucianism in China has never been recognized as a state religion, and yet for more than one thousand years the civil service entrance exams were based on how familiar the candidate was with Confucian texts. To learn Chinese philosophy, to start from the most basic notions, means to question what philosophy and religion are.

Such questions come at a precarious time in Western society. Nationalist movements and xenophobic sentiments are rising. Social conflict and international wars, healthcare and immigration, these issues are debated daily but the situations are hardly improving. This makes many call into question the sanctity of long-standing institutions that were once unquestionable: government based on democracy, economies based on capitalism, philosophies based on Greek ideas of logic, a legal system based on Roman ideas of justice. Our grand institutions and traditions are revealing their inherent flaws. It is enough to make some sinologists offer up Chinese-inspired alternatives.

Confucius’s philosophy can be divided into two levels: political theory and personal cultivation. Of the more than four hundred passages in the Analects of Confucius, every passage treats on at least one of the two. Modern sinologist and political theorist Daniel Bell’s sinological research focuses on Confucian political theory. His corpus of work invokes traditional Chinese ideas on governance as possible alternatives to the two-party, electoral democracy systems visibly struggling in the West. The sinological work of Henry Rosemont Jr. focuses on Confucian personal
This time, this age we are in, is no longer a time when Chinese philosophies are regarded as quaint and exotic, to be found on bookshelves next to transcendental meditation and crystal gazing. China is the largest country in the world by population. It is influencing the entire world, and its ancient sages have something to say about what real philosophy is. There is, however, one large problem: they say it in Chinese.

Thus we come to the theme of this issue, Translating Chinese Philosophy. The language of ancient Chinese philosophy is so different from modern English, I would dare say there is no exercise that could better capture the definition of translation as being “re-creation.” To express what faraway Chinese philosophers preached to a twenty-first-century Western audience takes more than immense skill. It takes great imagination and creativity. This is why, no doubt, Edward Shaughnessy declares unequivocally in his present contribution, “translation is an art, not a science.” In this issue each contributor approached the topic in his or her uniquely valuable way.

In the first article, Roger Ames’s starting point is Wing-tsit Chan’s *Source Book in Chinese Philosophy.* As Ames points out, Chan set the standard in 1963 for many of the terms we use for translating Chinese philosophy. However, conditions change. Bars are raised. Standards need recalibrating. Ames has spent much of his influential career advocating that students learn to take Chinese philosophy “on its own terms.” This means situating Chinese philosophies and philosophers in their native context, and making a conscious effort not to express Chinese philosophy using Western philosophical terms. Western philosophy is not the only philosophy. Western religion is not the only religion. Translating Chinese philosophical works using the lexicon of logic, rights, individualism, abstract principles, and other philosophical terms we are comfortable with does a great disservice to a tradition that does not contain such ideas. It infuses Chinese philosophy with our Western biases, and in turn takes us further away from, rather than closer to, Chinese wisdom. Ames is compiling a new *Source Book in Confucian Philosophy,* and his contribution to this issue is a discussion of his many considerations for doing so, and why, almost sixty years after Chan, it is time to do it again.

In the second article, Tian Chenshan, a former student of Ames and currently director at the Center for East West Relations at Beijing Foreign Language University, expands on Ames’s proposition by claiming that the most central terms in Chinese philosophy have no direct correspondence with any words in English. Importantly, the difficulty of translating goes beyond incompatible lexicons. He observes, “The languages of China and the West are imbued with their own unique world views, epistemologies, ideologies, and value systems.” To force a pairing of Chinese-to-English translations would be to wrongly insinuate the West’s legacy of logic (philosophy) and monotheism (religion) into Chinese philosophy. This is why Tian, Ames, and so many of their collaborators place more importance on the well-thought-out explanation of key Chinese philosophical terms than on their actual translation.

The third article comes from three young Western scholars working in China, at the forefront of a massive soft power push by the Chinese government. As America rose to global dominance after World War II, it made aggressive efforts to exert its political, economic, and military influence around the world. Ironically, it was America’s culture that spread more rapidly and was accepted more readily. This was America’s “soft power.” Now that the Dragon is rising, it wishes to copy this success story. Though this time around, the exportation of culture will not be an involuntary side effect. It is deliberate. “Part of the Chinese Communist Party’s Soft Power Initiative is to expand the country’s cultural appeal, and academically speaking, philosophy is at the top of the list. The proclaimed aim of translating and publishing one hundred books a year and the booming number of translation projects funded directly by the central government at one yuan per character, are indeed telling.” In this sense, the study and translation of something as classical as Chinese philosophy could not be more closely linked to the geopolitics of the twenty-first century. Again, timeliness. Dimitra Amarantidou, Daniel Sarafinas, and Paul J. D’Ambrosio are not translating Chinese philosophy. They are translating the works of modern Chinese scholars who write about Chinese philosophy. This endeavor presents them with unique translation obstacles. Not only are ancient Western and Chinese philosophy fundamentally different, the way that modern Western scholars and modern Chinese scholars talk and think about philosophy is different. Even ideas on something as seemingly insignificant as academic protocol vary greatly across the oceans. Plagiarism is the cardinal sin in the halls of Western academia, but repeating what one has previously read is a sign of learning in the East. As the bridges between these two foreign lands, how can these translators most appropriately translate a Chinese professor’s words for a Western reader? They are the first to admit they do not have all the answers. But like any sincere scholar, they are asking the right questions.

Edward Shaughnessy has been well known for decades in sinology circles for his work on early Chinese historiography, particularly because of his use of all the methods and tools available to a classical scholar. He utilizes the ancient commentary of received texts (the ancient texts that have been handed down generation to generation), the documents that have been unearthed at modern archeological sites, and a slew of textual criticism subfields to reconcile the two. His present contribution to this newsletter chooses three highly specialized fields of textual scholarship—etymology, paleography, and sonics—to demonstrate the subtle considerations which enter into translating ancient Chinese. But despite the technicality of this article, it also speaks to larger themes and trends in contemporary translation. The material that is available to us today is not the same material available to us yesterday. Or tomorrow. Shaughnessy points out that advances are constantly being made in the theoretical fields, and new physical discoveries are being made which broaden...
the sea of material we have to work with. His present contribution, and his larger body of work as a whole, teach us an important lesson: a translation, no matter how good, always has an expiration date. The passage of time and the changing of conditions will always require improved translations for new generations. If nothing else, even an excellent translation from one hundred years ago will use an English vernacular that reads uncomfortably to us. This does not even touch upon advances in the field. Shaughnessy’s work embodies this principle. There are at least eleven translations of the I Ching, Book of Changes, going back as far as 1867. The majority of them are translations based on the received text. In 1998 Shaughnessy published I Ching, The Classic of Changes: The First English Translation of the Newly Discovered Second-Century B.C. Mawangdui Texts (Ballantine Press). Then in 2014 he published Unearthing the Changes: Recently Discovered Manuscripts of and Relating to the Yi Jing (Columbia University Press). Conditions change. Resources change. Needs change. All translations have a time stamp on them, and they all have an expiry date.

This issue culminates with an important article that goes beyond reviewing where we are. It previews things to come. Carl Gene Fordham, a promising doctoral student in Peking University’s Chinese Department, opens up the possibility of a new branch of scholarship: “premodern text translation.” It is a disciple defined by its interdisciplinary nature. “[T]his field has the potential to explore many new and exciting questions that cannot be adequately addressed by sinology or translation studies as singular disciplines.” The best translator of ancient Chinese philosophy texts would ideally be well versed in translation technique and theory, modern Chinese, ancient Chinese, textual criticism, and philosophy. It is rare that the translators of the past can claim to be expert in all these fields. This may seem like too much to expect from a translator, but it is not unrealistic. In fact, it is a trend that is already emerging. “Over the past two decades during which this interdisciplinary field has developed, universities throughout China have been training premodern text translators in the hope that China may, one day, produce its own high-quality translations of the literature it rightly prizes as part of its intangible cultural heritage.” The time is ripe for this new field to establish itself. China is making a push to export its literary culture in the form of high quality translations accessible to Western readers. The Chinese government and universities provide educational programs to educate new generations of Chinese and foreign scholars, they give generous grants for such work, and they provide platforms for publishing. On the Western side, conditions are also becoming more advantageous for this enterprise. In the twentieth-century, China scholars did not always have easy logistical access to China and its academic resources. Today’s generation of sinologists can and often do come to China to live and learn for as long as they desire. The sharing of academic work through the internet and the digitization of ancient texts deliver a world of resources to anyone’s desk. Furthermore, since a wave of China studies has captured the attention of mainstream Western academia over the last two decades, with independent Chinese studies or East Asian studies departments popping up in our best institutions of higher learning, young Western scholars have the opportunity to completely dedicate themselves to these fields. Long gone are the days when sinological debates in the West were monopolized by scholars who began their formal training in Western subjects, and then at some point later in their career decided to “move on” to sinology. This new discipline, premodern text translation, then, is a very natural outgrowth of its time.

I will make the assumption that all of the contributors herein agree with me when I say, I hope this issue on Chinese philosophy is not an instance of preaching to the choir. If the only readers of this newsletter on questions surrounding Chinese philosophy and its corresponding translation work are students and professionals who already specialize in Chinese philosophy, then I am afraid this issue will not achieve its potential. Our hope is that those who do not specialize in Chinese philosophy will begin to see the position that Chinese philosophy is due in the vast array of world philosophies. The new theories and techniques that arise from Chinese philosophical translation can be critically absorbed into the translation work of any other language. True philosophers, lovers of wisdom, looking for insights to improve their lives, can reach out to Chinese wisdom with great benefit. Teachers and students of Western philosophy can enhance their grasp of world philosophy more fully and reflect on their own specialty more clearly by understanding what philosophy means in China.

NOTES

1. It is hard for the Western mind to reconcile a religion with no institutions. Robert Bellah was on to something when he called Confucianism a “civil religion.” See Philip J. Ivanhoe and Sungmoon Kim, eds., Confucianism, a Habit of the Heart (Albany: State University of New York Press, 2016).


5. A typical example of this would be the 45-page introductory section titled “Philosophical and Linguistic Background” that precedes the main text of Ames and Rosemont’s translation of the Analects. See Roger T. Ames and Henry Rosemont Jr., The Analects of Confucius: A Philosophical Translation (New York: Ballantine Books, 1998).

ARTICLES

Preparing a New Sourcebook in Classical Confucian Philosophy

Roger T. Ames
Peking University
RTAMES@HAWAII.EDU

AN INTERPRETIVE STRATEGY

Over the past few years I have been keeping out of mischief by working on a new bilingual Sourcebook in Classical Confucian Philosophy designed to serve the needs of both Chinese and Western students of Chinese philosophy. This project has given me the occasion to reflect on some of the translation challenges we must face in doing our best to move between very different philosophical traditions. In Wing-tsit Chan’s pioneering contribution to this important initiative, A Source Book in Chinese Philosophy (Princeton, 1963), his choice was to provide his readers with a considerable volume of translated textual materials organized chronologically, with a minimum of philosophical commentary and little by way of providing an interpretive context. The Chan Source Book is foundational in the sense that, in what it includes and what it excludes, it has circumscribed the parameters of the corpus for a generation of students of Chinese philosophy. For example, pre-Qin philosophy is much emphasized while Han dynasty texts and figures have minimal coverage. Again, in setting a high standard in the quality of his translations, Chan has also galvanized a specific formula of translations of key philosophical terms, promoting what scholars have since come to regard as a standard if not “literal” rendering of the classical Chinese philosophical vocabulary. For its time, Chan’s Source Book was a quantum advance both in coverage and in quality on what had serendipitously been translated from the Chinese philosophical canons. But then again, each generation is called upon to build its own connector to the generation that follows, and we need a new sourcebook for ours.

In my efforts to compile a new Sourcebook, looking back on where we are, I have two closely related worries. I will give an account below of why I believe that we have not paid sufficient attention to locating these philosophical classics within their own interpretive contexts. Indeed, by default we have inadvertently transplanted these texts into a worldview and a commonsense not their own. Fully aware that worldviews are in the fullness of time invested in our natural languages, the consequence of this omission is the danger of uncritically perpetuating the same formula for rendering key philosophical terms, where this now standard vocabulary has encouraged a sense of literalness and familiarity with an erstwhile “Chinese” philosophical vocabulary.

Friedrich Nietzsche in Beyond Good and Evil reflects upon how a specific worldview is sedimented into the very language that speaks it:

The strange family resemblance of all Indian, Greek, and German philosophizing is explained easily enough. Where there is an affinity of languages, it cannot fail, owing to the common philosophy of grammar—I mean, owing to the unconscious domination and guidance by similar grammatical functions—that everything is prepared at the outset for a similar development and sequence of philosophical systems; just as the way seems barred against certain other possibilities of world-interpretation.¹

Nietzsche is certainly not endorsing any theory of strong linguistic determinism—that is, the idea that our languages constrain us necessarily to think in certain ways. Rather, he is simply observing that a language and its syntax—in his example here, the Indo-European family of languages—over time become invested with a particular cultural narrative’s insights into what makes the human experience meaningful. Natural languages and their structures tend to reveal the default worldview of the cultures that speak them. Said another way, our languages “speak” us as much as we speak our languages, disposing us to entertain experience in one way as opposed to another, and prompting us to ask some questions rather than others.

Indeed, this same Nietzsche, reflecting on how languages such as French and German came to be gendered—la table and le soleil—allows that “when man gave all things a sex he thought, not that he was playing, but that he had gained a profound insight. . . .”² In fact, the corpus of Nietzsche himself is an object lesson in the very problem of translation and interpretation he ponders. Our languages are conservative in wanting to speak from within their own narratives, and tend to resist new ideas in proportion to the disjunction these ideas have with what has gone before. Commonsense is obstinate. Thus, when Nietzsche attempts to critique a persistent transcendentalism that has become entrenched within the languages and cultural experience of the Abrahamic traditions captured in his famous proclamation “God is dead,” he must himself become linguistically dexterous. He has little choice but to turn to and rely heavily upon rhetorical devices and literary tropes rather than the more “literal” expository language available to him precisely because he is frustrated, compromised, and even betrayed by the deeply committed language in which he is attempting to give voice to his revolutionary ideas.

A point that was drilled into me by my teachers was that different cultures think differently, and that philosophically, we elide important distinctions among them at our peril. The distinguished British sinologist, Angus Graham, for example, ascribes unique and evolving categories and conceptual structures to different cultural traditions, and in so doing, challenges the Saussurian structuralist distinction between langue (universal and systematic linguistic structures and rules governing all languages) and parole (diverse and open-ended speech acts in any of our natural languages).³ All the same, we might borrow and take liberties with Saussure’s distinction that abandons his structuralist assumptions to make Graham’s point. We can use langue (language) to contrast the evolved, theoretical, and conceptual structure of a given language
system shaped by an aggregating cultural intelligence over millennia that makes organized speech possible, with parole (speech) as the application of this natural language in the individual utterances we make." Graham and we fellow pluralists need such a distinction to reinforce our claim that the Chinese language has not developed and does not have available to it either an indigenous concept or a term that can capture the Abrahamic notion of "God," while at the same time allowing us to insist that the same Chinese language has all of the semantic and syntactic resources it needs to give a fair and robust account of such an idea. And likewise, there is no vocabulary available in our Western languages to do justice to the conceptual structure of Confucianism. We cannot say *li* in English, or in German either, although we can say lots about this key Confucian notion in both European languages, and get pretty clear on what it means.

Recently, and specifically in reference to the classical Chinese language, Graham concludes that in reporting on the eventful flow of Chinese *qi* cosmology, "the sentence structure of Classical Chinese places us in a world of process about which we ask ... "Where?" and also, since it is moving, 'At what time?" Like many (but not all) of us, Graham is persuaded that different populations within always changing cultural milieus appeal to different concepts and ways of thinking and living. And Graham spent his illustrious career doing his very best to bring some clarity to these differences. What Graham is saying here is that any perceived coherence in the emergent order of things that is assumed in Confucian cosmology, while being abstract and theoretical, is resolutely historicist and situated, and hence has to be qualified by a location, by a particular time in its evolution, and also by its applications.

When Graham asks after human nature within the context of Confucian cosmology, for example, he avers that firstwhile "human nature" is conceived of as an ongoing and evolving process rather than as some essential "timeless" property or universal endowment. Thus, beyond the question of "What is it?" he must also ask the other questions: "Where was it thought of in this way?" "Whence did we come to mean this?" "How did it serve us to think of it in this way?" and, perhaps most importantly, "Whither is its impetus in defining who we will become?" Indeed, we might invoke a distinction found in the first among the Confucian classics, the *Yijing* 易經 or *Book of Changes*. While cosmic order and all that emerges within it have certainly been understood in general and persistent terms (*tong* 通), at the same time, they must always be qualified by the local, the specific, and the transitory (*bian* 變). For Confucian cosmology, in the ongoing transformation of the world around us, we must respect the *where*, the *when*, and the who as aspects integral to this ineluctable process. The crucial implication of Graham’s insight into Confucian cosmology is that all of the rational structures that might be appealed to in expressing an understanding of the human experience—that is, whatever theories, concepts, categories, and definitions we might reference—are themselves all ultimately made vulnerable to change by the always shifting organs and objects of their application. In the flux and flow of experience, making sense of a changing world is itself a changing process.

In fairness to the new translations that have continued to appear over the past generation, we must ask the question: At the end of the day, can European languages, freighted as they are with a historical commitment to substance ontology—what Jacques Derrida has called "logocentrism" and "the language of presence"—actually "speak" the processual worldview that grounds these Chinese texts? Can these canonical texts such as the *Book of Changes* and the *Expansive Learning* (Daxue) be translated into English and still communicate the worldview that is invested in them? And more to the point, given my project as it is presently at hand, how does my Sourcebook propose to address the challenge of trying to locate these Chinese texts within their own implicit worldviews?

**RESOURCES FOR DEVELOPING A BASIC CHINESE VOCABULARY**

An astute Ludwig Wittgenstein insists that “the limits of our language mean the limits of our world.” If this is the case, in order to take Chinese philosophy on its own terms, we will quite literally need more language. The premise then is that there is no real alternative for these students but to cultivate a nuanced familiarity with the key Chinese vocabulary itself. The self-conscious strategy of this Sourcebook is to enable students of Confucian philosophy to read the seminal texts by going beyond simple word-for-word translation. This alternative is to provide these students with a pathway to develop their own sophisticated understanding of a cluster the most critical Chinese philosophical terms. Indeed, word-for-word translation can in the long run be counterproductive to the extent that it encourages students in reading these texts to inadvertently rely upon the usual implications of the translated term (for example, “Heaven”) rather than on the range of meaning implicit in the original Chinese term itself (for example, *tian* 天). When students read *tian* 天 as “Heaven” rather than as 天, they are certain to read the text differently, and in all likelihood, in a way heavily freighted with Western theological assumptions.

By way of analogy, if we reflect on our best efforts to read and teach classical Greek philosophy, most of us do not have an expert knowledge of the original language texts. But in developing a sophisticated understanding of an extended cluster of the most important Greek philosophical terms—*logos*, *nomos*, *nous*, *phusis*, *kosmos*, *eidos*, *psyche*, *soma*, *arche*, *alethea*, and so on—we are with imagination, able to get behind our own uncritical Cartesian assumptions and at least in degree, read these Greek texts on their own terms. In a similar way, by seeking to understand and ultimately appropriate the key philosophical vocabulary around which the Chinese texts are structured, students will be better able to locate these canonical texts within their own original intellectual assumptions. The only alternative to attempting to take the tradition on its own terms is to participate in a further colonializing of Chinese philosophy and the truncating of its long history. We have to resist the unconscious and patently spurious assumption that this tradition’s fairly recent encounter with the vocabulary of the Western academy has been its defining moment. Such an uncritical approach places the uniqueness, the heterogeneity, and the intrinsic worth of the Chinese philosophical tradition at real risk.
It is in this effort to take Chinese philosophy on its own terms then, that the first section of this *Sourcebook* is an attempt to explicate the tradition’s own indigenous presuppositions and its own evolving self-understanding. A careful reading of the introductory essay to our *Sourcebook*, “Chinese Natural Cosmology: An Interpretive Context,” will hopefully sensitize the reader to some of the ambient and persistent assumptions that have given the Confucian philosophical narrative its unique identity over time. It is these same presuppositions that inform the philosophical vocabulary and set the parameters from which their meanings in context must be parsed.

Are we then to understand that these generic, persistent cultural assumptions are essential and unchanging conditions of Chinese cosmology? Of course not. We have to unload this “essentialism” charge and respect responsible generalizations. As I argue in setting out this interpretive context for classical Confucian philosophy, making cultural comparisons without the hermeneutical sensitivity necessary to guard against cultural reductionism is undertaken at the risk of overwriting these same texts with our own cultural importances. In this insufficiently critical process, we inadvertently make a world familiar to us that is not familiar at all, and in this specious familiarity, effectively surrender much of the substance of the tradition’s own uniqueness and value.

Of course, in the *Sourcebook*, a critical version of the original classical Chinese texts is included for both the expert and novice alike to make whatever comparisons and evaluations they choose. To inform this comparative exercise, I and my collaborators D. C. Lau, David Hall, and Henry Rosemont in our earlier translations of the canonical texts have over the years compiled a rather substantial glossary describing the evolution of this extended cluster of key philosophical terms. In the new *Sourcebook* I have expanded upon this lexicon for students as they proceed in their readings to return on a regular basis. Just as providing the interpretive context is a self-conscious attempt to be as cognizant as we can about the assumptions we bring to the text, I think it is equally important to say up front why we have translated particular terms in the way that we do, and why we have abandoned many of the earlier formulations. In addition, in order to prompt and encourage students to reference this explanatory glossary, I include the Chinese characters and the romanization for these key terms along with their “placeholder” translations within the translated texts themselves, making the point that a new set of inadequate English translations to replace the earlier one is not the solution. Again, sometimes the same Chinese term in a different context is better served by a different English translation. But let me be clear about the expectations we have for the reader of the *Sourcebook*. At the end of the day, the project in this *Sourcebook* is not to replace one set of problematic translations with yet another contestable set of renderings. The goal is to encourage students to move between the translated texts, the philosophical introduction that provides the interpretive context, and the glossary of key terms, with the hope that in the fullness of time they will appropriate the Chinese terminologies themselves—*tian* 天, *dao* 道, *ren* 仁, *yi* 裕, and so on. In thus developing their own increasingly robust insight into these philosophical terms, the students will be able to carry this nuanced understanding over in their critical reading of other available translations. Ultimately for students who would understand Chinese philosophy, *tian* 天 must be understood as *tian* 天, and *dao* 道 must be *dao* 道.

THE INTERPRETIVE NATURE OF TRANSLATION

In our earlier forays into translating the Chinese canons, Lau, Hall, Rosemont, and I have developed a structure that, like this *Sourcebook*, includes a philosophical introduction, an evolving glossary of key philosophical terms, and self-consciously interpretive translations. I have given the rationale for including the extensive introduction and the glossary above. With respect to the translations themselves, in describing them as “self-consciously interpretive,” I am not allowing in any way that we are recklessly speculative or given to license in our renderings, nor that we are willing to accept the reproach that we are any less “literal” or more “creative” than other translators. On the contrary, I would insist first that any pretense to a literal translation is not only naïve, but is itself an “objectivist” cultural prejudice of the first order. Just as each generation selects and carries over earlier thinkers to reshape them in their own image, each generation reconfigures the classical canons of world philosophy to its own needs. We too are inescapably people of a time and place. This self-consciousness then, is not to disrespect the integrity of the Chinese philosophical narrative, but to endorse one of the fundamental premises of this commentarial tradition—that is, textual meaning is irrepressibly emergent, and that, like it or not, we translators are integral to the growth of the tradition, and as such, are not passive in the process of interpretation.

At a general level, I would suggest that English as the target language carries with it such an overlay of cultural assumptions that, in the absence of the “self-consciousness” that is herein represented by the introduction and glossary of terms, the philosophical import of the Chinese text can be seriously compromised. Further, a failure of translators to be self-conscious and to take fair account of their own Gadamerian “prejudices” with the excuse that they are relying on the existing “objective” lexicon—a resource that, were the truth be known, is itself heavily colored with cultural biases—is to betray their readers not once, but twice. That is, not only have they failed to provide the “objective” reading of the text they have promised, but they have also neglected to warn their unsuspecting reader of the cultural assumptions they have willy-nilly insinuated into their translations.

CHINESE PHILOSOPHY AS “EASTERN RELIGIONS”

It has become a commonplace to acknowledge that, in the process of Western humanists attempting to make sense of the classical Chinese philosophical literature, many unannounced Western assumptions and generic characteristics have been inadvertently insinuated into the understanding of these texts, and have colored the vocabulary through which this understanding has been articulated. To the extent that Chinese philosophy has become the subject of Western philosophical interest, it has often if not usually been analyzed within the framework of categories and philosophical problems not its own.
Indeed, over the last several centuries of cultural encounter, the vocabulary established for the translation of classical Chinese texts into Western languages has been freighted by an often-unconscious Christian framework, and the effects of this “Christianization” of Chinese texts are still very much with us. There are numerous examples of grossly inappropriate language having become the standard equivalents in the Chinese/English dictionaries that we use to perpetuate our understanding of Chinese culture: “the Way” (dao 道), “Heaven” (tian 天), “benevolence” (ren 仁), “righteousness” (yi 義), “rites” (li 禮), “virtue” (de 德), “substance” (ti 體), “principle” (li 理), “material substance” (qi 氣), and so on. Can a Western student read the capitalized “Heaven” as anything other than a metonym for the familiar notion of a transcendent God? Is living a life as someone’s son or daughter a “rite”? Should we reduce what is quite literally the image of cultivated, consummative human beings in all their aspects—cognitive, moral, aesthetic, religious, somatic—to a single, patently Christian virtue: “benevolence”? When and in what context would a native English speaker ever utter the word “righteousness?”

Chinese philosophy understood through this existing formula of translations has been made familiar to Western readers by first “Christianizing” it, and then more recently, by “orientalizing” it and ascribing to it a depreciating poetical-mystical-occult and religious worldview, as an alter image to our logical-rational-enlightened and decidedly secular self-understanding. Many of the more philosophically-inclined sinologists who have been involved in the recent translation of canonical Chinese works are now acknowledging that a fuller inventory of semantic matrices might be necessary for the translation of these philosophical texts, and are struggling to get beyond the default, “commonsensical” vocabularies of their native cultural sensibility.

AN INTERPRETIVE ASYMMETRY: VERNACULAR ASIAN LANGUAGES AND THE LANGUAGE OF MODERNITY

Beyond the “Christianizing” and “orientalizing” of the Confucian tradition, there is a further asymmetry that continues to plague our best attempts to make responsible comparisons between the Chinese and Western philosophical narratives. To state the problem simply, we have been given to relentlessly theorizing the Chinese tradition according to our Western philosophical assumptions, shoehorning Chinese concepts into categories that are not its own. We are given to pondering with some philosophical nuance: “Is Mohist utilitarianism agent-neutral or agent relative?” but it would not occur to us to ask if John Stuart Mill is an early or latter day Mohist. Again, we are given to a penetrating debate on: “Is Confucian ethics an Aristotelian aretic ethic or a Humean-inspired sentimentalist ethic?” but it would not occur to us to ask if Aristotle, and Hume too, are classical or perhaps later Confucians. Kwong-loi Shun has recently made much of this asymmetry:

[T]here is a trend in comparative studies to approach Chinese thought from a Western philosophical perspective, by reference to frameworks, concepts, or issues found in Western philosophical discussions. This trend is seen not only in works published in the English language, but also in those published in Chinese. Conversely, in the contemporary literature, we rarely find attempts to approach Western philosophical thought by reference to frameworks, concepts, or issues found in Chinese philosophical discussions.

As Shun observes, this problem is as true in the writings of contemporary Chinese intellectuals as it is of their Western counterparts, speaking as they do a vernacular language transformed by its encounter with a dominating Western modernity, and thus deploying a largely Western conceptual structure—a Western langue—even while speaking their own vernacular language. In the middle and late nineteenth century, the institutional apparatus of Western education was transplanted wholesale to reconfigure East Asian education to its very core. The institutions of European and American education—the public-school systems through to the universities with their disciplinary taxonomies and curricula—were imported whole cloth into the East Asian cultures of Japan, China, Korea, and Vietnam. First, the Meiji Japanese reformers and then the Chinese, Korean, and Vietnamese intellectuals, at once enamored of and overwhelmed by Western modernity, created their own Sinitic equivalencies drawn largely from traditional Chinese literary resources, to appropriate and give voice to the conceptual and theoretical language of the imported Western academic culture. The vocabulary of modernity with its liberating enlightenment ideas was translated into, and transformed fundamentally, the vernacular languages of East Asia, prompting these cultures themselves, then and today, to theorize their own traditions through a largely Western conceptual structure.

The complexity and the politics of this process of synchronizing the East Asian languages with the vocabulary of Western modernity, and the role that the Chinese literary tradition served as a resource for constructing this vocabulary, have been discussed in considerable detail by Columbia professor Lydia H. Liu. In thinking through the impact of this newly emerging conceptual structure as it surfaced and reconfigured the discourse of modern Chinese academic literature, Liu herself probes the “discursive construct of the Chinese modern.” “I am fascinated,” says Liu,

... by what has happened to the modern Chinese language, especially the written form, since its early exposure to English, modern Japanese, and other foreign languages. ... The true object of my theoretical interest is the legitimation of the “modern” and the “West” in Chinese literary discourse as well as the ambivalence of Chinese agency in these mediated processes of legitimation.

READING PHILOSOPHICAL TEXTS PHILOSOPHICALLY

The preparation of any new Sourcebook in Confucian philosophy must take account of another recent confluence of circumstances that is promoting a reevaluation of the
classical Chinese corpus. First, we are living in exciting times. A continuing series of truly dramatic archaeological digs in China are providing us with earlier versions of extant texts that have not suffered the distortions unavoidable over the course of some two thousand years of transmission. Again, these same finds are also offering us access to documents that disappeared from sight millennia ago. In many cases, these recovered texts as they surface are requiring a reassessment of our previous understanding of the principal philosophical works that are defining of the classical period, and that have served the tradition as its canonical core ever since. The recent recovery of these new versions of existing philosophical texts and the further discovery of many others that have been long lost, in occasioning the retranslation of many of the classics, have provided an opportunity for philosophers to step up and rethink our standard renderings of the philosophical vocabulary.

At the very least, these newly available archaeological resources provide a compelling reason for the retranslation of the selected excerpts from the seminal texts included in the Sourcebook. However, it is possible, in addition to the reevaluation made necessary by these new and exciting finds, there is yet an even more compelling reason to take up the project of offering a new Sourcebook in Confucian philosophy. Until recently, most professional Western philosophers have had little interest in claims on the part of proponents of Confucian philosophy that there is much of philosophical significance in the texts of ancient China. Indeed, it can be argued that geographical rather than philosophical criteria are being evoked in deciding what is philosophical. Again, an exclusively Western philosophical narrative rather than family resemblances that obtain among a plethora of philosophical traditions continues to be appealed to as a reason for excluding alternative philosophical traditions from proper investigation. As a consequence, texts that are profoundly "philosophical" are not being treated as such within the sanctum of professional philosophy. William James was almost right when he began his 1901 Gifford lectures at Edinburgh University by admitting that, "to us Americans, the experience of receiving instruction from the living voice, as well as from the books of European scholars, is very familiar... It seems the natural thing for us to listen whilst the Europeans talk." James is reporting on a self-understanding of the discipline of professional philosophy that is in important degree alive and well more than a century later. The only caveat offered here would be the he would have been more accurate had he included the Asian and African philosophers along with the Americans as the seemingly "natural" audience for European philosophy.  

**PHILOSOPHY IN REVOLUTION: OPENING A SPACE FOR CHINESE PHILOSOPHY**

But we need not invoke the profundity of Chinese philosophy to problematize some of the persisting assumptions within the Western tradition that have excluded Chinese philosophy from consideration as philosophy. Indeed, the revolution currently taking place within the Western philosophical community itself might be fairly described as an attempt to "think process" and to reinstate wisdom. It provides an opening and an invitation to take Chinese philosophy more seriously. That is, an internal critique continues to be waged within professional Western philosophy under the many banners of process philosophy, hermeneutics, post-structuralism, postmodernism, neo-pragmatism, neo-Marxism, deconstructionism, feminist philosophy, and so on, that takes as a shared target what Robert Solomon has called "the transcendental pretense"—idealism, rationalism, objectivism, formalism, logocentrism, essentialism, the master narrative, onto-theological thinking, "the myth of the given"—the familiar reductionistic "isms" that have emerged as putatively novel choices to allow philosophers to switch horses on the merry-go-round of systematic philosophy. In place of a Cartesian philosophical language that privileges the function of clear and distinct ideas in our erstwhile quest for an objective, apodictic certainty, vocabularies of process, change, particularity, creative advance, and indeed productive vagueness have increasingly come into vogue. These recent developments in Anglo-European philosophy have themselves begun to foreground an interpretative terminology more relevant to the articulation of Confucian culture.

The main problematic in a Cartesian dualistic worldview was one of closure articulated in the vocabulary of the quest for certainty guaranteed by clear and distinct ideas, the attainment of objective truth, and the reconciliation and ultimate salvation that follows from it. By contrast, a main problematic in the correlative cosmology we associate with "process turn" in Western philosophy in the many varieties enumerated above, is one of personal cultivation and disclosure—that is, of an aestheticism. To use Chinese terminology, the aspiration of such cultivation is wisdom, and the creative extension of an evolving cultural pattern of becoming consummately human (dao 道) that is ultimately derived from the uniqueness of those persons who contribute to it. There is a synergy in being shaped by and in turn shaping the world around one. Novelty emerges in the interface between the force of environing natural, social, and cultural conditions, and one's own creative contribution to one's context.

One of the most interesting ramifications of the increasing popularity of process language, from the perspective of our present project, is that the stimulation offered by the need to better understand Asian sensibilities, is in fact recursive. While process vocabularies are leading to increasingly productive interpretations of the classical Chinese world, these process interpretations of Chinese texts in turn provide us with new lenses through which to see our own Western sensibilities. Previously ignored or misconstrued elements within our own cultural self-understanding are beginning to receive new and decidedly more coherent interpretations.

Classical Chinese cosmology subscribes to the mantra, "the only kind of creativity is situated co-creativity." And, in the wake of the process thinkers A. N. Whitehead and John Dewey, a sustained reflection by philosophers on the alternatives to transcendentalism and the many dualisms that mark its presence that are to be found in the classical Chinese assumptions about cosmic order may pay us important philosophical dividends. The pervasive Chinese
assumption about the always emergent nature of order might at this particular historical moment provide us with a salutary intervention in the Western philosophical narrative. That is, in this classical Chinese worldview there is an alternative nuanced and sophisticated processual way of thinking about cosmology that can join this ongoing internal critique of transcendentalism taking place within the professional discipline of philosophy itself. Simply put, with the present surge of interest in Whitehead and particularly the American pragmatists, these newly emerging Western versions of process philosophy as they mature within our own philosophical culture can, with profit, draw both substance and critique from a Chinese tradition that has been committed to various forms of process philosophy since the beginning of its recorded history.

The happy conclusion that may be anticipated from these recent developments is that an era in which philosophy and philosophical thought have been considered essentially Anglo-European monopolies is drawing to a close. And further, while Western philosophy—primarily British, French, and German philosophy—constituted the mainstream curriculum for the discipline of world philosophy in the twentieth century, the revolution that is taking place within the Western academy itself presages a time when the process sensibilities pervasive in the long Chinese philosophical narrative may well become increasingly relevant in helping us find our way forward to a more inclusive understanding of world philosophy.

NOTES
2. Friedrich Nietzsche, *A Nietzsche Reader*, trans. by R. J. Hollingdale (Harmondsworth: Penguin, 1977), 86. One wonders what in the early days of these languages would prompt the French speakers to understand the sun as masculine and the moon as feminine, while their German cousins thought the opposite.
3. Saussure uses the analogy of a chess game, where *langue* are the fixed rules that govern the game while *parole* are the actual, varied moves made by different people that come to constitute any particular game.
4. I am “borrowing” this distinction from Saussure because I do not want to endorse any kind of structuralism that would allow for a severe separation between *langue* and *parole*, instead siding with the sentiments of a Zhuangzi or a Mikhail Bakhtin who would see these two dimensions of language as mutually shaping and evolving in their always dialectical relationship. Utterances gradually change the structure of language, and the changing structure orients and influences the utterances that it makes possible. For them, what we think about and how we think about it are co-terminous and mutually shaping.
7. Hans-Georg Gadamer uses “prejudices” not in the sense that prejudice is blind, but on the contrary, in the sense that our prejudices can facilitate rather than obstruct our understanding. That is, our assumptions can positively condition our experience. But we must always entertain these assumptions critically, being aware that the hermeneutical circle in which understanding is always situated requires that we must continually strive to be conscious of what we bring to our experience and must pursue increasingly adequate judgments that can inform our experience in better and more productive ways.

The Impossibility of Literal Translation of Chinese Philosophical Texts into English

Tian Chenshan  
BEIJING FOREIGN LANGUAGE UNIVERSITY  
CSTCEWR@BFSU.EDU.CN

This is a problem that has perplexed scholars for a very long time. How, indeed, may one approach the task of translating the works of Chinese philosophers into a foreign language? It is necessary to free oneself from the way of thinking that prompts questions like “How do we translate x term into English?” Word-for-word translation is not a prerequisite for the successful introduction of Chinese philosophy to the outside world. There is no term in any Western language that corresponds perfectly to its Chinese counterpart, as there is virtually no shared vocabulary between the two linguistic regions. The languages of China and the West are imbued with their own unique worldviews, epistemologies, ideologies, and value systems. One may observe all of these features in the non-dualistic (一多不分) nature of the Chinese philosophical canon and the script it was written in. The conceptual symbolic languages (概念符号语言) of the West are representatives of the dualistic (一多二元) system, which manifests its own version of these features. Their philosophical cultures are bound up with the singular numenon of the “God” concept and the idea that human beings are, like all things in the universe, individuals. The reason they differ so greatly from the Sinitic languages is because they feature a completely different cosmological narrative.
Thus, the languages of China and the West tell their own stories in completely different ways by making use of concepts, vocabularies, and systems of logic that are strikingly dissimilar. Since the two do not share any interchangeable functions, people have encountered serious problems when translating Chinese philosophy into Western languages, and have also brought about much confusion. This has particularly been the case when missionaries have been the translators; the comparative philosophers David Hall and Roger Ames described the English-Chinese and Chinese-English dictionaries they produced as “disasters.” With a little bit of reflection, one finds fundamental differences in almost any example of word-for-word translation between Chinese and English. Even everyday words like “everyone,” “goodbye,” and “sorry” differ greatly from their counterparts in Chinese. After all, is not a junzi君子, the common term, conveys its primary characteristics. Ames was able to explain and translate many Chinese classics by making use of this comparative method of philosophy.

When the Chinese translate canonical texts from the West, for example, Marxist works, a relatively word-for-word rendering is sufficient to produce a Chinese-language edition, because Marxism is in itself a manifestation of the Western non-dualistic (一多不分) way of thinking. However, many mainstream dualistic (一多二元) cultural works from the West, once translated into Chinese, seem to lack equivalence to a large extent. Admittedly, relevant terms have entered the Chinese lexicon during the translation process; for example, “liberty” and “freedom” have both been rendered as ziyou自由 in Chinese. However, Chinese lacks two different words for this concept as English does, and the concept in English is inseparable from a supposed existence of God which is seen as existing alongside individual human beings. The word ziyou in Chinese is thus suitable to the Chinese language but is in fact completely inequivalent when compared to the English concepts of liberty and freedom.

If we comprehend ziyou on the basis of its intrinsic meaning in Chinese, we need to consider the way these two characters were used in the Xiao Yao You 逍遥遊 (Carefree Wandering) chapter of the Zhuangzi 莊子 as denoting a free and unfettered state. Alternatively, one may consult the passage in the Analects of Confucius which reads: “When I was 70 years of age, I could follow my heart’s desires without overstepping the line”七十從心所欲而不踰矩. Ziyou in the Chinese classics carries this kind of non-dualistic (一多不分) philosophical meaning, with the implication that people may only attain a carefree state if they understand the relationship between human beings and the heavens, and become at one with every manifestation of nature. This concept is completely at odds with the idea of freedom and liberty in English which is supposedly bestowed to man by God. That concept of personal freedom is a freedom from all kinds of obstacles and is expanded to the social and philosophical realm; I do as I please, and I will not accept interference from others, especially the government, so that I may indulge in desires of the material world. Thus, we have a view that any kind of learning originating from Western philosophy and touching upon Chinese and Western philosophical systems must be based on sources written in a Western language; one may not rely solely on Chinese translations of Western materials.

Any kind of scholarly work in Chinese that has been translated from a Western language exhibits structural differences originating from the semantics of Western culture, and may differ substantially from its source text.
Translating Today’s Chinese Masters

Dimitra Amarantidou  
EAST CHINA NORMAL UNIVERSITY  
AMARANTIDOU@126.COM

Daniel Sarafinas  
UNIVERSITY OF MACAU  
DSARAFIN@GMAIL.COM

Paul J. D’Ambrosio  
EAST CHINA NORMAL UNIVERSITY  
PAULJDAMBROSIO@HOTMAIL.COM

A wave of translating contemporary Chinese thinkers is sweeping through sinology and philosophy. Whether we are collectively following the “spirit of the times” or privately tackling and enjoying a peculiar philosophical discourse, a few things seem to be certain: for the past decade Western scholars have been increasingly interested in engaging with contemporary Chinese thought. The Chinese government has been consistently and generously sponsoring efforts to make works by contemporary thinkers accessible in foreign languages as well. Part of the Chinese Communist Party’s Soft Power Initiative is to expand the country’s cultural appeal, and academically speaking, philosophy is at the top of the list. The proclaimed aim of translating and publishing one hundred books a year by the central government at one yuan per character, are indeed telling. China is serious about cultivating a growing Western interest which extends to not merely Confucius’s sayings and Tang poetry, but also what modern thinkers of a rising world power have to say. In this endeavor, Western academia is also increasingly attentive to contemporary Chinese philosophical discourse, and China is envisioning the possibility, or the necessity, of making herself heard, recognized, and appreciated as a valuable contributor in global philosophical discourse.

But after the grants have been applied for, foreign press contracts secured, and the World Congress of Philosophy concludes, global cultural trends and generous funding projects boil down to the bittersweet nitty-gritty of translation. Along with the perennial issues of lexical equivalence and (un)translatability, there is a broader question that we as translators of contemporary Chinese thought are constantly faced with: How do we deal with a philosophical discourse vastly different from its Western counterpart? To put it more directly, would China’s Soft Power Initiative be better off leaving out monographs that so blatantly violate the most basic norms of scholarship in Western academia? Will our colleagues in North America and Europe simply laugh at the not so uncommon misquotations, blatant misreadings, and lack of references in books by China’s top philosophy professors? It is impossible to offer a single, definitive answer without a significant amount of reflection. Pointing to a larger cultural context, these questions underscore the need to think beyond word-level choices or grammatical adjustments and recognize what is uniquely Chinese about the contemporary Chinese philosophical discourse. Translation of modern Chinese thinkers is about more than managing the intricacies of a different linguistic medium—it is entering a different academic paradigm.

One characteristic of contemporary Chinese philosophy which might help reveal some of the problems unique to translating is that many of its authors can be, and in fact are, viewed not as academics, but as Masters. In the Western world it would be somewhat analogous to saying that they are not professors but philosophers—i.e., real philosophers. Continuing in the tradition of the Zhi or “masters”—those prominent thinkers and charismatic teachers of antiquity who led by example and are important insofar as they are personal role models in addition to offering philosophical reflections—contemporary Chinese thinkers demonstrate a predilection for philosophical discussion that often operates through the suggestiveness, ellipsis, and polysemy implemented in literary language. Before explaining away Chinese scholars’ relative disinterest in extreme precision of expression and abstract logical exposition, we might want to imagine ourselves for a moment on the other side of this philosophical coin. They are not trying to be academic commentators of philosophy—they are trying to philosophize. It would then perhaps be easier to see that Western notions of abstraction, argumentation, and reasoning are not rejected by Chinese thinkers. Rather, these methods are simply taken as “missing the mark.” Whereas solid, rational expositions are detached from and make no appeal to emotional realities, emotive truths contained within a line of poetry, for example, are considered as really getting to the heart of a matter. At the risk of making an “essentialistic” claim, we could say that this method of discourse has been the norm through most of the history of thought in China. The standard was set by the original Master himself, Confucius. Interspersed with quotations from the Book of Songs, the Book of Documents, and other classics, his claims were legitimized through emotive recognition of truth rather than through the impersonal syllogisms and logic of rational demonstration.

Grounded in an emotive reality, Chinese thinkers’ rhetoric is oriented towards what is considered “concrete significance.” From a “Master’s perspective,” extreme precision, abstraction, and rigor can only exist on a theoretical level—and are thereby of limited importance. While unquestionably robust, Chinese theories have always been consciously checked by practical considerations and the demand to address a concrete, contemporary reality. The concrete question of “how will this affect my life,” which really means “how will this influence my relationships with others,” is never far from a Master’s thought. Our challenge as translators lies in the decision to either effectively communicate this attitude to a Western audience with its preconceived notions of how philosophic discourse ought to be conducted in mind, or disregard the Western audience’s biases and reproduce the thinker’s ideas and style of presenting these ideas as close to the Chinese as possible, letting them stand on their own merits.

When, following the long tradition of Masters, renowned contemporary thinkers such as Li Zehou or Yang Guorong begin or end a paragraph with a quotation from the Book of Changes or a verse from the Book of Songs—two of the most obscure texts in the history of human thought, as translators...
we are left asking: To what extent should we explain this or that quotation? If that verse from the *Book of Songs* is never explained by the author, but used instead to evoke certain emotions or cue up a particular attitude from an audience presupposed to understand the reference, do we simply leave Western readers to figure out the associations for themselves? No matter whether we define a particular author’s intent or not, and no matter how we decide to define it, the simple fact remains: these quotations are meant to get a point across that is just as much emotional as it is rational. But how do we translators strike a balance between preserving a specific discursive style and helping readers get that point? The moment we explain or footnote these quotations are we not devaluing what the author, as a Master, is doing? Yet if we strive towards some degree of functional equivalence, how do we deal with the fact that our readers will likely be left confused? The challenge for Western readers, however, extends far beyond reflecting on what the passage means or why it was quoted. The very format employed may be seen as questionable, or, more cynically, as unprofessional. “Leave the wishy-washy emotions and feelings to poetry and literature,” one might protest, “philosophy is the realm of reason and clarity.” But in these instances the thinkers in question are not opting for one rhetorical trope over another, they are operating within an entirely different discourse that translators are asked to understand and take on its own terms.

Another more troublesome issue in contemporary Chinese academic practice is also explainable in terms of the age-old tradition of Masters attitude towards discussing philosophical issues. Disinterest in precision and appreciation of the general “feel” of a text occasionally result in blatant mistakes in the monographs we work on. How should we deal with those mistakes? Here is a characteristic example.

One of us has worked very closely with Li Zehou on his *Response to Michael Sandel and Other Matters*, not only translating the book, but publishing several papers and hosting a conference on this theme. This book is extremely popular in China—where Li Zehou is perhaps the most well-known Chinese thinker and Michael Sandel is easily the most popular Western philosopher. Li’s criticisms of Sandel are generally quite broad, but oftentimes inaccurate. For instance, Li makes some questionable comments regarding the discussion of the 1884 case of survival cannibalism “The Queen vs. Dudley and Stephens” in Sandel’s book *Justice*. Following a shipwreck, Tom Dudley, Edwin Stephens, and Queen vs. Dudley and Stephens make a decision to treat others as they would want to be treated. This means that if they were in the same circumstances, they would also have chosen to be cannibals. Li comments directly on Sandel, claiming that in this case the utilitarian principles should take priority over treating all humans as ends. Interestingly, however, he does not credit Sandel with steering his readers to exactly the same interpretation. Li ends his description by claiming the primacy of “consideration for the concrete circumstances, and the role of emotions in the decision.” He strongly implies, though does not directly state, that his view is an alternative to Sandel’s, when in fact it is almost identical.

In another example of conscious or unconscious misinterpretation, Li writes, “So why, as Sandel asks, should events for the revival of Nazism be forbidden but not events opposing racial segregation? This is not something that principles or suppositions such as the ‘veil of ignorance’ or ‘neutrality’ explain.” When completing the final draft of this translation for publication, the co-translator, Robert Carleo, was caught in a dilemma. Carleo wrote a comment in Trackchanges to co-author Paul D’Ambrosio: “This statement seems just too dicey for me, and I am tempted to delete it. (It’s wrong, isn’t it?) Li seems to think that these ideas work on a level below his own. . . . Is there any way to salvage this rare mention of topics actually close to the theme of this book?” The point that Carleo does not explicitly mention, but which adds greatly to his “dicey” impression, is that Sandel is in fact one of the most famous opponents of ‘veil of ignorance’ and ‘neutrality’ explanations, yet Li presents him as their champion. What should the translator do in this case? In which situations and to what extent are we legitimized to act as textual critics?

Such issues seem all the more difficult to handle when we reflect on the contradictory aspects they entail. They are private and shared, necessary while unanswerable, empowering and restrictive, a source of frustration as well as joy. Just as much eye-opening as they are maddening. They are common for anyone involved in the realities of contemporary Chinese translation. They reveal the challenges of interlingual communication in general, but, more particularly, they point to the need for translators to re-define, again and again, our role in each situation, and re-affirm our responsibility to decide for ourselves how to best tackle it. Carleo’s simple question “It’s wrong, isn’t it?” can help open up a larger discussion and increase awareness, but it does not, as in Carelo’s own experience, promise any definitive answer to our own, very specific problems. We cannot simply induce a widely applicable principle out of the individualized ways in which Carleo, or any translator, decides to deal with his or her dilemmas. These dilemmas are, and in fact *should* be, private and perpetual.

At the end of the day, holed up in our working spaces, staring at a blinking cursor, we need to answer these questions for ourselves. But if translation is an act of communication, we might not simply want to make sense to ourselves and our peers. We may need to be convincing, perhaps, to our own “ideal readers.” In our experience, whether we are dealing with the big question about the role of the translator or with those countless little questions about this or that word, it is helpful to bring to mind from time to time our own “Masters,” our “personal heroes” or sources of authority. How would I explain my choice to Henry Rosemont, Roger Ames, or Li Zehou and Yang Guorong?
Three Thoughts on Translating Classical Chinese Philosophical Texts

Edward L. Shaughnessy
NANKAI UNIVERSITY
ESHAUGHN@UCHICAGO.EDU

There are disquisitions on the art of translating enough to fill a good-sized library, written by scholars who have devoted years to thinking about the problems involved in turning literature written in one language into literature of another language. In China, many colleges and universities have departments or even whole schools dedicated to translation. I have read only a very small portion of the works in those libraries and have never had the good fortune to take a course in translation. The closest I have come to receiving an education in this art is a long-time friendship with David Tod Roy (1933–2017), one of the Western world’s greatest translators of Chinese traditional literature (his five-volume translation of Jin Ping Mei 金瓶梅, The Plum in the Golden Vase, is already recognized as a classic of this genre), but despite often discussing questions of translation with him I would not claim to have learned anything systematic about the topic. Nevertheless, throughout what has now grown to be a lengthy career of studying ancient Chinese literature I have been consistently concerned with problems of translation in my own work, and so the following thoughts on translating classical Chinese philosophical texts may not be entirely uninformed.

I propose to draw on my own past published translations to address three discrete and yet intimately related aspects of translating classical Chinese texts, especially what for want of a better term we can call philosophical texts. These three aspects are etymology, paleography, and sonics. By etymology, I intend both the etymology of the original Chinese word to be translated, and also the etymology of the English (or whatever other language one might be translating into) translation. Paleography, on the other hand, is strictly limited to the Chinese evidence. While it is often possible to use traditional dictionaries and other scholarly studies to trace ancient forms of the characters used to write particular words, in modern times it has become increasingly possible to augment this traditional evidence with newly unearthed evidence from throughout the first millennium that the Chinese language was being written. This is particularly important for me, since this is precisely the time when most of the literature that I study was first being written. By sonics, I mean at least some general attempt to approximate the sound of the original. Needless to say, it is not possible to make English sound like Chinese. Nor do I think it is generally advisable to try to replicate rhyme used in the original. However, I think it is often desirable to represent the rhythm of the original. And, as I will suggest at the end of this essay, there are certain special cases when it is essential to hear the sound of the text if one is to understand its sense. Attention to the sound of the original is particularly important when translating poetry, but it is by no means limited to the translation of poetry.

While the translations that I will offer here all take into account one or more of these aspects, in the end I often find the translation to be limited not so much by the flexibility of the English language (which, after all, is almost as flexible a linguistic medium as classical Chinese) as by the willingness of readers to tolerate locutions that are explicitly foreign to them (or perhaps by my assumptions of what readers might be willing to tolerate). This brings us to the question of the fundamental purpose of translation: is it to make the original Chinese text (or original text in any language, for that matter), which we might assume reads naturally in the original, read equally naturally in English (or in any other target language), or is it to hew as strictly as possible to the semantics and syntax of the Chinese original even at the cost of producing an unnatural-sounding English translation (though with the possible benefit of expanding—ever so slightly—what might be perceived as being natural-sounding English)? My answer to this question, as I will demonstrate with various translations, is inevitably something of a compromise. I strive to present as literal a translation as possible, one in which readers familiar with Chinese may still detect the Chinese original, but also that readers unfamiliar with Chinese might read as natural English. Needless to say, it is often difficult to satisfy both of these competing interests.

ETYMOLOGY

I suspect that most translators pay at least some attention to the etymology of key terms, both of the original language and of the target language into which they are translating. However, in the case of translations of classical Chinese philosophical texts, probably none has paid quite the attention that Peter A. Boedberg, the legendary professor of Chinese at the University of California at Berkeley, did so in two classic articles published in the 1950s. The first of these, entitled “The Semasiology of Some Primary Confucian Concepts,” began with a call for translators of literature—“philo-logists” all—to show the same receptivity
to foreign influence as their scientific brethren to use the riches of all involved languages to invent new words.

For the humanist, however, linguistic barriers still stand as inviolable as if they were property lines of demarcation between cultural autarchies. Proud of his own heritage, often to the point of self-sufficiency, the humanist—who is essentially a “philo-logist,” that is, a lover of his native logos, or, at best, of that of his cultural area or subcivilization—is loath to permit the infiltration of the tiniest logospore of foreign extraction into the well-guarded precincts of his native literary tradition. In moments of magnanimity or spiritual weakness, he concedes the possibility of admitting a few spores, but only for the purpose of observation in the test tubes of his botanical laboratory. They may sojourn there under quaint labels in semilearned transcription, transliteration, Romanization, or Latinization, or “in the native character,” so that their exotic origin, and suspected virulence, could be easily apprehended by visiting minors or other innocents. Under no circumstances can these cells be allowed to come into mixogamous contact with the jealously protected gynoecia of the carefully nurtured word-plants of his native soil. The risk of having the beautifully laid-out garden turn into a jungle of linguistic hybrids and neological monstrostities is too appalling, and the unhappy asthmatic horticulturist rededicates himself anew to the desperate holding actions against the clouds of coryza-laden pollen drifting from beyond his cultural horizon.\(^3\)

The thought of trying to translate this passage into Chinese, whether classical or modern, provides some notion of the difficulty of accepting the translations that Boodberg proposes for what he identifies as “some primary Confucian concepts,” such as junzi 君子, de 德, ren 仁, and yi 義, among others. For junzi 君子, Boodberg’s “lordling” is perhaps readily intelligible, and I myself have often used his compromise “lord’s-son” for this term. However, despite his vigorous argument in defense of neologism, his “indarrectitude” for de 徳 has, as far I know, never been used even by his most ardent disciples. While the translations “co-humanity” and “selfshipful compropriety” or “proper selfshipfulness” for ren 仁 and yi 義 are obviously intended to shock readers out of a too ready familiarity with these important virtues, Boodberg’s nods toward “humanity” and “propriety,” based on the graph’s fundamental components ren 人 “human” and wo 我 “we,” are certainly acceptable by any standard. Indeed, such commonly seen translations of ren 仁 as “benevolence,” “perfect virtue,” “altruism” or “Goodness,” none of which acknowledges the “human” in the word, not only fail to acknowledge the word’s original Chinese etymology, but are—each in their own way—far more limited in their range of meaning than the simple and straightforward “humane” or “humanity.”

It is obvious that in this 1953 article, Boodberg was being intentionally provocative—and also intentionally playful—in his more extreme etymological neologisms. The provocation—and the playfulness—was only magnified in his even better known 1957 Harvard Journal of Asiatic Studies article “Philological Notes on Chapter One of the Lao Tzu.” After providing etymological analyses of almost every individual word in the first chapter of the Laozi 老子, Boodberg offered the following translation of the entire chapter (the scansion of the Chinese text being that given by Boodberg himself).

道可道非常道
名可名非常名
無名天地之始
有名萬物之母
故
常無欲以觀其妙
常有欲以觀其徼
此两者同出而異名
同謂之玄
玄之又玄
眾妙之門

Lodehead lodehead-brooking : no forewonted lodehead;

Namecall namecall-brooking : no forewonted namecall.

Having-naught namecalling : Heaven-Earth’s fetation,

Having-aught namecalling : Myriad Mottlings’ mother.

Affirmable,

Forewont

Have-naught

Desired—for to descry in view the circuit-luminaria;

These pairing ones at-one

Egressing,

Diverse namecall :

At-one—bespeak such : Darkling,

Adarkling such, again adarkling

The thronging sublinaria’s gate.
The Jabberwocky quality of the translation is much of its fun, but Boobdberg was not just being needlessly provocative. Take his translation of the famous first two lines:

道可道非常道，
名可名非常名:

Lodehead lodehead-brooking: no forewonted lodehead;

Namecall namecall-brooking: no forewonted namecall.

These lines are usually translated as something like “The way that can be spoken of is not the constant way; the name that can be named is not the constant name” (D. C. Lau) or “The Tao that can be told is not the eternal Tao; the name that can be named is not the eternal name” (Wing-tsit Chan). Boobdberg is absolutely right in principle to note that the verbal dao 道 following ke 可 ought to have a similar meaning to the nominal dao 道 that begins the sentence, in the same way that the verbal ming 名 “to name” is the same word as the nominal ming 名 “name” in the second line. And he is also right in practice to note that the verbal form dao 帶 “to follow” or “to lead” (along a road) is not only strictly cognate with dao 道, but is indeed just another form of the same word. An ideal translation ought to reflect this relationship between the noun and verb of the first sentence here, just as the translations noted above do so for the second sentence. Whether Boobdberg’s “Lodehead lodehead-brooking” is the best way to reflect this relationship is perhaps best left for the individual reader to decide. Boobdberg’s “lodehead” is an elaboration of “leadway” or “lode,” for which he provides the glosses “(way,” “course,” “journey,” “leading,” “guidance”; cf. “lodestone” and “lodestar”), adding the “head” (shou 首) which is the phonetic component of the Chinese character dao 道 and which, as Boobdberg also notes, may also contribute to the meaning of the character, viz. “to head,” i.e., “to tend in a certain direction.” In my own published study and translation of this chapter, “On Inconstant Ways,” I attempted to do something similar, but less provocatively, offering for the two lines:

Ways that can lead are inconstant ways.

Names that can name are inconstant names.

Of course, on a superficial level, “way” or “ways” for the nominal dao 道 and “lead” for the verbal dao 道 (i.e., 帶) do not reflect the relationship between the two words any better than do “Way” and “speak” or “Tao” and “tell,” the translations of D. C. Lau and Wing-tsit Chan. I was tempted to use “roads” instead of “ways,” which has at least a closer sonic relationship with “lead” than does “way,” and which may even share some etymological relationship, however remote. But in the end I chose to use “way” (or “ways”) because it has become such an established translation for dao 道 (or the Tao or Dao) that it seemed necessary to maintain it. I can appreciate that Professor Boobdberg would regard this as an unnecessary compromise, but at least he would approve that it is not just a transliteration such as Tao or Dao.

My rendering of dao 道 as “ways” points to a perennial problem in the translation of classical Chinese philosophical texts. It is well known that the Chinese language, and especially classical Chinese, does not differentiate between singular and plural nouns. Moreover, it does not require an article before the noun, whether the definite “The Way” or “The Tao” of D. C. Lau and Wing-tsit Chan, noted above, or just the indefinite article “a” or “an,” not to mention that Chinese—whether modern or classical—knows nothing of English rules of capitalization. To me, the second of these problems is usually more important than the first (consider the difference in connotation between “the Way” and “a way”), and I find it is often possible to avoid it by using the plural without any harm to the meaning.

A related, and to me more intractable, question concerns the final clauses of these two sentences: fei chang dao 非常道 and fei chang ming 非常名. Should the fei 非 negate the entire clause (i.e., “not constant ways” and “not constant names”) or, as I have rendered it, just the following chang 常 (i.e., “inconstant ways” and “inconstant names”)? The difference between these renderings is so subtle that it may not matter, and is, in any event, probably impossible to decide (though I do think that the positive “inconstant ways” and “inconstant names” is philosophically more interesting). Once again, Professor Boobdberg:

The close juncture of fei with the following word often results in a certain syntactic ambiguity. There is, for instance, no grammatical criterion which would enable us to decide whether 非常人 also fei ch’ang jen yeh means “(X) is (not) (fei) a common (ch’ang) man,” or: “(X) (is an) uncommon (fei-ch’ang) man.” Idiom-conscious readers will incline towards the second choice, fei-ch’ang, “uncommon,” “extraordinary” being a common idiom (as an adjective), readers well-grounded in the syntactical pattern fei . . . yeh might insist on the essential correctness of the first. This dilemma must have affected the thinking of many a translator of the Lao Tzu with its two fei-ch’ang.

I bring this up here mainly as a means of segueing into the next section of these random thoughts: on the question of paleography. Professor Boobdberg did not fail to offer his understanding of the word chang 常:

S 1.5 and S 2.5 常 ch’ang, “constant,” “regular,” “common,” “ordinary,” “persistent,” “conventional,” “enduring,” never meant “eternal” or “absolute” in our sense, as wrongly by so many translators of the first couplet of the Lao Tzu. As an epithet, it is used to characterize “long-customed,” “long-vested” things and habits, both in the positive sense of “time-honored” (regular, “customary,” and “enduring”) and in the negative sense of “time-worn.” “Commonplace,” “ordinary,” and “routine.”

I agree with everything Boobdberg says here vis-à-vis the word chang, and yet the matter is not so straightforward. It has always been known that chang was the word used in the Han dynasty as a substitute for the word heng 恆 after use of that latter word was prohibited because it was
the name of Liu Heng (202–157 BC), who reigned as Emperor Wen 漢文帝 (r. 180–157 BC) of that dynasty. Heng does, in fact, have the sense “eternal” or “enduring.” More important, in both of the Mawangdui 馬王堆 manuscripts of the Laozi, both of which were copied before Emperor Wen died and thus before this taboo became effective, these clauses read fei heng dao ye 非恆道也 and fei heng ming ye 非恆名也. Thus, translators cannot be faulted for translating either the chang of the received text or the heng of the manuscripts as “eternal.” However, even though heng was almost surely the original reading of the passage, it does not require the sense of “eternal,” since it also shares with chang the sense of “constant” (but none of chang’s other senses of “common,” “ordinary,” “persistent,” “conventional”). My own rendering of “inconstant” was given in the full knowledge of the manuscripts’ reading.

In the end, I am reasonably confident that the translation I have offered above is not only philosophically more coherent than the traditional translations, but is also consistent with the earliest attested Chinese interpretations of the passage. Nevertheless, I am even more confident that it will not be the final attempt to render this enigmatic passage into English, or any other language, for that matter. But it must be the last word of this section on etymology.

**PALEOGRAPHY**

The twentieth century brought to China a new academic discipline that has had a very significant influence on the study of all aspects of Chinese literature: archaeology. The importance of archaeology was first impressed on Chinese paleographers due to the discovery of inscriptions on Shang oracle bones, but as archaeology developed over the course of the century, the awareness of this importance continued to grow. Especially since the last quarter of the twentieth century, often termed the Golden Age of Chinese Archaeology, the discovery not only of ever more inscribed oracle bones, but also thousands of inscribed bronze vessels, and—more or less for the first time—records from the Warring States, Qin, and Han periods written on bamboo and silk, prompted numerous calls to “rewrite” the history of ancient China, including also the history of early Chinese literature. More and more it is becoming recognized that studies of China’s received literary tradition should also be informed by unearthed documents of all kinds.

One of my own first forays into the translation of China’s received literature was a translation of the Mawangdui manuscript of the Zhou Yi 周易 or Zhou Changes. As noted above, the tomb from which these manuscripts were unearthed, Tomb 3 at Mawangdui, was closed in 168 BC. At the time of its discovery in early 1974, this manuscript was certainly the earliest witness available to the early text of the Zhou Yi. Probably copied in the mid-170s BC, it was some 350 years older than the otherwise oldest evidence of the text, the Xiping 西平 Stone Classics text, which was quite fragmentary. While the Mawangdui manuscript was nearly complete, at least for the hexagram and line statements of the “classic” text, and while the contents largely coincided with the received text, it perhaps not unexpectedly contained many variants vis-à-vis the received text. Many of these variants were phonetically related to the characters contained in the received text, and were generally perceived to be just different ways of writing the same words, what is known in Chinese as “phonetic loans” (jiajiez 假借字), essentially different “spellings” of the word. In the case of a line such as fei long zai tian 飛龍在天, which corresponds to the line fei long zai tian 飛龍在天 in the received text, it is clear that fei 蜂, which means “cockroach,” is indeed just a different way of writing the word fei “to fly,” now routinely written with the character 飛. Indeed, in the Han dynasty fei “to fly” was often written as 飛, not only in manuscripts unearthed from that period, but also in received texts from the time. In any event, any translation of the line that insisted on reading that word as “cockroach” instead of some permutation of “to fly” (such as “flying”) would surely be wrong.

However, not all variants are quite so easily related. Take the name of the hexagram in which the line above is found, the famous first hexagram in the Zhou Yi: Qian 乾. It is much debated just what this character, which in most contexts is usually read as gan “dry,” means in the received text. Most early (and also late) commentaries simply gloss it as jian 健 “vigorous.” This is certainly possible, but very hard to demonstrate from context since other than as the hexagram name, the character only occurs in the text once (well, actually twice, since it occurs as a duplicative binome: qianqian 乾乾): one clause of the Nine in the Third line of the hexagram reads junzi zhong ri qianqian 君子終日乾乾 “the lord-son throughout the day is (qianqian:) vigorous.” In the manuscript, the name of the hexagram is written as jian 鍵 “key.” Virtually all Chinese scholars whose work I have seen simply assume that this jian 鍵 is another way of writing jian 健 “vigorous,” which in turn is but another way of writing the qian 乾 of the received text. Thus, the circle is completed and the reading of the received text is confirmed. But at what price? The circle may be complete, but at the price of simply erasing the manuscript. In my translation of the manuscript, I insisted that if the surface reading of the manuscript can make sense, it was my responsibility as a translator to reflect that sense. To be sure, in the sentence ‘junzi zhongri jianjian 君子終日鍵鍵, the grammar requires that jianjian 鍵鍵 serve as a predicate adjective for which “vigorous” would be appropriate. However, there are also cases within the Zhou Yi in which the name of a hexagram is used in two or more different ways, as if the authors of the text were intentionally taking advantage of the inherent polysemy of the Chinese writing system. As such, jian 鍵 “key” seems to me to be a reasonable meaning for the name of the hexagram which, after all, opens the text. And so I translated the name of that first hexagram as “Key.”

If most Chinese colleagues simply shake their head at this translation (as I have seen them do on several occasions), the few who have taken note of my translation of a still earlier manuscript of the Zhou Yi, that in the collection of the Shanghai Museum, positively howl that I have transgressed against all norms of translation and exegesis. The Shanghai Museum manuscript was robbed from some unknown tomb, probably in the vicinity of Jingzhou 荊州, Hubei in the late autumn of 1993. It was purchased by the museum early in 1994. Although the archaeological context of the manuscript has been lost, as well as some two-thirds of its contents, the paleographic evidence suffices to show that the manuscript was copied sometime in the second
half of the Warring States period, probably about 300 BC. Thus, it is now the oldest witness to the text of the \textit{Zhou Yi}, even if a somewhat fragmentary witness.\footnote{The first hexagram, at least according to the sequence of hexagrams in the received text, preserved, albeit only partially, in the Shanghai Museum manuscript is the fourth hexagram, known in the received text as Meng 蒙, another word for which there are several meanings, including "type of plant, dodder;" "lush, luxuriant;" "to cover;" "to wear on the head;" "to trick;" "occluded;" "ignorant;" "confused;" "youth." In the Zhou Yi tradition, the character is variously explained as "youth" or "ignorance," with many commentators combining the two meanings (i.e., "the ignorance of youth" or "Youthful Folly"). Although the bamboo strip bearing the hexagram name as well as the first two line statements has been lost from the manuscript, the formulaic nature of the four remaining line statements shows that the manuscript must have written the name of this hexagram as 蒙. In conventional script, 蒙 is the standard graph for the word mang "long-haired dog." What could a "long-haired dog" have to do with the ignorance of youth? Chinese scholars whose work I have consulted concerning this question have invariably regarded mang 蒙 as a simple phonetic loan for meng 蒙, and thus understand its occurrence in the hexagram as meaning either "youth" or "ignorance," or one of the other meanings hallowed by the Zhou Yi tradition. For instance, Pu Maozuo 濮茅左, the editor of the manuscript, notes that the \textit{Jingdian shiwen 經典釋文} of Lu Deming 陸德明 (556–627) indicates meng 蒙 as an alternative pronunciation of 蒙, and thus assumes that both 蒙 and 蒙 stand for the same word, which he regards as meng "youth; ignorance." On the other hand, this shows that at least one text of the \textit{Zhou Yi} still in circulation by the beginning of the seventh century wrote the name of this hexagram (and the occurrence of the same graph in all of the line statements) as 蒙. There is no question that mang 蒙 and meng 蒙 were sufficiently similar in pronunciation that 蒙 could be used to write meng 蒙. Nevertheless, it stands to reason that the opposite could also be true; that is, meng could also be used to write mang 蒙. In fact, this is an argument made by the Japanese scholar Ono Yuji 大野裕司.\footnote{He notes in line statements of this hexagram such as kun mang/meng 固/蒙 and ji mang/meng 撃/蒙, "long-haired dog" makes better sense as an object of the verbs kun 固 “to bind” and ji 撃 “to hit” than does "the ignorance of youth." Even in the Six in the Fifth line statement, tong mang/meng 傳蒙/蒙, tong 傳 "young; youth" could describe a long-haired puppy just as well as a child, ignorant or otherwise. Examining another line statement in the received text of this hexagram, which, however, has not survived in the manuscript, may provide further support for reading meng 蒙 as mang 蒙, and thus as "long-haired dog." The Nine in the Second line reads:}

\begin{quote}
\textit{Wrapping the meng. Auspicious. Taking a wife: auspicious; a son can marry.}
\end{quote}

Of course, the image "bao meng" 包蒙 "Wrapping the meng" is one of those enigmas that has made the \textit{Zhou Yi} famous, and which has provided almost limitless license to translators and commentators alike. What could "wrapping youthful folly" possibly mean? The same \textit{Jingdian shiwen} that noted the existence of a text in which mang 蒙 appears where we would normally expect meng 蒙, quotes the text of Zheng Xuan 郑玄 (127–200) as reading this line as biao meng 彪蒙. To be sure, Zheng Xuan's text wrote the second character as 蒙, but what is interesting is that the original meaning of the first character, biao 彪, is "stripes (of a tiger)."\footnote{Although this line is missing from the Shanghai Museum manuscript, if it were to read both characters as biao mang 彪蒙, might we then understand it as "a striped long-haired dog"? It seems to me to make at least as much sense as "wrapping youthful folly." And so, in my translation of the Shanghai Museum manuscript, I translated mang 蒙 as "shaggy dog." I would certainly not insist that this is "the correct" reading of the \textit{Zhou Yi}, as if there were such a correct reading. However, I would suggest that it is not only a possible reading, but it is at least prima facie the "reading" of the manuscript. I would argue that as the first English-language translator of the Shanghai Museum manuscript of the \textit{Zhou Yi}, I had a responsibility to reflect in my translation how the manuscript might contribute to our understanding of the many possibilities of meaning inherent in the text. To insist that a possible reading is impossible, simply on the basis of tradition, seems to me not only to be a rejection of the polysemy that is a defining feature of the \textit{Zhou Yi}, but, perhaps more important in a general sense, it is a rejection of the call to "rewrite" early Chinese texts—a rejection of the contribution that archaeology and paleography can make in our understanding and translation of ancient Chinese texts.\footnote{The \textit{Zhou Yi} is by no means the only classical Chinese philosophical text for which unearthed manuscripts have complicated our reading—and our translations. The \textit{Li ji 鋼紀} or \textit{Record of Ritual} is usually regarded as one of the more straightforward of the Chinese classics, typically entailing relatively few problems of interpretation (at least in comparison with a text such as the \textit{Zhou Yi}). And yet here too matters are not so straightforward. Until recently, the "Zi yi" 紫衣 "Black Jacket" chapter was one of the more neglected texts in the \textit{Li ji}, despite being one of the so-called four "Zi Si" chapters, texts attributed to Zi Si 子思 (483–402 BC), the grandson of Confucius (551–479 BC), chapters that include the famous "Zhong yong" 中庸 "Doctrine of the Mean." This changed around the year 2000, with the publication first of the Guodian 郭店 manuscripts and then a few years later of the first volume of the Shanghai Museum manuscripts, both of which included manuscript copies of the \textit{Zi yi}. Questions of interpretation—and thus of translation—arise with the very first passage of the manuscripts (the second passage in the received text). In the received text, this passage reads as follows:}

\begin{quote}
\textit{子曰: 好賢如紫衣，惡惡如巷伯，則爵不濁而民作願，刑不試而民咸服。}
\end{quote}

\begin{quote}
\textit{The Master said: "If one loves worthies as does 'The Black Jacket' and hates evil as does 'The Elder of the Lane,' then entitlements will not be gratuitous but the people will do as wished, and}
\end{quote}
punishments will not be used but the people will all submit."

The Greater Odes says: "A proper model was King Wen, The ten-thousand countries acted sincerely."

This seems to be eminently straightforward. However, both of the manuscripts write the last clause of what "the Master" said differently from the received text (for the convenience of publication, I here use a relaxed transcription method):

夫子曰：好美如好緇衣，惡惡如惡巷伯，則民咸服而型不屯。

詩云：儀型文王，萬邦作孚。

M#1: Our Master said: "If one loves beauty as he loves 'The Black Jacket,' and if one hates evil as he hates 'The Elder of the Lane,' then the people will all be strong and the model will not crumble."

The Poetry says: "A proper model was King Wen, The ten-thousand countries acted sincerely."

There are several reasons to question the correctness of the received text: The phrase in the manuscripts is 型, which the editors of 郭店楚墓竹簡, the influential first publication of the Guodian manuscript, suggest be read as "form" and "reform" in the sense of the relationship between the two words might be gained by translating them as "form" and "reform" in the sense of "reform school!", respectively. The final character of the phrase in the manuscripts is 呉, which the editors of 郭店楚墓竹簡, the influential first publication of the Guodian manuscript, suggest be read as a protograph or phonetic loan for 陈 "to move," thus suggesting that the 型 of the manuscript be read in the sense of 刑 "punishment": i.e., "punishments will not (move:) be used." While this reading is certainly sensible according to most principles of Chinese textual criticism, it is important to note that it has surely been influenced by a desire to accord with the reading of the received text: 乖 乖 "to move," thus suggesting that the 型 of the manuscript be read in the sense of 刑 "punishment": i.e., "punishments will not (move:) be used." There are several reasons to question the correctness of this reading.

First, whether in the received text or in the manuscripts, the Shijing quotation that concludes this passage includes the graph 型 clearly used in the sense of "model":

儀型文王，萬邦作孚

A proper model was King Wen, The ten-thousand countries acted sincerely.

Since this is the one graph that links the two quotations of this passage, it would seem that a consistent reading for it might be preferable to one in which it is read in two different senses. Second, the general context of at least the first half of the manuscripts version of the Zi yi stresses that the ruler should serve as a model for the people; if this first chapter of the work is to serve as an introduction to this theme, it would be reasonable to expect 型 here to be read as "model," if an appropriate meaning of 呉 can be given. As it turns out, 呉 is the protograph of a word family most members of which have connotations of "confused," "dull," or "broken": e.g., 乱 or 冫 = "confused, stupid"; 冫 鈍 "dull"; 冫 頓 "worn, crumbled." Without any need to posit a phonetic loan, perfectly good sense can be made by reading the phrase "the model will not crumble." On the other hand, in the received text of the Zi yi, the sequence of the individual chapters is radically rearranged, such that the main theme of the first chapters is the proper use of punishments by government, and thus it is easy to understand why an editor of that text might understand 型, whether written as 型 or as 刑, as "punishment." And in the political context of the Han dynasty, when the received text of the Zi yi was undergoing its final editing, the most influential arguments at court were against the use of corporal punishment; it is easy to see why an editor would be tempted to read this line as "punishments will not be used but the people will all submit."

By considering the early forms of the two graphs in question, 呉 and shi 試, it is easy to see how such confusion could have arisen. The 呉 of the manuscripts is written, while the seal script form of 乖, the protograph of 試 (by way of 型) is written. The two graphs are so similar in shape that it would have been natural for a reader who interpreted 型 to mean "punishment" to assume that was but a variant form of , thus making—for him—good sense of the phrase: "punishments will not be (tried:) used." Good sense, indeed, but a very different sense from the sense that the manuscripts wished to convey.

I review all of this paleography for this discussion of translating classical Chinese philosophical texts because most of the studies that I have seen employ various paleographic arguments to read the manuscripts in the same way as the received text: some variation on "punishments will not be (tried:) used." This is true too of the most authoritative English translation, by Scott Cook, which adds several words to read: "then the people will all submit, and yet the implements of punishment will not be blunted [through overuse]." I think this sort of translation is all but indefensible in terms of the paleography of the manuscripts. What is more, based on what I said above about the main themes of the manuscripts' version of the Zi yi, on the one hand, and the Li ji received text, on the other hand, although the translation certainly reflects the sense of the received text, it is quite incompatible with the sense of the manuscripts. If one had approached the manuscripts without reference to the corresponding received text, I am sure it would have been a relatively easy matter to give the sort of translation that I have offered: "the people will all be strong and the model will not crumble." But, just as in the cases of the Zhou Yi manuscripts examined above, because scholars steeped in the tradition of the Chinese classics assume that the received text is somehow "correct," then they view it as necessary to adapt their reading of any new manuscripts that might be unearthed to conform to their prior understanding of the received text. It seems to me, once again, that this misses the great opportunity that
archaeology and paleography offer us to “rewrite”—or at least to re-think—classical Chinese philosophical texts. It goes without saying that this rewriting or re-thinking should also inform our translations of those texts.

Having argued for the importance of unearthed manuscripts, I should hasten to add that any translation that one might offer of classical Chinese philosophical texts will be contingent upon one’s own purpose in translating. If one proposes to translate a text as it has been transmitted over the ages and as it is found in the received literature, then the evidence of any manuscript that might be discovered will only be incidental to this, and need not be reflected in the translation. On the other hand, if one proposes to translate a manuscript version of the text, then of course the evidence of the manuscript is crucial. The matter becomes more complicated if one proposes to translate the text according to its “original meaning,” even assuming that there is such a thing as its original meaning. In this case, it would seem necessary to take account of all relevant evidence, both the received text and any manuscripts that might be available. While the manuscripts are likely to be the earliest evidence, and thus according to most canons of literary criticism have the greatest weight, this will not invariably be true. In the end, the translator will often be called upon to make her own decision between two or more possible readings. After all, despite Professor Boodberg’s hopes, alluded to above, translation is an art, not a science.

SONICS
The final topic that I propose to discuss in these thoughts on translating classical Chinese philosophical texts is the matter of sonics: to what extent it may be possible or desirable to replicate the sound of the Chinese original. Needless to say, modern Chinese sounds very different from English—and from most other languages into which translators might seek to translate Chinese—not to mention the manifest difference between the sounds of modern Chinese and the sounds of ancient Chinese, if these latter can even be retrieved with any confidence. And yet the sounds of a text can play an important role in the way the reader understands the text. Let me clear, I am certainly not proposing the wholesale use of transliteration instead of translation. While there are occasional Chinese words that have been taken over into the English vocabulary, such as yin and yang for which translations such as “sunny” and “shady” would probably be less intelligible to most readers than the transliteration, transliteration generally represents little more than what Peter Boodberg has called “the mummification of foreign terms in the sarcophagus of transliteration” and “a latent xenophobia as an evidence of a cautious and precise scholarship.” What I intend by sonics has more to do with the rhythm of the original.

In classical Chinese philosophical texts, rhythm probably manifests itself most commonly in parallel structure. Parallel structure is pervasive throughout almost all genres of writing. The parallels are sometimes drawn explicitly in what Greek rhetoricians called anadiplosis (when the last word or words of one clause is or are repeated as the first word or words of the following clause) and the related sorites (a chain linking the tenets of an argument, such as the opening passage of the Da Xue 大學 or Great Learning).

I think many translators would agree that anadiplosis and sorites are such important structures that they should be reflected in any translation. On the other hand, many translators avoid replicating parallel structure at all costs. True, in the hands of a clumsy writer parallel structure can devolve into a fugue-like redundancy, and while it may have the pretense of eloquence in Chinese, in English, at least, it just sounds redundant. However, in the hands of a master writer, the parallels can sound almost Bach-like. I would suggest that a proper translation needs at least to try to sound out these parallels.

Pervasive as the rhythm of parallel structure is in classical Chinese philosophical texts, it probably goes without saying that rhythm is most important in the translation of poetry. Let us take the most famous poem in all of classical Chinese poetry—and one that has been translated perhaps more than any other poem (though not nearly as often as the Laozi) as an example: the poem “Guan ju” 關雎 from the Shi jing 詩經 or Classic of Poetry.

Let me give a not quite random sample of published translations, arranging them by date of publication.

---

**James Legge, Kwan Ts’eu (1871)**

*Kwan-kwan go the ospreys,*

*On the islet in the river.*

*The modest, retiring, virtuous, young lady: —*

*For our prince a good mate she.*

*Here long, there short, is the duckweed,*

*To the left, to the right, borne about by the current.*

*The modest, retiring, virtuous, young lady: —*

*Waking and sleeping, he sought her.*

*He sought her and found her not,*

---

Kwan-kwan go the ospreys,
On the islet in the river.
The modest, retiring, virtuous, young lady: —
For our prince a good mate she.
Here long, there short, is the duckweed,
To the left, to the right, borne about by the current.
The modest, retiring, virtuous, young lady: —
Waking and sleeping, he sought her.
He sought her and found her not,
And waking and sleeping he thought about her.
Long he thought; oh! long and anxiously;
On his side, on his back, he turned, and back again.
Here long, there short, is the duckweed,
On the left, on the right, we gather it.
The modest, retiring, virtuous, young lady: —
With lutes, small and large, let us give her friendly welcome.
Here long, there short, is the duckweed;
On the left, on the right, we cook and present it.
The modest, retiring, virtuous, young lady: —
With bells and drums let us show our delight in her.

James Legge, The Kuan Chü (1876)

Hark! From the islet in the stream the voice
Of the fish hawks that o’er their nest rejoice!
From them our thoughts to that young lady go,
Modest and virtuous, loth herself to show.
Where could be found, to share our prince’s state
So fair, so virtuous, and so fit a mate?

See how the duckweed’s stalks, or short or long,
Sway left and right, as moves the current strong!
So hard it was for him the maid to find!
By day, by night, our prince with constant mind
Sought for her long, but all his search was vain.
Awake, asleep, he ever felt the pain
Of longing thought, as when on restless bed,
Tossing about, one turns his fevered head.

Here long, there short, afloat the duckweed lies;
But caught at last, we seize the longed-for prize.
The maiden modest, virtuous, coy, is found;
Strike every lute, and joyous welcome sound.
Ours now, the duckweed from the stream we bear,
And cook to use with other viands rare.
He has the maiden, modest, virtuous, bright;
Let bells and drums proclaim our great delight.

V.W.X., The Ospreys Woo (1878)

As the ospreys woo
On the river ait,
So the graceful lass
Has her manly mate.

As the coy marsh flowers
Here and there do peep,
So the graceful lass
In his wakeful sleep.

But he seeks in vain,
Brooding night and day,
Ah me! ah me!
Tossing rest away!

As the coy marsh-flower
Chosen here and there,
So the graceful lass;
He in tune with her.

As the coy marsh-flower
Gathered here and there,
So the graceful lass,
Bells now ring for her.

William Jennings, Song of Welcome to the Bride of King Wen (1891)

Waterfowl their mates are calling,
On the islets in the stream.
Chaste and modest maid! Fit partner
For our lord (thyself we deem).

Waterlilies, long or short ones,
Seek them left and seek them right.
’Twas this chaste and modest maiden
He hath sought for, morn and night.
Seeking for her, yet not finding,
Night and morning he would yearn
Ah, so long, so long! – and restless
On his couch would toss and turn.

Waterlilies, long or short ones,
Gather, right and left, their flowers.
Now the chaste and modest maiden
Bell and drum shall hail as ours.
Long or short the waterlilies,
Pluck them left and pluck them right.
To the chaste and modest maiden
Bell and drum shall give delight.

Arthur Waley, Fair, Fair, Cry the Ospreys (1937)

“Fair, fair,” cry the ospreys
On the island in the river.
Lovely is this noble lady,
Fit bride for our lord.

In patches grows the water mallow;
To left and right one must seek it.
Shy was this noble lady;
Day and night he sought her.

Sought her and could not get her;
Day and night he grieved.
Long thoughts, oh, long unhappy thoughts,
Now on his back, now tossing on to his side.

In patches grows the water mallow;
To left and right one must choose it.
Shy is this noble lady;
With gongs and drums we will gladden her.

V.W.X., The Ospreys Woo (1878)
Bernhard Karlgren, *Kuan ts’ü* (1944-45)

1. *Kwan-kwan* cries the *ts’ ü-kiu* bird, on the islet of the river; the beautiful and good girl, she is a good mate for the lord. 2. Of varying length is the *hing* waterplant, to the left and right we catch it; the beautiful and good girl, waking and sleeping he sought her: wished for her; he wished for her but did not get her, waking and sleeping he thought of her; longing, longing, he tossed and fidgeted. 3. Of varying length is the *hing* waterplant, to the left and right we gather it; the beautiful and good girl, guitars and lutes befriend her: hail her as a friend. 4. Of varying length is the *hing* waterplant, to the left and right we cull it as a vegetable; the beautiful and good girl, bells and drums cheer her.

Ezra Pound, *Hid! Hid!* (1954)

"Hid! Hid!" the fish-hawk saith,  
by isle in Ho the fish-hawk saith:  
"Dark and clear,  
Dark and clear,  
So shall be the prince’s fere."

Clear as the stream her modesty;  
As neath dark boughs her secrecy,  
reed against reed,  
tall on slight  
as the stream moves left and right,  
dark and clear,  
dark and clear.

To seek and not find  
as a dream in his mind,  
think how her robe should be,  
distantly, to toss and turn,  
to toss and turn.

High reed caught in *ts’ai* grass  
so deep her secrecy;  
lute sound in lute sound is caught,  
touching, passing, left and right.  
Bang the gong of her delight.


Waterbirds on  
River islands!  
Shy the nymph our  
Shepherd’s chosen.

Waterlilies  
Wreathe around her,  
Shy the nymph he  
Waking, sleeping,  
Never reaches;  
Waking, sleeping,  
Longing, longing,  
Turning, tossing.

Waterlilies  
To adorn her,

Shy the nymph we  
Greet with zither.

Waterlilies  
To array her,  
To the shy nymph  
Bell, drum bring glee!

Poetry being poetry, I suspect that different readers will appreciate different translations of this single Chinese poem (although I suspect that only the most tone-deaf reader could appreciate Bernhard Karlgren’s rendition). As is well known with respect to the poetry of the *Shi jing*, its most notable sonic quality is the staccato four beats per line (or eight beats per couplet). A second feature that marks most of the poems in the collection, and certainly this poem “Guan ju,” is the end rhyme linking each couplet. Some of the translators featured above have sought to highlight one or another of these features. Thus, Arthur Cooper in the last of the translations featured here has opted to replicate the staccato beat of the original, with four syllables per line. To accomplish this, he has had to omit much from his translation. For instance, he translates the very first line, “*guanguan jujiu*” 关关雎鳩, as “Waterbirds on,” “waterbirds” translating *jujiu* 雎鳩, but neglecting entirely the first two characters *guanguan* 关關, usually thought to be the sound that the “waterbird” or “waterbirds” make. Such a free “translation” is obviously an aesthetic choice on the translator’s part, but, as I will suggest below, it may make a difference in how we understand the poem to hear what the bird or birds are saying. Also taking considerable liberties with the text is the great modernist poet Ezra Pound. The couplet beginning his second stanza:

Clear as the stream her modesty;  

As neath dark boughs her secrecy,

is a beautifully rendered image, but it seems to rearrange the original poem, taking the third line of the first stanza and more or less combining it with the first line of the second stanza.³¹

Pound makes a bow toward rhyme in his first stanza:

"Dark and clear,  

Dark and clear,  

So shall be the prince’s fere."

However, probably wisely, he did not try to sustain the rhyme throughout. For sustained rhyme, we have to turn to James Legge and William Jennings, both writing in the nineteenth century. Jennings’ second stanza would, by most standards of translation, have to be regarded as quite a success.

Waterlilies, long or short ones,  
Seek them left and seek them right.  
’Twas this chaste and modest maiden  
He hath sought for, morn and night.  
Seeking for her, yet not finding,
Night and morning he would yearn
Ah, so long, so long! – and restless
On his couch would toss and turn.

On the other hand, his first stanza seems forced:

Waterfowl their mates are calling,
On the islets in the stream.
Chaste and modest maid! Fit partner
For our lord (thyself we deem).

It is compromises such as this that have caused most translators of poetry to refrain from trying to replicate the rhyme scheme of the Chinese original.

If a translation is to be faulted for too faithfully adhering to the sonic quality of the original, both in terms of beats per line and also for the rhyme, what is the translator to do? I believe it is possible both to translate every word of the original and to approximate at least the first of these sonic effects, by translating into an English verse of modified iambic tetrameter. The following is my own rendition.

Join cries the osprey
On the island of the river.
Shy and slender is the chaste girl,
A loving mate for the lord’s son.

Up and down the water lilies,
To the left and right drifting them.
Shy and slender is the chaste girl,
Awake and asleep seeking her.

Sequencing without success,
Awake and asleep thinking of her.
Longing, oh, longing oh,
Tossing and turning to and fro.

Up and down the water lilies,
To the left and right picking them.
Shy and slender is the chaste girl,
With harp and lute befriending her.

Up and down the water lilies,
To the left and right choosing them.
Shy and slender is the chaste girl,
With bell and drum amusing her.

To be sure, the first line does not seem to adhere to the eight beats of every other line, but if I were to read this out loud, I would be sure to extend the “Join’s” across two beats apiece, replicating at least somewhat both the way the line would be recited by a traditional Chinese reader, and also approximating the long high-pitched call of the osprey. True, too, there is no rhyme, but I have tried to compensate by translating the rhyming two-character description of the girl, yaotiao 窈窕, which has long been held to be one of the keys to the poem, with an alliterative compound in English: “shy and slender.” Arthur Waley’s “Shy was this noble lady” seems entirely too understated, and also weirdly disembodied for the girl of the protagonist’s dreams.32

Before closing this discussion of the translation of “Guan jù,” I should also like to make a plea for taking seriously the cry that the osprey makes at the beginning of the first line. The words guanguan 關關 have traditionally been regarded as onomatopoeia, the sound that the osprey or ospreys make on the island in the river. Of the translations surveyed here, three (VWX, Jennings, and Cooper) neglect the sound entirely, two (Legge 1871 and Karl(gren) render it in transliteration (kwan-kwan), and one (Legge 1976) gives the functional translation “Hark.” Only two of the translators hear in the sounds a Chinese word: for Waley “Fair fair” and for Pound “Hid hid.” I can imagine that Pound was inspired by the modern meaning of guan 聯: “to close, to latch,” and as such he would have heard the fish-hawk to be commenting on the girl’s being “hidden.”33 Pound is often roundly criticized by Sinologists, when they even bother to read his translations at all, but in this rendition he combined his poetic sensibility with a certain concern for etymology, and in this respect, at least, I think his effort is the most admirable of any translation published to date.34 In the Shi jing, when birds or animals are quoted as crying out, the cries are not just pure onomatopoeia; the sounds are necessarily written in Chinese characters, and it is my contention that these characters are almost always intended to convey meaning.35 In this case, whatever sounds the ospreys, or as I would argue, just a single osprey, may have made in reality (the best description that I have read of the cry of an osprey is that it sounds like a door turning slowly on a rusty hinge),36 the poet heard it to be saying in Chinese guan-guan. While Pound is correct that guan 聯, the character used in the received text to write the sound, now means “to close,” and thus might have an extended connotation of “hidden” behind a gate, the etymology of this word has different connections. The key component of the character is guan 縮, which serves as both the phonetic element and also as a pictograph of the original meaning of the word: it is the cross-bar that locks the two doors of a double-doored gate together. In the character, the “gate” component (門) is a later addition, contributing meaning, to be sure, but not basic to the character. The original sense of the word, in addition to “to close,” is also “to join.” That this is the meaning that was intended in this poem, or at least was understood to be intended by the earliest copyists of the poem, can be seen in the way the character is written in all of the earliest manuscripts that quote the poem: guan /. Here too the gate component is secondary; the key component is , which is a pictograph of two objects or bodies linked together. This is the original form of the Chinese characters guan 縮 and guan 聯, which mean “to penetrate” or “to link together” (as strings of Chinese coins). I think it is no coincidence that this was also the standard ancient Chinese euphemism for sexual union. Whatever the osprey may have been saying, the poet heard it to be saying “join-join,” which, as the rest of the poem shows, was the main thing on the poet’s mind. If we fail to hear this, we miss an important aspect of this poem.37 I would urge that it is the translator’s responsibility to try in some way to replicate if not the sound of the Chinese, at least the sense of the osprey’s cry—as heard by the poet.

CONCLUSION
According to the Mao zhuàn 毛傳 or Mao Tradition, the canonical source for understanding the Shi jing, the sound guan-guan is simply a “harmonious sound” (sheng 和聲). Elsewhere this is expanded to be “the paired bird calls...
mutually harmonizing and pleasing each other” (niao sheng zhi liang xiang he yue ye 鳥聲之兩相和悦也). This is probably a fitting note on which to end this discussion of translating classical Chinese philosophical texts, with the hope that the original text and the translation might mutually harmonize, even if not to please each other. However, to move from the realm of aesthetics and music back to the humanities and linguistics, perhaps I should return to Peter Boodberg’s translation of the opening couplet of the first chapter of the Laozi: Dao ke dao fei chang dao, ming ke ming fei chang ming 無可道非常道，名可名非常名. After offering four possible renditions:

I. (If/when/though) the way is way-able (i.e., brooks/admits treatment as a way, it is no common way. 

(If/when/though) the name is namable (i.e., brooks/admits treatment as a name, it is no common name.

II. (When we) way (i.e., treat-as-the-way) the wayable (i.e., what brooks/admits treatment as a way, it is no common way.

(When we) name (i.e., treat as a name/its name) the namable, it is no common name.

III. If the way is wayable, (we) gainsay the common ways; if its name be namable, (we) gainsay the common names.

IV. Waying the wayable, (we) gainsay the common ways; naming the namable, (we) gainsay the common names.

To these translations, Boodberg adds the following explanation:

The polysyntactical ambiguity and seeming confusion and contradiction of the above renderings will prove, after a moment’s reflection, more apparent than real. It is not impossible moreover to construct a crude English replica of the original text reproducing all of its ambiguities.39

That “crude English replica” is of course “Lodehead lodehead-brooking; no forewonted lodehead; Namecall namecall-brooking: no forewonted namecall.” Much as I admire the spirit animating Boodberg’s translation, I think it is possible to combine his etymological skill with recent advances in paleography and also a sensitivity to the rhythm of Chinese texts to translate classical Chinese philosophical texts into other English replicas, doubtless no less crude, but perhaps closer to the reading experience of both Chinese and English readers.

NOTES
1. Much of the material in this article has already appeared in various of my more specialized writings. I repeat it here in the hopes that it may be new to readers of the APA newsletter.
5. The first of these translations is that of D. C. Lau, Lao Tzu: Tao Te Ching (Harmondsworth: Penguin Books, 1963), 57; the second from Wing-tsit Chan, The Way of Lao Tzu (Tao-fe ching) (Chapel Hill: University of North Carolina Press, 1983), 97. For “175+ translations of this chapter, almost all of them similar to these two,” see Bureaucratic Public Secrets, Lao Tzu: (175+ Translations of Chapter 1), http://www.bopsecrets.org/gateway/passages/tao-fe-ching.htm.
8. Note Boodberg’s caustic comment on the use of transliterations: “Without necessarily dislocating the delicate patterns of our traditional linguistic growth, we could surely essay, with some profit and edification, a few systematic and controlled experiments in the ingrafting of seemingly alien concepts on our linguistic texture. Healthy neology is not incompatible with literary norm, while mummification of foreign terms in the saccrophagy of transliteration, with no gesture whatever toward the courtesy of translation, is as often the symptom of a latent xenophobia as an evidence of cautious and precise scholarship; and a well-executed calque is not necessarily to be condemned without hearing as a counterfeit caconym and denied probation as literary tender; Boodberg “The Semasiology of Some Primary Confucian Concepts,” p. 319.
10. Ibid, 603–604.
11. In my article “On Inconstant Ways,” in addition to providing a close grammatical analysis of the phrases, I also examine four different passages from the Wenzhi 文致, all of which explain the Laozi as an argument in favor of flexibility and against the perpetuation of time-worn precedents. I argue also that these Wenzhi passages should be regarded as earlier than the corresponding passages in the Huainainzi 淮南子, and that they are consistent with several notable pre-Qin attitudes concerning precedents.
12. The call to “rewrite” ancient Chinese history is especially associated with Li Xueqin 李學勤 (1933–2019). He issued it first in a speech delivered in 1992, first published as Li Xueqin, “Lun xin chuan bo yu yuxue yan jiu” 深入現實與學術研究, Zhongguo chuanbo wenhua yu xian daihua 中國傳統文化與現代化 1993.1; rpt. in Li Xueqin, Dangdai xuezhe zhi xuan xue xue shi: Li Xueqin juan 當代學者自選文庫: 李學勤卷 (Hefei: Anhui Jiaoyu chubanshe, 1998), (the call is on p. 364). Li himself subsequently highlighted the word as the title of one of his books: Chongxie xueshu shi Chongxie xueshu shi: li xueqin juan bao yu chu jian bo yu xueshu yan jiu 論新出簡帛與學術研究 (Shijiazhuang: Hebei Jiaoyu chubanshe, 2002).
13. Consider the following influential statement by Martin Kern on the importance of unearthed documents for the study of the Shi jing 詩經 or Classic of Poetry, perhaps the most important of all classical Chinese literary texts, if not philosophical texts:

As a result of these finds—some of them coming from archaeologically controlled excavations, others looted from unknown sites and then sold to prestigious institutions that nevertheless vouch for their authenticity—we live, with the Poetry, in a time like no other time. Gradually, a Poetry before the Mao Poetry is emerging, and with it a “new”—if in fact earliest known—approach to the ancient songs. For scholars today is a defining moment in the history of the text, a moment on par with the Mao determination of the text; the Song challenge, the Qing inquiry, and the May Fourth departure. Yet our moment goes beyond
In my 2014 book *Unearthing the Changes*, I concluded this section on variants in the Shanghai Museum manuscript of the *Zhou Yi* with the following statement about their implications for understanding the text:

> What sort of exegetical principle can we derive from this discussion of the word *jīng* in the *Zhou Yi*? I believe the fundamental philosophical thought of the *Zhou Yi* lies in change. The linguistic usage of the text would seem to be no exception. The diviners who created the hexagram and line statements seem to have very much appreciated the different senses of individual words, in the different line statements of a single hexagram often emphasizing different aspects of a single word family. In his 1982 doctoral dissertation, Richard Kunst highlighted this polysemous quality of the text.

[[In the case of the *Yi*, it may often be the case that a given word in a single context or in several adjacent contexts was meant from the beginning to be ambiguous. Or rather, put differently, it was the polysemous inherent in a word which gave it a numerous quality and led to its incorporation in the text in contexts capable of more than one reading.]]

In the Western Zhou or Spring and Autumn periods, when the *Zhou Yi* was being composed, the words of these word families were often written with a single character. I suspect the ancient diviners were already well aware that words are variable, changeable, and that in creating the *Yi*, the Changes, they sought to exploit this feature of their language. However, later in the process of the transmission of the Changes, different schools of interpretation, which is to say the scribes of different times and places, were only able to choose a single one of these meanings to understand any given character; moreover, based on the writing conventions of their times, they were only able to use a single "correct" character to write it; Shaughnessy, *Unearthing the Changes*, pp. 65-66.

I continue to think that this is a reasonable approach to the *Zhou Yi* in particular and also to the question of translating classical Chinese philosophical texts in general.

---

14. One such example is to be seen in the hexagram called *Jing Xi*. "Pit" in the received text and *Jing Xi* is "Pit" in the Shanghai Museum manuscript. For my discussion of the polysemous inherent in this hexagram, see Edward L. Shaughnessy, *Unearthing the Changes: Recently Discovered Manuscripts of and Relating to the Yi Jing* (New York: Columbia University Press, 2014), 59–66.


16. For instance, the *Xu gua* 序卦 commentary, basing itself on the traditional sequence of the hexagrams, moving from *Zhen* 屯 hexagram (R3), which is explained as "a plant first sprouting," explains the name of the hexagram: *wu sheng bi meng* 萬生必萌 "when things are born they are necessarily making." The commentary of Wang Bi 王弼 (226–49), which serves as the basis of the orthodox exegetical tradition, explains the hexagram name as "a youth wishing to resolve that which confuses him" (*yu jue suo huo ye su huo meng*). *Zhou Yi* zhengyi (Sibu beiyao ed.). 1.19a. This is reflected in the English translation of Richard Wilhelm’s German translation of the hexagram name: “Youthful Folly”; see Richard Wilhelm, *The I Ching* of Book of Changes, trans. Cary F. Baynes (Princeton, NJ: Princeton University Press, 1950), 20.

17. According to the phonetic reconstructions of Axel Schuessler, *mang* 餕 had an Old Chinese pronunciation of mág and a Middle Chinese pronunciation of mǎng whereas *meng* 娘 had Old Chinese and Middle Chinese pronunciations of mǎng and mù respectively; see Axel Schuessler, *Minimal Old Chinese and Later Han Chinese: A Companion to Grammata Serica Recensia* (Honolulu: University of Hawai’i Press, 2009), 169.

18. Shanghai bowuguan cang *Chu zhu shu* (san), 137.


21. For a study of these chapters, done well before the discovery of the manuscripts that will be the topic of this discussion, see Jeffrey Kenneth Riegel, *"The Four ’Ts’ Ssu’ Chapters of the Li Chi: An Analysis and Translation of the Fang Chi, Chung Yung, Piao Chi, and Tsu*” (Ph.D. diss., Stanford University, 1978).


23. For a fuller discussion of this passage, together with citations of eighteen Chinese and Japanese studies that had been published prior to its publication, see Edward L. Shaughnessy, *Rewriting Early Chinese Texts* (Albany, NY: SUNY Press, 2006), 70–77. For a Chinese translation of this study, see Xia Hanyi 夏含易, *Chongqie Zhangguo gudai wenxian zhongwen chu jian* 郭店楚簡古代文獻 (Shanghai: Zhongguo guji chubanshe, 2012), 64–69.

24. Guodian *Chu mu zhujuan*, p. 131 n. 4.

25. For this word family, see Bernhard Karlgren, *Grammata Serica Recensia* (1957; rpt. Kungsbacka, Sweden: Elanders, 1972), #427.

26. This interpretation was first suggested to me by Li Ling 李零, in *Shi jing tong* 三代詩三家義集疏, *Shi jing* (Wuqiubei zhai Yi jing ji chûgoku tetsugaku, 映画, 1957; rpt. Kungsbacka, Sweden: Elendars, 1972), #427.

27. Early glosses for the word in this context are 32. For an excellent discussion of the sexual overtones of *yaojiao* 老僕, especially with its phonetically similar binomes *yaotiao* 老俏, *youshou* 窮俏, and *yaoshao* 窮遜, see Wang Xianqian 王先謙, *Shi san jiao yi jishu* 《詩三家義朱注》 (rpt. Beijing: Zhonghua shuju, 1987), 71–72.


31. On the other hand, to my mind all of his liberties are redeemed when the *Zhou Yi* was being composed, the words of these word families were often written with a single character.


33. Early glosses for the word in this context are *tong* 通 “to penetrate” and *jiao* 交 “to conjoint”; see Wang Xianqian 王先謙, *Shi san jiao yi jishu* 《詩三家義朱注》 (rpt. Beijing: Zhonghua shuju, 1987), 71–72.
4. Without wishing to anticipate the discussion below, I might note that both of these words were in ancient China standards euphemisms for sexual congress.

34. As for most Sinologist’s choice, Waley’s “Fair fair,” I suspect that only Waley knew whence it came, and he is no longer available to ask.


36. For the cry of the osprey, see https://www.youtube.com/watch?v=_KTpxps7mrw. The only discussion of bird calls that I have seen in the Chinese commentarial literature is that of Zheng Qiaoyang 鄭秋陽 (1108–1166) in the Tong zhizhuan tongzi. In all species of geese and ducks, since their beaks are flat their sound is guan-guan; in species of chickens and pheasants, since their beaks are pointed, their sound is yao-yao; these are natural sounds. The beak of the osprey resembles that of ducks and geese, therefore its sound is like this, also getting the sense of the water’s edge**: quoted in Shi jing cidian, ed. Xiang Xi (Chengdu: Sichuan Renmin chubanshe, 1986), 144. This is nonsensical for two reasons. First, geese and ducks are quoted in the Shi jing as making both sounds like guan-quan (yong-yong 橋鳴) and also yao-yao (ao-ao 敲敲). Second, the beak of the osprey is, in fact, pointed, and not flat like that of geese and ducks.

37. By analogy, in Edgard Allen Poe’s poem “The Raven,” the poet heard the raven to be saying “Lenore” and “Nevermore.” If we were not to hear this, we would fail to understand the psyche of the poet.

38. See Wang Xianqian, Shi Sanjia yi jishu 史三甲術, 4.


Introducing Premodern Text Translation: A New Field at the Crossroads of Sinology and Translation Studies

Carl Gene Fordham
PEKING UNIVERSITY 1366974083@QQ.COM

Over the past two decades, China has witnessed the rapid development of premodern text translation as a new, interdisciplinary field situated somewhere between traditional Chinese philology and translation studies, but also taking in elements of comparative literature, cultural studies, and other related disciplines. A plethora of books and journal articles have been published, a biennial conference has been held since 2002, and a number of universities have even started to offer doctorates in the specialization. Despite this, premodern text translation has not become a discrete field in Western academia, which is somewhat ironic considering that the vast majority of high quality, premodern text translations have hitherto been produced by scholars outside of China. And yet, unbeknownst to sinologists and translation scholars in the West, this field has the potential to explore many new and exciting questions that cannot be adequately addressed by sinology or translation studies as singular disciplines. Thus, this paper does not attempt to answer questions, but ask them instead. This may seem contrary to the typical way an academic article is written, but the fact is many of these questions have not been asked before in Western academia; in many cases, there are no answers available yet, only unexplored terrain for future scholars. This paper is the first time an English-language name for this discipline “premodern text translation” has even been put into print.

In preparation for this paper, I undertook two tasks. One was to review the major research produced in the field of premodern text translation, with the intention of providing a snapshot of the development of this interdisciplinary field to Western academics who would not otherwise have known of its existence. Two was to survey as many scholars who have published premodern text translations into English as possible, with the aim of understanding their attitude towards the development of this interdisciplinary. I wished to identify challenges they themselves faced during their translations of premodern Chinese texts, and determine how approaches adopted by interdisciplinary researchers could be used to better understand those challenges, and ideally formulate some relevant solutions.

Many of the most prolific translators of premodern texts are in their twilight years and difficult to contact, or have sadly passed away in recent years. I was fortunate, however, that over a dozen professors were able to take time out of their busy schedules to respond to questions I believe are relevant to the status of premodern text translation as a field in its own right. Some were even so helpful as to direct me to book chapters and articles discussing relevant topics. Based on their responses, I have summarized four aspects that I believe are most relevant for the introduction of this new field to Western academia, namely, the name and nature of premodern text translation, translator competence, preparatory work, and translation processes and methods.

THE NAME AND NATURE OF PREMODERN TEXT TRANSLATION

“Premodern text translation” is the tentative translation I offer for the Chinese term dianji fanyi 典籍翻譯. Although clunky, the advantage of “premodern text translation” over the rendering “classics translation” that is commonly used in China is that it acknowledges that the word dianji 典籍 denotes more than just the classic texts of the pre-Qin period—it includes virtually any text formed and circulated in China before the New Culture Movement of the mid 1910s and 1920s, and even encompasses texts written in non-Sinitic languages. Dianji is somewhat synonymous with gudian wenxian 古典文獻 (“ancient documents”), since the Chinese language does not clearly delineate between “ancient” and “imperial” the way English-speaking historians traditionally have in discourse on Chinese history. I welcome any other suggestions on how dianji fanyi might be translated into English; however, I have yet to come across a better rendering.

To give the reader some sense of the popularity and scope of premodern text translation as a research field in China, allow me to direct your attention to a book published in May 2019 entitled Selected Collected Commentaries on English Translations of the ‘Xue’er’ Chapter of the Analects, with the
First Chapter of James Legge’s ‘The Chinese Classics’ as a Base Text. Evidently, the field has matured to a point where translation critique of not only a single text but a single chapter can be treated to in-depth analysis and is able to attract funding for publication. The publication of this work was made possible by nearly two decades of hard work on the part of scholars in China to develop premodern text translation into a field in its own right. In the interests of space, it is not possible to enumerate every single piece of research that has been produced during that period, or even provide a definitive list of the most influential works. Rather, my aim is to introduce some of the key topics that have attracted the interest of scholars in China who do work in premodern text translation, and provide the relevant bibliographic information for those interested in reading more, and perhaps picking up where they left off.

To begin, let us take a look at some important works that provide historical background that informs our examination of premodern text translation. There are three scholars who have done some remarkable work in this area: David E. Mungello, David Honey, and Ming Dong Gu. Each of them has brought entirely different, but equally useful, dimensions to the table. Mungello has published extensively on the very first encounters between China and the West, and provided a much needed context; Honey has written Incense at the Altar, an extensive yet highly accessible introduction to the development of sinology and sinologists; while Gu’s Sinologism: An Alternative to Orientalism and Postcolonialism attempts to identify key historical trends at the macro level, and integrate them into a theoretical framework in which translation plays a key role. For a varied discussion on how early sinologists translated Chinese works into foreign languages, Sinologists as Translators in the Seventeenth to Nineteenth Centuries provides detailed examinations of specific case studies. A number of books have been published in China which catalogue the history of foreign translations of Chinese literature. While there is no room to list them all here, Hanji Waiyishi 漢籍外譯史 is a good starting point as it covers a wide range of texts, periods, languages, and regions. Likewise, many works on missionary-translators have been published by Chinese scholars. Lastly, it is worth pointing out that most of the well-known premodern text translators have had biographies written about them, many of which provide insight into how they approached the task of translation. By reading about the life of, for example, James Legge, and then learning about that of Bernhard Karlgren or Ezra Pound, it is clear that premodern texts from China have been understood and rendered in a myriad of ways.

As regards the actual practice of premodern text translation, it has been examined to some degree by a handful of Western scholar-translators. However, it is worth mentioning that these examinations have been mostly restricted to the odd book chapter here and there. As insightful as they are, there has not yet been a full-length book treatment on the subject by a premodern text translator in Western academia. By contrast, the practice of premodern text translation has been analyzed at length by scholars in China. Notably, this interest in research has been accompanied by the publication of various textbooks aimed at training students in premodern text translation at the graduate level, for example, Zhongguo Dianji Yingyi 中國典籍英譯 by Wang Rongpei 汪榕培 and Wang Hong 王宏.

The sophistication of research in this field has also led to the development of some theoretical frameworks which attempt to ascribe overarching trends, principles, and ideologies to premodern text translation, often with a specific focus on Chinese poetry. It has also allowed for the publication of a large number of books dedicated to examining a particular translated edition of a premodern text, such as the full-length book treatment in Chinese by Wang Fanglu 王方路 of the Shijing 詩經, or by Wang Hui 王輝 in Chinese of the Zhongyang 中庸, to name a few. Much research has been done on the transmission and influence of specific translated editions of premodern texts, both from a bird’s eye view, and from the perspective of a particular ideology, such as Confucianism or Zhu Xi thought. Lastly, there is no lack of publications which explore a special topic related to premodern text translation that does not fall into the aforementioned categories. One of the most prevalent of those topics is how premodern text translation fits with the strategic aims of government and non-government organizations in China, for example, the Going Out policy.

However, the fundamental question of whether premodern text translation is essentially different from other kinds of translation—for example, literary translation, or translation of contemporary texts—has been left unexplored in Western academia. This question may not be as simple as it seems; let us consider the layers of complexity at work here.

To ask someone to translate a text written before the advent of modernity, or written in medieval or ancient times, is to ask him or her to render something written not only in a foreign language but in a foreign time. Translating modern Chinese into modern English is already an onerous task involving the transfer of the lexicon, grammar, and shared assumptions of two languages with almost no shared history or culture. And, yet, premodern text translation takes the task one step further: it asks us to translate texts written centuries, or even millennia, apart.

I note that The Stanford Encyclopedia of Philosophy (Fall 2019 Edition) dedicated an entire entry to the subject of “Translating and Interpreting Chinese Philosophy.” It was written by the late Prof. Henry Rosemont Jr. who researched and translated widely on Chinese philosophy and religion. In the entry, Rosemont explains his rationale for singling out the translation of Chinese philosophy as a topic worthy of discrete investigation and inclusion in the encyclopedia:

There are no entries under “Translating and Interpreting...” for Greek Philosophy, for example, nor German Idealism or French Postmodernism. Even Indian philosophy lacks such an entry. This fact should bring home not just the singularity of the classical Chinese written language, compared to contemporary languages derived from the proto-Sanskrit Indo-European family of languages written in alphabetic scripts, but remind us as well of the greater distance between Chinese and other cultures from past to present.
This is where we are reminded of the truly global nature of this new field. Although this paper, for the sake of brevity, does not discuss premodern text translation of texts written in languages other than Chinese but formed and circulated in China such as ethnic minority literature, or indeed the languages of antiquity outside China, such as Ancient Greek or Latin, similar issues would presumably be encountered during these translation processes, and comparative research looking at these different premodern languages and the challenges they impose on the modern translator would be highly illuminating. In this way, premodern text translation has the potential to grow from a Chinese-focused field to an international one. Perhaps an eventual corollary would be the inclusion of the Chinese and other overlooked ancient civilizations in the discipline of classical studies, so that, for example, undergraduates would learn about Confucius and Laozi alongside Plato and Aristotle.

However, the consensus I discovered among the Western scholar-translators that I surveyed was that, while premodern texts have features which make them very different from modern texts, there is nothing about premodern text translation per se that sets it apart from other kinds of translation. Thus, when pressed to decide whether there were any intrinsic problems to premodern text translation, the professors for the most part expressed a concern for the needs and expectations of the audience, whether they be an academic or general one. Admittedly, this is a concern shared by almost any translator of almost any text.

Prof. Eric L. Hutton of the University of Utah, who among a myriad of academic achievements produced the translation Xunzi: the Complete Text (Princeton University Press, 2016), explained this point most eloquently by way of a “model” metaphor which may be applied equally to both premodern text translation and translation of other kinds of texts. Hutton approaches translation like an engineer would produce a model for the purposes of study and investigation, for example, for an aircraft or automobile. Given the complexity of most models, different engineers would add or omit features in their model depending on the aim of their investigation. Likewise, when rendering a text from Chinese into English, it is not possible for the translator to reproduce every possible feature of the source text in English. Instead, the translator is forced to select a limited number of the features of the source text to include in his or her “model” depending on the desired effect of the translation and its intended audience. This general principle applies equally to the translation of modern texts as it does to premodern texts, since the translator has objectives and audiences in both instances.21

Having read the numerous articles written by other premodern text translators, and studied the feedback I have received via email from many others, I believe this is a common approach by scholar-translators in Western academia. However, I also discovered some challenges posed by premodern text translation that do not seem to be shared by modern text translation. Here I am aided by Prof. Michael Nylan from the University of California, Berkeley, whose extensive contributions include a complete translation of the Fayan 法言 (Exemplary Figures / Yuejueshu 越絕書 (The Glory of Yue: Brill, 2010)). Among other practical aspects, Nylan has written of the “scissors-and-paste” style in which premodern texts were composed and how it influences the translation process:

Translators must be conscious of early conventions of writing in the period they are working on. Good writing, over and over again in the sources, is described by one of two synonymous binomial phrases: zhuwen 属文: putting together related passages and zhuiwen 綴文: compiling a text from preexisting units. . . . Compositions on a given theme are strung together from other passages (not necessarily devoted to the same topic) that contain the same vocabulary items or, more rarely, the same grammatical patterns. The repetition of particles conveyed to the reader/listener similarities in story lines, conclusions, and so forth. And since rhetoric often posited regularities, and parallelism by its very nature implies such regularities, without asking for a precise formulation liable to proof, parallel passages are ubiquitous.22

Nylan goes on to point out that modern readers of Classical Chinese often look for logical connections where they are none. However, the transmitters of premodern Chinese texts did not consider logic important; rather, texts were designed to “establish the bona fides of its compilers by the display and repetition of well-worn tropes of high cultural literacy, memorized sayings, and some snippets of authoritative texts or proverbial wisdom, after which the compilers were free to try to formulate their most persuasive arguments.”23

Another aspect peculiar to premodern text translation is it spans multiple genres and times. Taking in the available literature at a bird’s-eye view, it is evident that different genres present entirely different concerns for the translator. Take, for example, Classical Chinese poetry. Prof. Wilt L. Idema of Harvard University has spoken at length of its specificity: the way translators through the ages have grappled with preserving (or choosing not to preserve) rhyming schemes, metrical equivalence, and parallelism; the simplicity and directness of Classical Chinese poetry which is at odds with the highly rhetorical language of its counterpart in Western antiquity; the problem of annotations in poetry translation which read like jokes that have to be explained, and so on.24 And yet this is but one genre among the many in premodern Chinese literature: What of pre-Qin philosophical works, the fu 赋 rhapsodies of the Han dynasty, Yuan dynasty plays, or the famous novels of the Ming and Qing dynasties? These all come under the umbrella of premodern texts, and yet they each present their own distinctive challenges for the translator, and thus are deserving of separate investigations into topics such as translation process and method.

Here I wish to share the input of Olivia Milburn, professor of Chinese Seoul National University and translator of the ancient Chinese text Yuejueshu 越絕書 (The Glory of Yue: Brill, 2010). Milburn rightly points out that a key difference between translating modern and premodern texts is that, when it comes to the former, the author and his or
her contemporaries are usually available for consultation and clarification when problems arise. However, this is impossible for the latter.25 Prof. E. Bruce Brooks from the University of Massachusetts Amherst and translator of The Original Analects (with A. Taeko Brooks: Columbia University Press, 1998) expressed a similar sentiment, describing premodern texts as anything but “authorial” in nature, which makes examination of them through the lens of translation theory incredibly problematic. Additionally, unlike most modern texts, premodern texts are products of long and continuous processes of adaptation. Thus, in Brooks’s view, to represent the premodern text as coherent and consistent is to lie about the text.26

Milburn was also generous enough to elaborate on some of the specific challenges encountered by translators of pre-Qin texts, although the discussion could, to some extent, apply to premodern texts more broadly. Milburn explained that pre-Qin texts will at some point have been translated at least once into standard Qin-Han characters; however, due to problems such as manuscript errors, omissions, censorship, and any number of other issues, it is very difficult to determine the accuracy of the translation. In some cases, the received text accessible to the modern reader has become so garbled as to be incomprehensible. Again, I would posit that issues of this nature pose challenges for the translator that are unique to this field.

Lastly, a common concern among the professors surveyed was the importance of philology when translating premodern Chinese texts. Professor Emeritus of Chinese literature at the University of Washington David R. Knechtges, for instance, emphasized the point in an article he wrote which deconstructed his ongoing translation of the Wenxuan 文選 into English (Wen Xuan or Selections of Refined Literature in various volumes: Princeton University Press from 1982 onwards), concluding: “If we do not properly understand the language of our texts, we cannot hope to translate them correctly.”27 Yet, in the view of Nylan, however much translation depends on philological work, translators must never equate translation work with philological work:

For the academic philologist preoccupied with lexicography and grammatical analysis, the single goal is clarity of meaning, but clarity often undermines the preservation of the most interesting features to be found in a historical or philosophical persuasion piece: the precise way that the text sets about to “intimate its meanings” in a compelling fashion. What damage a philologist equipped with a tin ear for English can do to a translation can be readily appreciated by reference to Bernhard Karlsgren’s truly awful (i.e., incomprehensible and aesthetically offensive) translations of several of the Chinese classics. Widely admired and emulated prose in Han, pre-Han, and the immediate post-Han period conveyed the sheer delight some of the author-compilers took in seeding their texts with intertextual allusions, larding them with enigmas and ambiguities, and devising dizzying verbal copia and reduplicative binomes of redoubtable hypnotic power but hazy meaning. Translators should aim to reproduce some measure of that same marvelous and magnetic richness, even if it seems sometimes to come perilously close to an over-the-top excess to our tastes today.28

TRANSLATOR COMPETENCE
To many scholar-translators in the West, the question of translator competence may seem irrelevant to them. Perhaps, in their minds, a premodern text translation is produced by extensive and intensive reading of the text and its related sources, then translated into English as faithfully as possible, and that is that. Thus, they would argue, to investigate the question any further would be mere theorization. However, in the context of China today, this answer is probably inadequate. The reality is that premodern text translation is already a mainstay of translation programs at both the undergraduate and graduate level in China. Over the past two decades during which this interdisciplinary field has developed, universities throughout China have been training premodern text translators in the hope that China may, one day, produce its own high quality translations of the literature it rightly prizes as part of its intangible cultural heritage. And, regardless, what is the harm in examining more closely what is needed to produce better quality translations? Surely the better informed we are about how competence is formed in translator education, the more time can be saved in preparing students for this mammoth task.

On this particular question, I believe the scholar-translators from the West can find agreement with their counterparts in China; during my correspondence with them, I noted their shared emphasis on the importance of expertise in Classical and Literary Chinese, in particular a competence that goes beyond literate second-language acquisition. According to Nylan, not only are premodern text translators equipped with a level of premodern Chinese that few command, but perhaps equally important, they have an in-depth understanding of historical context without which a nuanced translation is not possible. Without these competencies, it seems true comprehension of the source text cannot be guaranteed, let alone an adequate translation.29

However, the difficulty of attaining literacy in classical and literary Chinese should not be underestimated. Here I am aided by Prof. Robert G. Henricks from Dartmouth College, who contributed to English speakers’ understanding of ancient Chinese philosophy with a new translation of the Daodejing 道德経 based on the Mawangdui Silk Texts (Lao-tzu: Te-tao ching: A New Translation Based on the Recently Discovered Ma-wang-tui Texts, Ballantine, 1992). In Henricks’s view, what makes classical and literary Chinese so difficult is its incredible terseness; much meaning is expressed in very few words, and even those words used may be polysemous or even imprecise. Additionally, grammatical structures may not define syntactical relationships with any kind of precision, and authors often do not even use the particles at their disposal.30

One should not forget, either, the issue of linguistic ability in English. In an address at Guangdong University of Foreign Studies on June 19, 2017, Prof. Idema set forth the common assumption that “literary translations are best produced by highly-educated native speakers of the target language with an intimate knowledge of the source language.”31 Idema notes that, despite occasional exceptions such as the prolific Chinese scholar and translator Lin Yutang
Ming Dong Gu, a professor of Chinese and comparative literature at the University of Texas at Dallas, goes one step further by suggesting that the ideal translator of premodern texts runs a gamut of roles during his or her practice:

Translation is not simply an act of rendering a source text into a target language; it is a complex hermeneutic act with the aim of producing a performative continuum in which the translator assumes the multiple roles of reader, scholar, critic, thinker, and writer. The outcome of such a hermeneutic act is a multiple textual spectrum with readerly translation at one pole and writerly translation at the other. . . . In the final analysis, the ideal translator is not merely a competent reader who has a mastery of both source and target languages or a sensitive reader who is able to discover hidden connections in a source text. He or she should be a well-trained scholar who has intimate knowledge of source-text culture and target-text culture; a discerning critic who possesses a high literary sensitivity and can tell the strengths and weaknesses of a translated text; a practical thinker who can apply insights derived from reading, scholarship, and criticism to translation; and a creative writer who is worthy of being ranked among first-rate authors.14

Prof. Gu touches upon something here that I believe to be central to premodern text translation and how it may enlighten us as to the core competencies of the budding premodern text translator, which in turn has implications for how we might train literary translators at the tertiary level who specialize in premodern texts. Should a premodern text translator be merely a competent bilingual, the resources available at the department of Chinese studies or translation studies would surely suffice; and yet considering the complex, multifaceted nature of the task at hand, these disciplines are not adequate in and of themselves. In light of this, it is necessary to consider an interdisciplinary approach.

In my interactions with the professors, I also asked whether they considered expertise in Western philosophy, or a classics education, to be a prerequisite for someone to translate a premodern Chinese text into English. The responses for this question were the most divided out of all the other questions. For that reason, I believe more research could, and indeed should, be done on whether this particular competence is essential for the aspiring premodern text translator.

As it happens, there has already been much written on the topic of comparative philosophy more broadly, and its implications for premodern text translation. Undoubtedly, the most vocal supporter of philosophical understandings would have to be Prof. Roger T. Ames of Peking University and the University of Hawai‘i at Mānoa whose translations include the Dao de jing 道德経, Lunyu 論語, Zhongyong 中庸, Xiaoqing 孝經, and Sunzi Bingfa 孫子兵法, with some being published with the subtitle “A Philosophical Translation.” There is little space in this paper to provide a detailed overview of Ames’s significant contributions to this field;
those interested in reading further need only consult the explanatory texts contained in any of these translated editions.

For the other scholars I consulted such as Hutton, a background in Western philosophy is not an absolute prerequisite for the premodern text translator, though he acknowledges that some familiarity with Western philosophy and classics is helpful. In his view, since various English terms have a history of being used in connection with particular ideas, it is important to develop a sensitivity to the ways in which rendering Chinese terms one way or another brings them closer to or further away from the “baggage” that various English terms carry. For example, the notion of an agent’s “will” is particular to the Western philosophical and religious tradition, and to render the Chinese zhi 志 as “will” without careful consideration is risky since it is a loaded concept that carries with it certain philosophical and religious associations.\(^\text{16}\) However, Hutton concludes that the required sensitivity can be gained through general education and experience, and that it is not necessary to become an expert in Western philosophy and classics to become a successful premodern text translator.

For Henricks, philosophical training may only aid in the translation of some premodern texts, for example, Neo-Confucian writings. He provided the example of Angus Graham’s translations of some of the early texts using logic as a good example of that expertise in action. Nonetheless, Henricks emphasized that a philosophical education is not necessary for many of the early Chinese classics such as the Laozi, Zhuangzi, Lunyu, and Mengzi. In his view, there is no need to make these translations overly abstract because these texts are not inherently philosophical.\(^\text{17}\) Other professors acknowledge the usefulness of philosophical expertise, but deny its value as a prerequisite. For Prof. Edward L. Shaughnessy of the University of Chicago, for example, expertise in the variety of Chinese being translated—whether it be classical, literary, or vernacular—is more important.\(^\text{18}\)

One topic that merits further investigation is whether one’s understanding of Western philosophy—and the preconceived notions that inform part of that understanding—may actually prevent an accurate understanding of the Chinese source text. I am grateful to have received input on this point from Prof. Paul R. Goldin of the University of Pennsylvania, who published a translation of the Xiaojing 孝經 in 2005 (in *Hawaii Reader in Traditional Chinese Culture*, University of Hawaii Press), and will publish the first complete translation into English of Xinyu 新語 this year with Elisa Levi Sabattini entitled *Lu Jia’s New Discourses: A Political Manifesto from the Early Han Dynasty*. Goldin refers to this kind of accuracy as an “emic understanding,” which I believe to be an appropriate term for this phenomenon, as it denotes an understanding of the text on its own terms, rather than imposing systems of thought that have originated outside of China onto the text in question. There have been articles published on particular translators’ perceived biases, but I am unaware of any systematic research on emic versus etic understandings of premodern texts and their translations into English.

PREPARATORY WORK

In my view, an important part of premodern text translation should be considering the preparation required when taking on a translation of a premodern Chinese text. As I expected, the scholars I surveyed for this paper provided particularly detailed explanations on this subject. For all of them, consultation of documentary sources and traditional commentaries was viewed as an essential part of the preparatory work required for the translation of a premodern text, or at least a classic or canonical one. For Knechtges, even this is not enough:

Linguistic knowledge alone is not sufficient in most cases for a proper understanding of the meaning of a text. Although one may know the literal sense of all the words in a line and be able to explain their grammatical function and even to reconstruct their putative ancient or medieval pronunciation, such knowledge may not produce a correct translation. For example, in translating the *fu* on the capitals, which occupy the first six chün of the *Wen hsüan*, I spent much time reading historical and archaeological studies of Changan, Loyang, Chengtu, Chien-k’ang, and Yeh. I became intimately familiar with such works as *San-fu huang-t' u 三輔黃圖*, the reports on archaeological excavations, and studies of early Chinese architecture. The capital of *fu* of Chang Heng and Pan Ku are replete with accounts of ritual, and it became necessary for me to acquire a thorough knowledge of the ritual classics *Li chi 禮記* and *Chou li 周禮* as well as the monographs in the *Hou Han shu* on ceremony and official dress.\(^\text{19}\)

Prof. Goldin was also generous in providing his perspective on preparatory work. In his view, translations into Vernacular Chinese and foreign languages may be consulted during the preparatory process as long as the translator is aware of the potential influence they may exert on comprehension and expression. Notably, this kind of preparation would probably not be necessary if one were translating modern, European texts, for example a novel by Honoré de Balzac.\(^\text{20}\) Hutton went one step further, arguing that it would be irresponsible or even disrespectful to proceed with a translation without consulting the previous scholarship of East Asian scholars, regardless of whether one agrees or not with how those scholars interpreted the text being translated.\(^\text{21}\)

From this topic, we can consider the question of vernacular Chinese translation: Should premodern text translators consult one,\(^\text{22}\) and if so, what effects does that vernacular translation have on the process and product of translation, either beneficial or detrimental? Idema advises premodern text translators to base their rendition on the source text and not on a translation into contemporary Chinese as was the case with many texts included in the *Library of Chinese Classics*.\(^\text{23}\) Again, I am interested in how this hypothesis can be proven true, rather than assuming it to be a fact, and this is where premodern text translation can assist us with its interdisciplinary approach.
Some professors emphasized the importance of understanding the intellectual context of the source text before beginning translation. Prof. Shaughnessy particularly emphasized this point. Shaughnessy gave the example of the Yi Jing (易經), a text that he himself has produced two translations for (I Ching: the Classic of Changes. Ballantine, 1998; Unearthing the Changes: Recently Discovered Manuscripts of the Yi Jing (I Ching) and Related Texts, Columbia University Press, 2014). Since it took almost a millennium for the Yi Jing to take form as a text, and then another two millennia for it to be used and interpreted by different people at different times, it has been understood in radically different ways. Since it is not possible to produce a synoptic translation that would account for all of these different interpretations, Shaughnessy concludes that one has to decide in advance which time period and interpretation to adopt, and then strive to remain true to this context.44

The importance of making a glossary of key terms and their translations was also emphasized by many of the professors. Hutton, for example, pointed out that large projects are usually worked on continuously over long periods of time, meaning that translators are liable to forget translation choices made earlier on, resulting in inconsistent translations of the key terms, with the translator often unaware of the issue until much later on in the project. Hutton even laments that if he had made a list as he had progressed through the project it would have been easier to maintain consistency throughout the translation, rather than discovering inconsistencies at the end and then having to work back through the whole manuscript to try to address them.45

Some professors emphasized the importance of investigating the linguistic background of key terms, especially high frequency ones. In Goldin’s view, many translators fail to do this, largely because to do so requires an ability that cannot be acquired without linguistic training. As a result, inadequate stock translations often persist in premodern text translations due in part to a lack of critical examination.46 This is, perhaps, a whole topic in itself. It would be illuminating to investigate what linguistic tools and procedures may assist the premodern text translator, especially those that are often neglected by scholars whose primary specialization is, for example, philology, historiography, or poetics.

Prof. Milburn recommended running key words and phrases from the source text in major databases for Chinese literature. This kind of cross-referencing not only ensures the accuracy of the translation, but also determines whether parts are quoted in later sources. These parts, which provide similar content but in a different form, may then assist the translator with understanding problematic aspects of the text being translated.

Once these concerns at the word level are addressed, the professors were quick to stress that there is still a range of questions that needs to be resolved in terms of the purpose of the translation before getting started on the translation itself. In my communication, I identified some purposes that premodern text translators commonly strive for. I am indebted to Prof. Idema for assistance in brainstorming these:47

- A new translation that would update the antiquated style of previous translation(s), for example, those published by James Legge over one hundred years ago.
- A new translation that would correct perceived mistakes in previous translation(s), for example, those done by scholars without philological training.
- A new translation that would incorporate new sources or materials, for example, excavated texts, commentaries, or recent scholarly discussion.
- A new translation based on a different edition that has been overlooked by previous translators, for example, one that has only recently been made available.
- A new translation for an audience that has been neglected by previous translators, for example, a non-specialist one.
- A new translation that brings out the literary qualities of the text, where others have focused on philological concerns, or vice versa.

TRANSLATION PROCESSES AND METHODS

Lastly, I asked the scholar-translators to elaborate on the processes and methods they used in their work on premodern texts. I believe that the following view of Prof. Jonathan Chaves of the George Washington University, an accomplished translator of classical Chinese poetry, may be taken as representative of the typical attitude towards the process of translation by premodern text translators in Western academia. Chaves identifies with the description of Buddhist sutra translation provided by the Song poet Mei Yaochen 梅堯臣 (1002–1060) who views translation as a ritualistic activity. In the view of Chaves, the inner workings of this activity cannot be expressed in words. He writes:

*The fact is that the essence of the [translation] process has always, somewhat paradoxically perhaps, eluded verbal expression. Given a solid knowledge of the language in question, the literary translator should not go about his task with a preconceived theory of how to translate. After the job is done, he can, of course, analyze what has happened to his heart’s content. But while engaged in what must be at least partially a creative endeavor, he is best advised to emulate the monks of Mei Yao-ch’en’s poem and enter upon a ceremonial relationship with the original writer, to perceive as the poet has perceived, to experience what he has experienced.*48

This attitude towards translation could not be more strikingly different than that of translation studies researchers who, while admitting the difficulties involved in determining how exactly translators translate, would not claim the question unanswerable; indeed, the premise of their discipline more or less relies on the very proposition that at least some answers to this question may be revealed to the researcher. It is one of the key questions that signals the end of sinology and the beginning of translation studies—or, in our case, the beginning of the interdisciplinary field of premodern text translation.
I asked the professors to consider the translation process and methods, not only because they are two aspects that concern the daily work of every translator, but also because they are commonly neglected in translation studies or academia at large due to a common focus on the product, i.e., the target text. For those professors who believed the question answerable, they expressed a preference for faithful translation as a minimum standard, and idiomaticity and readability as secondary considerations. This is probably not surprising given that their translations are mostly for scholarly readers.

However, a common theme that I was able to identify in the professors’ feedback was a shared concern for the needs of the reader. For Idema, these needs may be addressed by providing substantial introductions to place the source text in its own culture, adding footnotes where unavoidable, or even including illustrations as visual aids to the reader. For Hutton, one way the needs of his readers can be addressed is by sharing his translation drafts with his undergraduate students to find places where they misunderstand what he is saying in his translation, sometimes in unusual ways he had not anticipated. The professor reported that the questions his students raised were helpful in identifying places where the translation was unclear or difficult to read.

Nylan has also written at length on aiding the reader in “decluttering” the main text of the translation by avoiding parenthetical remarks in the main text:

To this same end, translators should consider using separate glossaries or appendices that list (a) persons mentioned in the text . . . (b) official titles . . . (c) all place names . . . (d) key terms that appear in the main text . . . (e) book titles mentioned in the text. . . . The provision of such appendices allows both translators and readers to focus upon the English in the main text, which is surely the proper focus of their attention, while engorging or disgorging as many of the details as they choose.49

Nylan also urges premodern text translators to provide introductions to their translations which include a biography of the author, brief history of the text, review of key terms used in the text, and comparisons of the work to masterworks widely available in English, which she believes allows readers to “immerse themselves in the world known to the original author.”50

For scholar-translators like Prof. William H. Nienhauser, Jr., Halls-Bascom Professor of Classical Chinese Literature at the University of Wisconsin–Madison, attending to the needs of the reader should ideally be in line with the implied reader of the source text. Nienhauser describes how this principle applies to the translation series of Shi ji 史記 of which he has been chief editor:

There may have been many things which struck the reader about this translation (neologisms such as “Grand Duke” and “Nobile Scion”, for example). However, I hope that it has become clear through the sample translation that the Shi ji is not only an integrated work in the sense that the various sections—annals, tables, treatises, hereditary houses, and biographies—are meant to be read and cross-referenced, but that Ssu-ma Ch’ien has also considered it an integral part of the other source materials then available on this subject. It may not be true that every Han reader could fall back on knowledge of every event or narrative when Ssu-ma Ch’ien demanded, but it is apparent that Ssu-ma Ch’ien expected his ideal reader to be able to do so. For this reason, it is vital to provide an annotative context for all translations intended for the modern scholarly reader. Our modern reader has, of course, the option of referring to or neglecting the relevant context given in our apparatus. While we recognize the tremendous loss in style and flavor the Shi ji suffers in any translation, we hope in this way at least to retain some of the spirit of Ssu-ma Ch’ien’s historiography in our rendition.51

Notably, while Nienhauser’s approach to premodern text translation has been adopted by many of the other scholar-translators, it is also at odds with translators like Burton Watson who became well-known for producing translations that sacrificed philological accuracy for fluency and reader accessibility.52

PROSPECTS FOR PREMODERN TEXT TRANSLATION AS AN INTERDISCIPLINE

Interdisciplines are not created in a vacuum; they are born out of necessity. Where there are important questions in a given field that cannot be answered by a singular discipline, new tools for analysis are made to meet the challenge, eventually resulting in the formation of a new body of knowledge. At the heart of every interdisciplinary is not just intellectual curiosity, but an overt challenge to specialization.

In the case of premodern text translation, scholars are neither interested in the overly technical aspects of traditional Chinese philology, nor a study of premodern China through translation that makes no acknowledgment of the bilingual transfer process. While some scholars in those respective fields may show disinterest, or even aversion towards this new discipline, as this paper has demonstrated, premodern text translation has evolved into a field in its own right in universities in China. Since China is one of the very few nations today that boasts a continuous literary tradition spanning from antiquity to modernity, it is understandable that scholars and leaders in China alike have supported the development of this field with the hope that more research of this nature may be undertaken on a global level.

Having witnessed the rapid development of this discipline, I anticipate that discovery and exploration by international academia is somewhat likely in the near future. Parallels can be drawn here between premodern text translation and digital humanities, and how the latter is applied specifically to Chinese studies in the field of digital sinology. As fields like these continue to develop, it is anticipated that more opportunities for cooperation and consultation will
emerge, especially between academicians and institutions in and outside of China. For example, projects which bring together scholars from China who specialize in a particular text with academics from English-speaking countries with a background in translation would be able to produce high quality translations by integrating skill sets from both disciplines that individual scholars rarely (if ever) possess in equal quantities. Such projects would result in not only more research in this area but, crucially, more high-quality translations of premodern texts. It is this potential for application that sets premodern text translation apart from its related fields in which published translations are viewed as vehicles for study, and not objects of study in their own right.

Thus, let us persuade academics in the West currently producing research in their respective disciplines to consider the benefits of working in this dynamic, interdisciplinary field and, should they do so, show to them what progress has already been made. As it stands, premodern text translators outside of China rarely engage in constructive dialogue with their counterparts in China, and vice versa. By facilitating exchange between them, it is hoped that more, better quality premodern text translations may be produced, which aligns with the interests of both parties, not to mention the loftier cause of making traditional Chinese culture more accessible to a global audience. It will also have the benefit of preventing one side from having to reinvent the wheel each time a new study or translation project is carried out.

NOTES
1. The first conference on premodern text translation was held in October 2002 at Hebei Normal University, Shijiazhuang. The last one was held in October 2019 at Xi'an University of Technology, Xi'an. The next one—the twelfth session—is planned to be held at Ludong University in Yantai, Shandong province.
4. See David Honey, Incense at the Altar: Pioneering Sinologists and Incense at the Altar: Pioneering Sinologists and Premodern Text Translation under the Protection of Missionaries in the Modern Period.
5. See Ming Dong Gu, “Readerly Translation and Writerly Translation: A History of Premodern Text Translation from Chinese into Other Languages.”
6. For an overview of the current state of research, as well as some in-depth analyses of influential translated texts, see Yue Feng 岳峰, Zai Shisu yu Zongjiao zhijian Zou Gangsi: Xi Jindai Chuanjiaoshi dui Rujia Jingdian de Fanyi yu Quanshi [At the Threshold of Religion: A History of Premodern Text Translation from Chinese into Other Languages].
10. See Wang Hongyin 王慧印, Xin Yixue Lungao 新興學論稿 [Collected Papers on New Translation Studies]; Li Yanxia 李燕霞, Qin Liyan 秦麗艷, Yuanying-Moxing Fanyi Lilun yu Dianji Yingyi Yanjiu 與因—模式翻譯理論與典籍譯譯研究 [The Prototype—Mode Theory of Translation and Research into Premodern Text Translation into English].
13. See Li Yuliang 李玉良, Luo Gongli 羅公利, Rujia Sixiang zai Xifang de Fanyi yu Chuanbo 儒家思想在西方的翻譯與傳播 [The Translation and Transmission of Confucian Thought in the West].
17. See Li Xiaoxia 李小霞, Wenhua “Zou Qu Fu” Yujing zhong de Fanyi yu Chuanbo 文化“走出去”語境中的翻譯與傳播 [Translation Studies in the Context of the Cultural Aims of the Going Out Policy].
20. Personal communication.
22. Personal communication.
23. Ibid., 122.
25. Personal communication.
26. Personal communication.
29. Personal communication.
30. Personal communication.
31. Wilt L. Idenia, “How to Make Our Selected Chinese Texts and Their Translations Accepted by Publishers and Readers in the West: A Sinologist’s Reflections.”
32. Ibid.
33. Personal communication.
35. Or sinology, East Asian studies, etc. depending on the institution in question and the way it organizes its disciplines.

36. Personal communication.

37. Personal communication.

38. Personal communication.

39. David R. Knechtges, “Problems of Translation: The Wen hsüan in English,” 47. Knechtges provides even more in-depth examples in the article.

40. Personal communication.

41. Personal communication.

42. Nowadays, there are Vernacular Chinese translations for virtually all premodern texts that are well-known in China, and they are regularly consulted by general and specialist audiences alike.

43. See Witt L. Idema, “How to Make Our Selected Chinese Texts and Their Translations Accepted by Publishers and Readers in the West: A Sinologist’s Reflections.” Provided by personal communication. The Library of Chinese Classics is a large-scale, multi-press translation project that has produced a large number of full-length translations of classic Chinese texts into various foreign languages.

44. Personal communication.

45. Personal communication.

46. Personal communication.

47. For an example of such a critique, see E. Bruce Brooks and A. Taeko Brooks, “The Unproblematic Confucius: Review of Burton Watson’s ‘The Analects of Confucius’.”

REFERENCES


SUBMISSION GUIDELINES AND INFORMATION

GOAL OF THE NEWSLETTER ON ASIAN AND ASIAN-AMERICAN PHILOSOPHERS

The APA Newsletter on Asian and Asian-American Philosophers and Philosophies is sponsored by the APA Committee on Asian and Asian-American Philosophers and Philosophies to report on the philosophical work of Asian and Asian-American philosophy, to report on new work in Asian philosophy, and to provide a forum for the discussion of topics of importance to Asian and Asian-American philosophers and those engaged with Asian and Asian-American philosophy. We encourage a diversity of views and topics within this broad rubric. None of the varied philosophical views provided by authors of newsletter articles necessarily represents the views of any or all the members of the Committee on Asian and Asian-American Philosophers and Philosophies, including the editor(s) of the newsletter. The committee and the newsletter are committed to advancing Asian and Asian-American philosophical scholarships and bringing this work and this community to the attention of the larger philosophical community; we do not endorse any particular approach to Asian or Asian-American philosophy.

SUBMISSION GUIDELINES

1) Purpose: The purpose of the newsletter is to publish information about the status of Asians and Asian Americans and their philosophy and to make the resources of Asians and Asian-American philosophy available to a larger philosophical community. The newsletter presents discussions of recent developments in Asian and Asian-American philosophy (including, for example, both modern and classical East-Asian philosophy, both modern and classical South Asian philosophy, and Asians and Asian Americans doing philosophy in its various forms), related work in other disciplines, literature overviews, reviews of the discipline as a whole, timely book reviews, and suggestions for both spreading and improving the teaching of Asian philosophy in the current curriculum. It also informs the profession about the work of the APA Committee on Asian and Asian-American Philosophers and Philosophies. One way the dissemination of knowledge of the relevant areas occurs is by holding highly visible, interactive sessions on Asian philosophy at the American Philosophical Association’s three annual divisional meetings. Potential authors should follow the submission guidelines below:

i) Please submit essays electronically to the editor(s). Articles submitted to the newsletter should be limited to ten double-spaced pages and must follow the APA submission guidelines.

ii) All manuscripts should be prepared for anonymous review. Each submission shall be sent to two referees. Reports will be shared with authors. References should follow The Chicago Manual Style.

iii) If the paper is accepted, each author is required to sign a copyright transfer form, available on the APA website, prior to publication.

2) Book reviews and reviewers: If you have published a book that you consider appropriate for review in the newsletter, please ask your publisher to send the editor(s) a copy of your book. Each call for papers may also include a list of books for possible review. To volunteer to review books (or some specific book), kindly send the editor(s) a CV and letter of interest mentioning your areas of research and teaching.

3) Where to send papers/reviews: Please send all articles, comments, reviews, suggestions, books, and other communications to the editor: Minh Nguyen (atnguyen@fgcu.edu).

4) Submission deadlines: Submissions for spring issues are due by the preceding November 1, and submissions for fall issues are due by the preceding February 1.

5) Guest editorship: It is possible that one or more members of the Committee on Asian and Asian American Philosophers and Philosophies could act as guest editors for one of the issues of the newsletter depending on their expertise in the field. To produce a high-quality newsletter, one of the co-editors could even come from outside the membership of the committee depending on his/her area of research interest.
INTRODUCTION: PARENTING AND PHILOSOPHY

Lauren Freeman
UNIVERSITY OF LOUISVILLE

Several years ago, I attended a small conference for women in philosophy. One week before the conference, I learned that I was pregnant with my second child. Given that I had an almost two-year-old, another on the way, and that I had just completed my first year in a tenure-track job, I was excited to see a panel on the program called (something like) “Philosophy and Motherhood.” I showed up enthusiastically, pen in hand, notebook splayed, eager to write down all of the tips that were going to be offered on how to manage and balance children and a promising career (not to mention everything else).

But I was not so excited to learn that the unanimous—and, quite frankly, surprising—conclusion from all of the participants on the panel was that it was more or less impossible to achieve the trifecta of success in philosophy, parenthood, and (even minimal) happiness. One panelist spoke about how she’d had children well before beginning graduate school and how she was more or less absent during their early childhood in order to finish her degree. As a result, she regretted it, either entirely, or for having them.

The conclusion I drew was that if you have children as an academic, you regret it, either entirely, or for having them before tenure. But even if you don’t come to regret your decision, it’s still impossible to succeed at research, have a family, and be happy. Needless to say, as someone who had one child, was newly pregnant, and was already nervous about bringing a second child into our family, I left the panel and the conference feeling downtrodden, disappointed, and fearful for what my future had in store for me and my partner.

But, as I’ve come to see six years later, what the panelists described on that day is not the only outcome of being a parent and a philosopher, nor should it be the only or the dominant narrative floating around. In retrospect, it was this panel, and the very negative picture it painted, that was the initial impetus for conceiving of the current issue.

So far, my own experience as a parent has been positive. I say this with full acknowledgment and awareness of all of my many privileges. I had my son (who is now almost eight) while I was just about to begin the third of a three-year limited-term teaching job in Montreal, Quebec, in Canada. Even though it was not a permanent job, because Quebec has the most generous parental leave policy of any province in Canada, I still received a full year of parental leave along with 100 percent of my salary (which, as far as I’m concerned, should be the norm everywhere). Because of this, we were able to move as a family to the city where my partner had secured a job for the year. It was during this year that I landed a tenure-track job after four years on the market. (Note: doing a series of more than five on-campus interviews while nursing, and leaving a six-month-old at home with his dad (who was working full time, during the winter of several “Snowpocalypses” on the east coast which caused many flight delays and some canceled flights which caused me to stay away from them for longer than planned) required lots of juggling, lots of paid childcare, lots of requested accommodations from search committees, and several embarrassing and awkward moments pumping milk during my campus visits. Though being on the job market while nursing an infant is not ideal, it is certainly possible and I am happy to report that all of the departments where I interviewed were as understanding and accommodating of all of my requests as they possibly could have been. This is a point that needs to be broadcasted loudly, since every department far exceeded my expectations.

When my daughter was born a year and a half into my tenure-track job, not surprisingly, things were not as easy parental-leave wise as they were in Canada, since I was then living in the United States and working at a large state school with no good parental leave policy of which to speak. But with tremendous and relentless effort and negotiation, I did manage to secure a leave with which I was completely happy. With the input and savviness of colleagues who had successfully navigated this terrain before and who served as invaluable mentors (thanks, Tracy K.!) and a remarkable amount of creative thinking on the part of my chair at the time (thanks, Bob K.!), I ended up tweaking my annual work plan by ramping up my research and service duties so that I did not have to teach for an entire semester. With this time off teaching, combined with the summer months...
following that semester, I did not go back to teaching until my daughter was eight months old. Another ingredient in the mix then and now is a family-friendly department and current chair (shout out to David O.), which makes an enormous difference when scheduling classes, department meetings, and other obligations.

Though we live in a city with no family anywhere close by, my parents and my partner’s parents are still very involved in the lives of our children and help us in whatever ways they can. But the geographic distance means that we have no family on whom we can rely or depend for extra or emergency childcare—for snow days, sick days, or children’s (countless) school breaks that do not align with our work calendars. This means that we’ve spent and continue to spend rather alarming amounts of money on extra childcare. But, as seems to be common in the lives of working parents in general, and of philosophers in particular—see the narratives in this issue by Samantha Brennan and Amy Allen—this is often a necessary tactic in order to get it all done (again, a huge privilege that we are able to afford this extra childcare—thanks to Kristen, Grace, Rachel, and Gabby for saving us more times than I can count!). Most of all, and in addition to the other privileges I’ve just mentioned, what has been a necessary ingredient in being able to juggle it all is that I have an incredibly devoted and involved partner (also a philosophy professor) who is also entirely committed to feminist ideals of child-raising and to an equal division of household and family responsibilities.

The topic of parenting and philosophy is not discussed enough in any formal way. As an antidote to the panel I mentioned above—which painted a rather grim picture of motherhood/parenthood and philosophy, to say the least—I conceived of this issue as a way of broadcasting a variety of diverse narratives of success. In the early stages of thinking about the issue, I made an interesting observation, namely, that when I approached several potential contributors, there was a real and explicit reluctance (a) to think of themselves as successes (though the people I have in mind are all full professors and/or deans, and all have three or more children) and (b) to tell their stories in a way that centers their success. In varying ways, they all feared, in the words of one of the philosophers who I asked to contribute, “coming across as an asshole.” The fact that narratives of success in parenting and philosophy are not the norm in the field, combined with the fact that people who are successful are hesitant and sometimes even ashamed to tell their stories, made me even more persistent in convincing them to contribute.

Some did, some did not.

Before giving a sneak preview of what you’ll find in this issue, a few caveats are in order. First, I intentionally did not want to call this issue “Motherhood and Philosophy.” Many issues surrounding children and careers focus on women, and there are good reasons for this—reasons that are rooted in systemic misogyny, sexism, patriarchy, and the myriad ways in which women have been and continue to be systematically discriminated against with regards to their reproductive choices (or, lack thereof), pregnancy, childbirth, raising children, and doubly so when these issues are combined with their choice to pursue careers. In this issue, I wanted to feature diverse philosophers who are mothers, but also fathers, and parents of all sexes and genders and to provide them with a venue in which to share their stories. Focusing on women can be helpful, but it can also eclipse the stories of other primary caregivers who do not identify as women. Focusing on parenting more generally acknowledges that trans and non-binary folks also face specific barriers in their parenting journeys. Saray Ayala-Lopez’s contribution, for example, is not only written from the perspective of someone raising a child as a non-binary person and parent, but it also defends the choice of raising a gender-open child.

Second, and more specifically, fatherhood and success in philosophy is not spoken of nearly enough (if at all) in the way that it should be, by which I mean fathers who are explicitly committed to feminist ideals of parenting. I am sorry that I do not have more contributions from fathers, but I am very grateful for the words and stories of Joseph A. Stramondo and Matthew Lindauer (Lindauer co-wrote his contribution to the issue with his partner, Serene Khader). Both of these contributions underscore what Khader and Lindauer call the “daddy dividend,” namely, the ways in which even in relationships committed to feminist ideals of parenting and equal division of household labor, fathers are still benefiting from having to do far less than their partners (in the eyes of the public) to be considered superlative parents.

This volume is robust, but not comprehensive. I hope that you enjoy the many voices, perspectives, and stories that you read in what follows. Let me give you a small taste of what’s to come.

Serene Khader and Matthew Lindauer’s contribution, “The ‘Daddy Dividend’: The Gender Division of Labor and Regression Towards Patriarchy,” articulates something that, I think, so many of us have experienced and spoken about with one another, yet that I’ve never seen articulated so perfectly and eloquently. They begin with an observation, namely, that many well-meaning gestures, comments, and responses to their goal of egalitarian, feminist-informed parenting have ended up reinforcing the very habits that they’ve wanted to guard against. In order to explain this phenomenon, they coin the term “daddy dividend,” which refers to the following phenomenon: while fathers tend to be praised for doing, minimally, what any parent would be expected to do (i.e., for changing a diaper, talking to their baby, or even just holding their baby), mothers are regularly scolded in public by strangers who point out, unsolicited, for example, that their baby is hungry, cold, gassy, or has a runny nose. Khader and Lindauer show how the daddy dividend is pernicious, both in the short- and in the long-term. They argue, “[i]n praising fathers for doing what is simply required of mothers, strangers reinforce the attitude in men that the fatherly duties assigned to them by the traditional gender division of labor are all that should be expected of them.” Khader and Lindauer’s essay goes on to focus on three lessons about distributing household work that they’ve learned from conceptualizing the gender division of labor in parenting: (i) that social rewards for
fathers can exacerbate the tendency of women to engage in shadow labor, (ii) that gender socialization faces would-be egalitarian households with significant startup and maintenance costs, and (iii) that striving for equality may not actually be the best way to achieve equality.

Joseph A. Stramondo’s narrative essay, “Gender, Disability, and the Violent Undercurrents of Parenting Inspiration Porn,” illustrates how the daddy dividend carries over into the experience of parenting while disabled. Specifically, he brings an intersectional lens to the discussion by considering how the daddy dividend is combined, in his experience, with being objectified as “inspiration porn,” a term coined by Stella Young, to designate instances when disabled people are framed as inspirational merely because of the presence of disability. Stramondo explains: “The problem of inspiration porn is that it gives voice to the very low expectations the world has for disabled people. By treating the completion of ordinary tasks as if they were monumental accomplishments, it shows just how incompetent people assume we are. For most disabled people, this experience is pervasive.” Yet, as Stramondo goes on to discuss, experiencing the objectification of inspiration porn is not universal. For example, his partner, Leah, a cisgender disabled woman, has never had to contend with it. Whereas Stramondo is often considered to be father of the year by strangers, simply for existing as a disabled father, Leah is disparaged for being a disabled mother. His essay goes on to consider the ablest presumption that disabled parents are inherently inadequate parents and some of the ways in which this presumption is not only offensive, but dangerous.

Sarah Ayala-López’s essay, “(Philosophizing about) Gender-Open Children,” considers another kind of non-normative parenting, namely, the decision to raise a gender-open child. Raising a gender-open child means that parents do not reveal the gender of their child, or better, reject the assumption that their baby/child already has a gender. This means that when caring for a gender-open baby, parents don’t use binary gender pronouns (like “she” or “he”) to refer to their child; rather, you use “they” singular; and when introducing their baby, they do so as a human who doesn’t have a gender yet. It also means that they don’t reveal their baby’s genital status, which most people assume to be straightforwardly associated with gender and the only factor in determining one’s sex (both false). To some, raising a gender-open child might seem like an odd (or even a harmful) move. But Ayala-López questions these assumptions. In order to do so, they envision three conceptual spaces in which to philosophize on this matter.

The first is the space of parents, in which they respond to what they’ve called the Activism Olympics objection, which questions whether gender-open parenting is just a way for parents to expand their activism, perhaps even at the expense of their child. The second space is the space of children, where Ayala-López responds to the questions of how we can be sure that being raised gender open is good for the child; how not being assigned a (binary) gender can also be good for the child; and whether the absence of an assigned gender at birth has its own set of problems: taken together, what they call the Missing Identity objection. Third, they engage the space of those who theorize about gender-open children. There, they are forthright about the ways in which their decision to raise their child gender-open is not alien to the academic work they’ve done on sex and gender. While being related to their work seems to give their parenting decision a kind of academic street cred, they nevertheless want to keep their theorizing in check. Ayala-López firmly believes that when theorizing about other people’s identities, there are certain ethical guidelines which must be made clear to ourselves, both as individual researchers and as a society. One of these guidelines requires us to answer the question of why we are interested in researching the questions we do. They conclude by developing some of the ethical considerations that should guide our philosophizing about other peoples’ identities. Given various debacles that have plagued our discipline in the past few years surrounding theorizing about people’s identities, we ought to take seriously Ayala-López’s guidance on this matter, for these ethical guides can be generalized beyond the specific question of raising gender-open children. I hope that more people take heed and recognize that philosophizing about identities is a very different matter—with profoundly different consequences—than philosophizing about abstract concepts. People’s lives are on the line. I very much look forward to Ayala-López’s future work on this topic, since there is clearly far more to be said on this matter.

Quill Rebecca Kukla’s paper, “Taking Children’s Autonomy Seriously as a Parent,” considers the culture of parenting older children as they transition into adulthood. This paper is an important follow-up to their 2005 book, Mass Hysteria, which argued that we exist within the strong grip of a damaging cultural myth that dictates that being a good mother requires us to remove the boundaries between mother and infant: This myth still seems to be just as true today as it was fifteen years ago. Within the context of their own eighteen-year-old child (who is also a contributor to this issue!), Kukla has become interested in another agency- and boundary-obliterating parenting mythology. As a point of departure for their project, they assert that “we fundamentally fail to respect the basic autonomy rights that children have as people, and indeed that we equate ‘good parenting’ with severe violations of these autonomy rights. We accept and even approve of subjugating children to their parents’ will in ways that would be shocking if enacted against other sorts of people.” Within this context, they argue for “a radical reorientation of how we think about children’s autonomy and parents’ duties to enable and respect it” with the goal of underscoring what they take to be profoundly wrong with the way that we tend to generally understand parenting and its goals.

Kukla calls attention to a myriad of ways in which we as a society systematically undermine children, in ways that tend to be sidestepped, denied, or outright ignored in mainstream discussions about parenting: for example, that we tend not to actually see children as people (as in, they are considered only to be partial people, or on-their-way-to-being people); that we tend to gaslight children and not think that this is morally wrong; that as parents, we often restrict where children can go and what they can do even when there is no evidence-based argument for doing so.
In her essay “Parenting in Trauma,” Melissa Burchard takes as her point of departure the claim that if we believe and continue to assert that philosophy is important for understanding reality, the world we live in, and our experiences within it, then serious discussions and engagement with issues of parenting and childrearing is important for philosophy as a discipline. She has come to believe this based on her own very difficult experience of parenting, which began fifteen years ago when she and her partner adopted a four- and nine-year-old from foster care, both of whom had been traumatized in multiple ways, by various people in their lives, over an extended period of time. Burchard’s parenting and philosophical interests have informed one another: as a philosopher-mother, she realized that she needed a philosophy of trauma—and more specifically, one that paid attention to the developmental trauma of children who have been traumatized under circumstances like her children—but that such a thing was almost entirely lacking in the philosophical literature. Burchard has recently published a book on this topic, Philosophical Reflections on Mother in Trauma (Routledge 2018), which is a more comprehensive account of the issues she considers here. What she offers here are some implications of trauma for parenting and philosophical understanding.

Specifically, Burchard discusses the ways that trauma affects basic interactions of children in various stages of development, both psychologically and physiologically and the various implications of trauma that philosophy needs to take seriously. Her argument is this: “if infants and children do not get the proper kinds of interactions, they do not develop certain relational abilities, which may prevent them from forming satisfying, mutually trusting relationships throughout their lives.” Generally, such a line of thinking is something feminist philosophers have taken seriously for a while, though focus has not been directed to children. But Burchard wants uptake for this claim beyond feminist circles. She goes on to discuss trauma’s effects on children’s ability to play—what she calls “post-trauma play”—and some of the surprising and difficult things she learned from her own kids that cast into relief what we tend to take for granted. For instance, she questions the assumption that it is always easy for children to engage in imaginative and free play and considers how trauma in early childhood compromises this capacity. Burchard also discusses trauma’s effect on children’s sense of reality and the various ways in which her children were not living in the same world or inhabiting the same reality as her and her partner: in the world of her children and of so many children who have endured early childhood trauma, parents are not people to whom to turn for love and nurturing; rather, parents are people, first and foremost, to be feared (in addition to be loved and depended upon). Before concluding, Burchard considers some of the implications that parenting children who have experienced trauma has for social, political, and moral philosophy, among others, in particular that it can help us articulate the wrongs of oppressions or oppressive actions. Finally, she considers the upshot of her discussion for considering the good life and how even a basic conception of a good life may be challenging when children have been severely traumatized, especially if that trauma has affected or disrupted their developmental processes.

Amy Allen’s narrative, “On Muddling Through,” chronicles how she found success both as an academic and as a mother. I do not want to downplay the success part of her narrative; Amy, on the other hand, would like to focus on all of the ways that the stars have aligned throughout her career to make her decision to have four children commensurate with eventually becoming a full professor and now chair of a robust department with a PhD program, in addition to the author and editor of many books. One of the ways in which Allen’s and Samantha Brennan’s narratives overlap nicely is that while not ignoring the daily difficulties of parenting (young and older children) while working full time, they both recognize that of all careers, academia is probably one of the friendliest toward parents, in the sense that for the most part, as a full-time tenure-track or tenured faculty member, it offers us greater autonomy (and oftentimes flexibility) over one’s work schedule (this is obviously not so for adjunct teaching). Allen still underscores the myriad structural problems that make it difficult to juggle everything, including, but not limited to the fact that there is no mandatory paid parental leave in the United States; no universal, affordable daycare system; and the fact that many people live far away from their extended family. As a result, her words of wisdom are that until the structural problems (of oppression) are resolved, at best we can hope to muddle through it all, where admittedly, having more privilege helps. But in this muddling through, Allen shares some of the strategies that have been particularly helpful to her over the years and still.

As is clear from the title of Samantha Brennan’s narrative, “Parenting, Feminism, and Academic Life: My Happy Story,” her decision to have three children toward the very beginning of her career (beginning, that is, in graduate school) has resulted in a happy story (though obviously the story is still evolving). Brennan is the Dean of the College of Arts and Professor of Philosophy at the University of Guelph in Canada. In the last few years, she co-founded and co-edited Feminist Philosophy Quarterly, an online, open access journal; she also founded a very successful fitness blog that she still runs and that also resulted in a popular book called Fit at Mid-Life: A Feminist Fitness Journey (https://greystonebooks.com/products/fit-at-mid-life). Last year, she ended her three-year term as a member of the Canadian Philosophical Association executive (Vice-President, President, and Past President). While not doing academic things, she also rides her bike long distances, goes back country canoe camping, and races small sailboats. Within the context of such a rich life in which she also has close relationships with her three adult children (ages twenty-one, twenty-three, and twenty-seven), she provides the reader with ten thoughts on what has worked for her and has enabled her to achieve all of her many successes.

In sum, as they write, “we violate all of the ground-level rules we normally recognize for how to treat people as self-determining agents with dignity.” Kukla concludes by introducing a very different framework for engaging with children, one guided by the idea that insofar as children are people, we have a moral duty to inhibit their autonomy and invade their privacy as little as possible.
Carol Gray’s contribution, “Bridging the Divide: Thoughts on Parenting as a Grad Student,” shares some important insights about parenting while a grad student. Gray had a career as a lawyer before deciding to do a PhD in political science. At that time, her son was nine; he's now fifteen. She discusses how she has managed to find the time to be with him (even if it means combining work and play); how bringing him to school with her has been a rewarding and enriching experience for both of them (even if it was done out of necessity); how she has always prioritized family activities and has scheduled classes around family time (even if it’s meant that she has not always taken the classes that most interested her); and, importantly, how she always makes it up to her son when her school conflicts with a family commitment. A perspective on parenting while a grad student is one that I really wanted to include in this issue, since it is fairly common and, again, one that is not discussed nearly enough.

Leigh Viner’s contribution, “Anthropologists from Mars,” discusses her experiences single-parenting her now grown son (who, maybe not coincidentally, was a philosophy major in university). She thinks of this experience as being like an anthropologist from Mars, with him as her assistant. She tells us how the philosophers who have had the biggest impact on her—philosophers who value art, play, freedom, dignity, and openness—have also been the ones that have shaped her parenting and how it was, in fact, her son who led her to feminism. Her message is simple, namely, that if we are doing it right, then “we learn at least as much from our children and from our students as they do from us.”

Speaking of philosophy-parents and philosophy-sons, the penultimate contribution to the issue is by the son of another philosopher included in the newsletter, namely, Eli Kukla (Quill Rebecca Kukla’s son), who discusses what it’s been like to be raised by philosophers. I’m excited to be including in this issue the perspective of a child of philosophers—one that we definitely don’t hear enough of. What excites me most about including these two contributions side-by-side is that much of what Quill Kukla discusses in their article from a theoretical and normative perspective (i.e., how we should treat children, the virtues of actually respecting the autonomy of children, even if our society fails to see them as full people) is exemplified in Eli Kukla’s thoughts and is visible in his achievements and robust philosophical outlook and perspectives. What I mean is that Quill Kukla has not only presented an argument in favor of a certain type of parenting; in Eli Kukla, we have proof, in the flesh, that the type of parenting they are advocating works!

“Raised in Philosophy” discusses what it’s like to be what Eli Kukla calls a “pure-bred philosophers’ kid,” meaning that both of his parents and his grandfather are philosophers. Eli Kukla gives us a number of examples of the ways in which his philosophical upbringing has shaped who he is today and how having been immersed in philosophy “has affected seemingly every aspect of [his] life and personality.” As Eli Kukla writes and, later, expands upon, “Most people, at some point in their lives, reach a point where they question many of their assumptions about life. People question religion, sexuality, monogamy, traditional family structures, gender, et cetera. But for me, being raised by philosophers, I never started out with any of these assumptions to begin with.” Ultimately, seeing the profession through the eyes of someone who has grown up immersed within it, Eli Kukla’s essay reminds us of many of the wonderful things about the profession, which (depending on the day, the scandal, etc.) tend to be easy to forget.

Finally, I wrote a short narrative piece for this issue, entitled “Children, Parenting, and the Nature of Work,” where I share some feminist observations from my recent experience of doing philosophy with children.

In addition to the contributions just mentioned, in this issue you will find the following book reviews: Lisa Tessman’s review of Eva Feder Kittay’s Learning from My Daughter: The Value and Care of Disabled Minds, Claire A. Lockard’s review of Overcoming Epistemic Injustice: Social and Psychological Perspectives, edited by Benjamin R. Sherman and Stacey Goguen, Jina Fast’s review of Sophie Lewis’s Full Surrogacy Now: Feminism Against Family, and Kathryn Lafferty-Danner’s review of Loretta Ross and Rickie Solinger’s Reproductive Justice: An Introduction.

I am so grateful to all of the authors in this issue. It’s been a really fun, fulfilling, and illuminating issue to put together. I’m also grateful to Brian Robinson and Sabrina Little for their help reviewing articles. Finally, I am forever grateful to my mentor and friend (and also Chair of the APA Committee on the Status of Women) Kate Norlock for her editorial advice on this issue.

I hope that you enjoy this issue!

ABOUT THE NEWSLETTER ON FEMINISM AND PHILOSOPHY

The Newsletter on Feminism and Philosophy is sponsored by the APA Committee on the Status of Women (CSW). The newsletter is designed to provide an introduction to recent philosophical work that addresses issues of gender. None of the varied philosophical views presented by authors of newsletter articles necessarily reflect the views of any or all of the members of the Committee on the Status of Women, including the editor(s) of the newsletter, nor does the committee advocate any particular type of feminist philosophy. We advocate only that serious philosophical attention be given to issues of gender and that claims of gender bias in philosophy receive full and fair consideration.

SUBMISSION GUIDELINES AND INFORMATION

1. Purpose: The purpose of the newsletter is to publish information about the status of women in philosophy and to make the resources of feminist philosophy more widely available. The newsletter contains discussions of
recent developments in feminist philosophy and related work in other disciplines, literature overviews and book reviews, suggestions for eliminating gender bias in the traditional philosophy curriculum, and reflections on feminist pedagogy. It also informs the profession about the work of the APA Committee on the Status of Women. Articles submitted to the newsletter should be around ten double-spaced pages and must follow the APA guidelines for gender-neutral language. Please submit essays electronically to the editor or send four copies of essays via regular mail. All manuscripts should be prepared for anonymous review. References should follow The Chicago Manual of Style.

2. Book Reviews and Reviewers: If you have published a book that is appropriate for review in the newsletter, please have your publisher send us a copy of your book. We are always seeking new book reviewers. To volunteer to review books (or some particular book), please send the editor, Lauren Freeman (lauren.freeman@louisville.edu), a CV and letter of interest, including mention of your areas of research and teaching.

3. Where to Send Things: Please send all articles, comments, suggestions, books, and other communications to the editor: Dr. Lauren Freeman, University of Louisville, lauren.freeman@louisville.edu.

4. Submission Deadlines: Submissions for spring issues are due by the preceding November 1; submissions for fall issues are due by the preceding February 1.

ESSAYS

The “Daddy Dividend:” The Gender Division of Labor and Regression Towards Patriarchy

Serene Khader and Matthew Lindauer
CUNY Graduate Center and Brooklyn College

When we welcomed our baby last winter, there were some things everyone seemed to want to know. Was she sleeping? How were we feeding her? What kind of casserole did we want? But, really, was she sleeping?

One of the biggest challenges we faced was one that no one asked about: deciding how we would divide parenting and household labor. That no one asked about it, of course, did not mean that our division of labor was wholly private. We knew from feminist scholarship, and from the surprising amount of work it had taken to divide household labor in an egalitarian way when we were childless, that entrenched social inequalities would rear their heads in our home. We were less prepared, however, for the extent to which many well-meaning gestures, including ones that are intended to promote egalitarianism, on later reflection, would turn out to reinforce habits we wanted to guard against.

We coined the term “daddy dividend” on one of the many days a stranger on the subway told Matt (a white man) that he was the “best daddy ever.” The thing he had done to receive this accolade was to wear his baby in a carrier, and perhaps not seem utterly miserable doing so. By contrast, Serene (a brown woman) has never been told by a stranger that she’s the best mom ever, or even a decent one. It’s difficult to disaggregate the effects of race and gender in our case, as we know that women of color are widely perceived as inferior mothers. In fact, because our baby has lighter skin than Serene does, sometimes people do not perceive Serene as her mother at all.

The daddy dividend is a cluster of gestures—again, well-meaning in many cases—that negatively tax the social esteem of men when they publicly perform basic parental duties: taking their children places, feeding them, changing their diapers, talking to them, and so on. It is the corollary of the positive tax on moms for their public parenting; moms are regularly scolded by strangers stating confidently (sometimes indignantly!) that their baby is hungry, cold, gassy, or has a runny nose. Of course, this scolding may take on different tones and intensities, and social meanings depending on the mother’s race, class, sexuality, ability, and other social inequalities, and so may the compliments given to fathers. But our experience has been that momming in public carries a cost; dadding in public carries a benefit.

We do not wish to suggest that the gender division of childcare is sustained entirely, or even primarily, by such microcompliments and microaggressions. However, we have found it helpful to think of the traditional gender division of labor in parenting as a social practice that is supported, in part, by norms governing the assignment of esteem, visibility, and value. In what follows, we will focus on three further lessons about distributing household work that we have learned from conceptualizing the gender division of labor in this way: that social rewards for fathers can exacerbate the tendency of women to engage in shadow labor, that gender socialization faces would-be egalitarian households with significant startup and maintenance costs, and that striving for equality may not be the best way to achieve equality.

Dividing childcare and household labor is not a snapshot in time, even if selfies with the kids claim otherwise. It is an ongoing practice that is impacted by our society’s preexisting norms and expectations around what fathers and mothers are expected to do. To borrow H. L. A. Hart’s term, it is possible to take the “internal point of view” on the traditional gender division of labor, wherein participants accept these norms and govern their behavior accordingly. This can be done through direct, self-conscious action—openly declaring “I will not wash the dishes, it’s not a man’s job!”—or through more subtle means—thinking to oneself, “I played with the baby for twenty minutes, so that counts for at least twenty minutes of cleaning the dishes. Back to reading.” The internal point of view on the traditional gender division of labor has it that mothers and fathers bear distinct parenting and household duties. Any engagement by fathers in the duties associated with mothering is therefore supererogatory—beyond the call of...
fatherly duty. Hence, the same behavior is praiseworthy for one in virtue of being more than what they are required to do—for the other, it is middling at best. For a mother answering a work call while pushing the stroller, negative comments are required to remind her of the duties the practice prescribes her. She had better not be getting the idea that she should instead adopt the external point of view on the practice, the position of an observer that does not accept or even rejects its norms.

One might be forgiven for thinking that in offering Matt the daddy dividend strangers encourage him to engage more in the activities associated with mothering. But this misses the centrality of obligation and supererogation to the gender division of labor. In praising fathers for doing what is simply required of mothers, strangers reinforce the attitude in men that the fatherly duties assigned to them by the traditional gender division of labor are all that should be expected of them. Picking out an action for praise expresses the judgment that it is not just what you should be doing—it would be weird to praise someone for brushing their teeth, or wearing shoes in public. But it is common to give men the daddy dividend, in the form of public praise, for performing actions that are just as common and expected for mothers.

By seeing the gender division of labor as a practice, the further insight follows that any egalitarian revision to it will also be a practice, and one that is affected by its predecessor. That is to say, we cannot ignore the fact that partners trying to set up an egalitarian gender division of labor need vigilance and ongoing carefulness if they wish to avoid slipping back into the package of patriarchal norms that the inherited practice entailed. As the daddy dividend illustrates, there are signs all around pulling us back towards the old way of doing things.

Moreover, rewarding men for engaging in high-visibility parenting encourages women to engage in low-visibility parenting, or, as it is often called “shadow labor.” As most parents know, most of the important labor of parenting is not done in public. Much of the most physically and cognitively demanding labor associated with parenting is done outside of the jaunt in the park or down the street for coffee—it’s in the middle of the night, or in one’s vanishingly small leisure time. The daddy dividend encourages men to invest in public parenting. Since they are likely to take other men’s behavior and not equality as the relevant baseline, and they know that men who engage in high-visibility parenting are doing more than other men, the daddy dividend encourages women to shift to less visible, and often more taxing, labor that still needs to be done. This includes the more mundane, or even yucky, elements of parenting (cleaning up the food the baby has thrown all over the kitchen) and the much-discussed “cognitive load,” which is typically code for the managerial tasks associated with mothering (making sure that a baby has her next size of clothing, ensuring that paid childcare arrangements are made, keeping the household’s schedule, making doctor’s appointments, and so on).

Of course, the perception that men’s engagement in feminized parenting tasks is supererogatory does not begin with the daddy dividend. It begins in childhood socialization, and thinking of parenting as a practice through which value is assigned to tasks has also helped us to think through what it means to directly address the role of this socialization. Boys are generally socially habituated not just not to think about planning for the care work of babies, but also not to learn the skills associated with doing so. The perception that women are naturally better at this work also works to invisibilize the mechanisms by which women and boys acquire these skills, and to discourage men and boys from bothering to acquire them—or parents from bothering to teach them. This means that the gender division of labor in parenting in a mixed-gender couple is usually one where women have an advantage with respect to the relevant skills.

Thinking of socialization as producing specialization rather than just stereotypes has helped us think more concretely about what is involved in changing its inequalitarian effects. We know that it is easy and efficient (in some sense, at least) to do what we already specialize in. This means that there are startup costs for mixed-gender couples trying to achieve equality. Some of these costs fall on men, who need to learn, but it is important to recognize that the costs are often borne by women who have to teach them. For instance, Matt has taken up the task of contacting relatives for used clothing and tracking whether our daughter’s current clothing fits her. But for this to happen, Serene had to schedule a lesson on arranging and sorting clothing and anticipating future clothing needs.

Our third and last point here follows from the fact that, given all of this, the ongoing influence of the gender division of labor is more likely to produce inequalitarian than egalitarian outcomes. For this reason, aiming at equality may not be the appropriate framing to guide men’s action. Like any form of praise, the daddy dividend can be seductive, and we aren’t recommending confronting well-meaning individuals when such praise is being offered (they mean well, and there are a lot of people who might feel threatened—not good, especially when your kids are around). If you are a dad who really wants to do something about the daddy dividend, it’s probably also a bad idea to heap further internal praise on yourself for noticing it. When you’re faced with such a system of social rewards, the better strategy, we believe, is to overcompensate in some domains of household labor. Perhaps egalitarian men should do “more” of the cleaning, diaper-changing, laundry, clothing turnover and organization, and other unappealing tasks in the household. Because the perception that a man is doing “more,” when so much in our society is telling him that he has done more than his share already, is likely to just be a reflection that things are a little bit closer to equality.

Just as people’s selfies tell us little about whether they’re actually happy and living well, men’s selfies with their children tell us little about whether they are parenting and partnering well. By shifting our focus on snapshots to sustained parenting practices, we may just be less likely to confuse the two. And as a final note, if you are worried that what we have said seems too justice-oriented or juridical for such a personal area of one’s life, we’d like to invite you to think about our points in terms of love. If you love your
partner, what do you really owe to them? It’s probably not what an outmoded, patriarchal social practice says you do.

NOTES

1. The phrase “daddy dividend” has, unsurprisingly, been used by others, but not in the way that we do here, and sometimes in ways that are at odds with our commitments. For instance, Douglas Abrams describes a “daddy dividend” that children supposedly receive for having an involved father’s presence in their lives (“The Daddy Dividend,” Psychology Today, March/April 2002). As feminists and egalitarians, we have concerns about this sort of argument, but discussing them is beyond the scope of this brief article. Others use the term to describe a career boost that fathers seem to receive relative to fatherless men (see, e.g., https://www.theatlantic.com/magazine/archive/2014/01/the-daddy-track/355746/).

2. This stereotyping is particularly pronounced and well documented in the case of Black women. For a discussion of stereotypes of Black motherhood, such as the association of Black women with unwed motherhood, excessive sexual desire that prevents virtue, and “welfare queens,” see Dorothy Roberts, “Racism and Patriarchy and the Meaning of Black Motherhood,” https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1594&context=faculty_scholarship

3. See Joe Stramondo’s piece in this newsletter, “Gender, Disability, and the Violent Undercurrents of Parenting Inspiration Porn,” for a discussion of how a context of disability oppression transforms the daddy dividend into an occasion for inspiration porn when the dads in question have visible disabilities.

4. See Ann Cudd’s Analyzing Oppression (Oxford University Press, 2002) for an analysis of how social structures work to force women to “choose” to engage in disproportionate caregiving.

5. H. L. A. Hart, The Concept of Law (Oxford: Clarendon Press, 1961). We are borrowing and using this term, as well as the “external point of view,” for our distinctive purposes here, which vary considerably from Hart’s own in developing a theory of law.


Gender, Disability, and the Violent Undercurrents of Parenting Inspiration Porn

Joseph A. Stramondo
SAN DIEGO STATE UNIVERSITY

Her bright purple jacket, pink tights, silver tutu, and honey-colored hair present a sharp contrast to my black coat, black pants, black beard, and black power wheelchair, as we zoom down the sidewalk together, chatting away. She first learned to ride on my chair when she was eighteen months old. We started in the safety of the manicured grounds of our apartment complex and slowly expanded our range to the shopping center down the block and, eventually, the trolley system that opened nearly the entire city to our adventures. At first, I kept a bag of fruit snacks in my pocket to use for bribes when something was just too interesting to keep her toddler hands off of, but I haven’t had to use that trick in several months now. Having her constant stream of questions answered seems reward enough to keep her perched on my lap. She is my companion nearly every time I leave the house with a destination other than my office, my classroom, or the airport.

This Saturday morning, we are running an errand: picking up some items her mom needs from the pharmacy. There is a small group of people loitering by the entrance and I can immediately tell they want to say something to us. “Hey man, look at you! Is that your little girl? You’re a great dad! You’re father of the year!” One of them tries to give me a high five as we roll past, but my hands are occupied; one is on my chair’s joystick and the other clutches Hazel loosely to help her balance. She ignores them completely and I mutter some nicely. Neither of us want to interrupt our trajectory into the store. I’m thinking I might want a bottled iced coffee to help me keep up with her and her brother all day and she has already been practicing the tone she will use to convince me to buy her some chocolate at 10 a.m.

Maybe we appear rude to an outsider, spurning their words of encouragement in our haste to acquire a jar of acetaminophen. The truth is that I’m quite proud of Hazel’s intuitive ability to recognize and, when appropriate, shrug off ableism. It was, in fact, ableism we were encountering.

In particular, we were being used as real life “inspiration porn.” Stella Young coined this term to designate instances when disabled people are framed as inspirational merely because of the presence of disability: “I use the term porn deliberately, because they objectify one group of people for the benefit of another group of people. So, in this case, we’re objectifying disabled people for the benefit of non-disabled people. The purpose of these images is to inspire you, to motivate you, so that we can look at them and think, ‘Well, however bad my life is, it could be worse. I could be that person.’” She speaks freely of her own frequent experiences with this phenomenon, and I’d guess her words hit home with many, if not all, visibly disabled people: “I’ve lost count of the number of times that I’ve been approached by strangers wanting to tell me that they think I’m brave or inspirational, and this was long before my work had any kind of public profile. They were just kind of congratulating me for managing to get up in the morning and remember my own name. And it is objectifying.”

The problem with inspiration porn is that it gives voice to the very low expectations the world has for disabled people. By treating the completion of ordinary tasks as if they were monumental accomplishments, it shows just how incompetent people assume we are. For most disabled people, this experience is pervasive.

However, experiencing the objectification of inspiration porn isn’t universal when it comes to parenting. In fact, it is an experience that my partner, a cisgender disabled woman, has never had. To be sure, many people still regard her parenting with what Silvers calls the presumption of incompetence. However, it is not expressed in the same way. Rather than being told that she is mother of the year, she is asked why her toddler isn’t wearing a coat.
Of course, this difference has everything to do with gender. When I am celebrated as a disabled father and Leah is disparaged as a disabled mother, I am being rewarded by what Serene Khader and Matthew Lindauer refer to in this same newsletter issue as the Daddy Dividend, “a cluster of gestures . . . that negatively tax the social esteem on men when they publicly do what they simply should do as basic parental duties.”

Yet, at their root, these experiences are not so different. In both cases, there is this presumption that the disabled parent is inherently an inadequate parent. In both cases, the stakes are incredibly high. How long would it take for me to lose father of the year status, for example, if one of my kids were to even minorly injured while under my care in public?

From our very first conversations about starting a family, Leah and I discussed the need to ensure our children’s health and safety to a much greater degree than what would be required of non-disabled parents. I practiced changing diapers on stuffed animals and lifting bags of flour in and out of our adapted crib because we believed that it was not a question of if child services would be called on us, but when. We knew that all of our racial, educational, and economic privilege would not protect us from the hostility directed toward disabled parents.

While this fear has not yet been realized almost four years into our parenting journey, it was certainly not unfounded. In the United States, there is a long history of violence against disabled parents, especially disabled mothers, beginning with the eugenics movement and its emphasis on forced sterilization to prevent disabled people from becoming parents in the first place. These same attitudes toward parenting with a disability are now expressed via discrimination in access to advanced reproductive technologies and child custody, which has been our greatest fear.

In a way, an analysis of the inspiration porn associated with parenting, specifically, may be what is needed to illuminate exactly what is so problematic about inspiration porn more generally. In my view, inspiration porn is not merely offensive, but it is dangerous. As Stella Young suggests, it is an expression of the statement, “Well, however bad my life is, it could be worse. I could be that person.” However, it seems to be more than that. It is also an expression of the statement, “If you fail to accomplish this mundane task in a way that we approve of, you will be subject to disproportionate ridicule, censure, control, and, if the stakes are high enough, punishment.”

ACKNOWLEDGMENTS
I would like to thank my spouse, Leah Smith, for helping me develop the ideas in this essay with a countless number of conversations about all of the issues raised here.

NOTES

(Philosophizing about) Gender-Open Children

Saray Ayala-López
CALIFORNIA STATE UNIVERSITY, SACRAMENTO

I’m at the playground with my baby, and a smiling adult inquires, “Is it a boy or a girl?” Scientific studies show that if I say X, they will see my baby as doing A, being A, feeling A—versus if I say Y. They’ll likely make different assumptions about whether my baby is able to climb up the playground structures and sit without support, and they’ll encourage my baby to engage in different activities. And of course, they’ll respond to them differently depending on whether they think the baby is a boy or a girl. What do I do if I don’t want that to happen? One way to respond to the question is not to reveal the gender, or rather, to reject the assumption that my baby already has a gender, which is arguably a weird move, breaking out of a smoothly functioning, well-oiled social exchange. Here I want to talk about this weird move: How weird is it? What are some of the reasons to make the move anyway, and what are some of the concerns?

Gender-open children are children who are not assigned a specific gender at birth. When caring for a gender-open baby, you don’t use binary gender pronouns (like “she” or “he”) to refer to them (you can use, for example, “they” singular), and you introduce your baby as a human who doesn’t have a gender yet. The best way to protect the baby from being treated and perceived according to the gender binary (as either a girl or a boy) is not to reveal their genital status, which is commonly taken to be critical in the determination of sex and to be straightforwardly associated with their gender. This can be done to different degrees, depending on the specific family and social context. When thinking about gender-open children, I see three conceptual spaces to philosophize: the space of parents; the space of children; and, finally, the space of those who theorize about gender-open children. Let’s start with the parents.

1. PARENTS
In this section I explore three different questions around the decision to raise a gender-open child.
1.1. ACTIVISM OLYMPICS

When I first put together my plan of having a baby and thought about the issue of gender, my first thought was “I would like gender to have no role in the baby’s life—to the extent that I can control this—until they themselves can decide how much gender they want in their life, and what kind of gender.” Later, I discovered that gender-open parenting is actually a thing. Just like attachment parenting, and many other parenting styles that surround any parent, like a nightmare landscape full of high-stakes choices. And pretty much any choice brings about social judgments and the accompanying consequences.

On the one hand, gender-open parenting being a thing makes it easier to explain it to relatives, friends, and strangers, as you constantly face the question, “Is it a boy or a girl?” When I respond: “I’m doing gender-open parenting,” people assent, often with a poker face, and the conversation ends there. The power of a label! Whether or not they understand what it means, they understand that it is a thing, and its metaphysical, if totally mysterious, weight gives everyone at least a sense that the question about the baby’s gender has been answered, even if not in the way that they had expected. When I try the long explanation instead, it almost never works out, and we all leave the conversation frustrated. An additional merit of gender-open parenting is the very fact of naming this practice. After learning the notion of hermeneutical injustice, and/or experiencing hermeneutical injustice yourself, you come to very much appreciate new concepts that fill the conceptual gaps around marginalized identities, their experiences, and the practices that are relevant for them. Instead of the perhaps vague thought I mentioned above (“I would like gender to have no role in the baby’s life—to the extent that I can control this—until they themselves can decide how much gender they want in their life, and what kind of gender”), I can straightforwardly say, “I’m doing gender-open parenting. This is what I want for my child,” perhaps modifying with this labeling my own understanding of what I’m doing. This label also makes it easier to find other parents and caregivers who are approaching their child’s gender in a similar way.

On the other hand, however, this label makes me cautious, because it puts the focus on the parents. And even though this couldn’t be otherwise (the baby is not the one who can decide whether or not to be assigned a gender at birth), I want to make sure this doesn’t become one of my lifestyle garments, like the fair-trade organic tea I consume, or the rainbow pin button I wear on my blazer. The specific question that concerns me is this: Is gender-open parenting a way for parents to expand their activism (perhaps, at the expense of their children)?

According to what I will call the Activism Olympics objection, gender-open parenting is not only consist of not assigning them a binary gender at birth. It is also about how they grow up, or how the characters in the tales I read to them interact and are referred to. That is, my child is being exposed to a set of ideas around gender, and in general, to a view of personal identity and personal relationships that is not determined and ruled by the dominant binary gender schema. Thus, I’m definitely trying to instill this more flexible view of gender and identity in my child.

It seems that having spent time thinking about gender and having strong ideas (and feelings) about it makes my parenting decision more vulnerable to the Activism Olympics objection. I do object to the gender binary, and I believe that people would be happier if freed from it. Raising a gender-open child sounds like a practical way of fighting against the gender binary, more eloquent than any paper I could ever write. But would the opposite (i.e., not having spent time thinking about gender, not having strong ideas about it) make my parenting decision less objectionable? It seems like an absurd result: the less involved you are intellectually, morally, and politically in this parenting decision, the less controversial it would be. To avoid the Activism Olympics objection, we need to find a different avenue: a possible way out of this objection is finding a place where we can raise our children according to our values, and this is not necessarily using them as instruments for our own betterment.

By taking this route, we are then confronted with the general question of whether it is ok to instill your values in your children. Raising my child gender open does not only consist of not assigning them a binary gender at birth. It is also about how they grow up, or how the characters in the tales I read to them interact and are referred to. That is, my child is being exposed to a set of ideas around gender, and in general, to a view of personal identity and personal relationships that is not determined and ruled by the dominant binary gender schema. Thus, I’m definitely trying to instill this more flexible view of gender and identity in my child.

If we decide that instilling your own values in your children is not a good practice, or needs independent justification (e.g., an argument that those values are good), then gender open parenting would need an additional defense, but the same could be said of raising children in the gender binary, which people commonly do. Being the default option doesn’t make binary-gender parenting value-neutral. Assigning babies a binary gender at birth and treating them, more or less intentionally, according to that gender, reveals a specific set of values (e.g., about what gender is and should be, what is best for the child, and how society is best organized). Someone could say that the fact that most people endorse those values and instill them into their children is already a good reason in favor of this default practice. But default options, as tempting as they are to our often lazy minds, need not be the best. They can even be detrimental to our well-being (there are plenty of examples in our history of default options that we now
acknowledge as detrimental to everyone involved, such as embracing slavery or opposing interracial marriage. In order to conclude anything about the goodness of the set of values surrounding the default practice of gender assignment at birth, we need independent reasons other than the sheer number of people doing it. I leave it for a different place to discuss what set of values is preferable.

1.2. FREE THE CHILDREN

We could say that when someone decides to raise their child gender open, they are doing it for the child’s benefit, so that the child can flourish without the limitations imposed by the dominant gender norms. But can we be sure that being raised gender open is good for the child? How is not being assigned a (binary) gender good? One quick answer is that at a minimum, it doesn’t have the negative consequences that assigning a binary gender has. You just need to glance at the studies on how the assigned (binary) gender of children mediates our perception of them (and their perceptions of themselves), to the extent of making us see different things depending on the gender we believe a child has (see references above). If you look at teenagers, and at your own teenage years, it’s easy to see how expectations for the specific gender you are assigned (whether or not you identify with it) can make you miserable, from not having the perfect body shape, to not having the right talents.

However, this response leaves unanswered the question of whether the absence of an assigned gender at birth has problems of its own. One common objection in this regard is what I call the Missing Identity objection, according to which gender-open children are missing a critical element to build their identities. Gender is very important to developing our identity, the reasoning goes, and gender-open children are being deprived of that. This objection has several problems. First of all, it misinterprets what raising a child gender open is: gender-open children are not being deprived of a gender identity. Rather, they simply do not have one imposed upon them, and are left free to pick one whenever they are ready, and change it as many times as they feel like. Moreover, it might be the case that eventually they do choose a gender within the binary. The difference is that it was a choice and specifically, their choice. Second, it assumes several problematic things: (i) that the gender binary is universal, as if all societies are organized according to this Western binary; (ii) that identity has to be necessarily mediated by this binary gender; and (iii) that reproductive capacities determine our identity, via determining our gender. Gender is essential to our identity, the reasoning goes, because it connects us to our communities by placing us in specific social roles, it gives us a place within that society, one that has to do with our anatomy, in particular our reproductive capacities. In sum, this reasoning assumes a Western, binary and cis status quo, spiced up with an essentialist view of gender. I won’t argue against any of those assumptions here; there are wonderful works out there doing that already. The take-home message I would like to underscore is that gender-open parenting does not deny the importance of gender as a dimension of identity, nor does it deprive children of it (or foreclose the possibility that one day the child might choose to embrace a gender within the binary); but rather, it calls into question those specific assumptions mentioned above. It contests the idea that in order to flourish as a person you need to give a very specific value to the dimension of gender, according to a strict, binary and very old set of norms.

Besides the above-mentioned problems, behind the Missing Identity objection seems to lurk the fear that if children are not assigned a gender at birth, then they might not grow into a stable binary gender identity. Instead, they might identify as non-binary, or who knows what. In response to this concern, first, another clarification is in order: raising a child gender open is not imposing upon them a non-binary gender identity. Rather, the idea is to free them from any imposed gender identity, until they have the knowledge and experiences to decide for themselves, in the same way they decide about other features of their identities. These children will grow up to identify in whatever way works for them, within or outside the binary, in a more or less fluid way. Second, this concern betrays a plain rejection of non-binary identities: Why are we assuming that only a stable binary gender identity is the right kind of identity? A different argument is needed here to arrive at the conclusion that non-binary identities are problematic. There are some arguments out there concluding that non-binary identities are not real. For example, someone might have a social-position account of gender and conclude that there is no non-binariness, for when you look at how our societies are structured and normed, the reasoning goes, there is no place for non-binariness; that is, non-binariness is not a social position in the way that binary genders are; therefore, by definition, it doesn’t exist.

There are at least three ways to respond to this reasoning. First, we can advocate for an account of gender that is not exclusively based on social position, but which also includes self-identification. With this other account in hand, a lack of nonbinariness as a social location does not invalidate non-binary gender identities, as the latter does not hinge on the former. Second, we can argue, with Bettcher, that mainstream social practices around gender are not the only ones in town. Instead, there are what she calls subaltern cultural practices around sex and gender that validate non-mainstream gender identity claims. In Bettcher’s proposal, as long as there are cultural practices around gender, even if in minority, resistant spaces, gender identity claims get validated. Third, we can expand Bettcher’s proposal and say that in order to validate a gender identity, established (even if in minority spaces) cultural practices are not necessary. While a (subaltern) cultural practice expands and supports individual patterns of resistance and efforts at identity-construction, a wholesale culture (e.g., practices, collectively shared concepts), even if subaltern, is not necessary to validate experiences that build identities. For example, someone might develop inchoate resistant concepts for their experiences in conversations with a close friend, in the absence of a collectively acknowledged space for those experiences. An interlocutor who listens and makes the effort to understand in the way prescribed by Medina could be enough to validate those experiences as identity-building.
2. CHILDREN

In this section, I explore the following question: How are gender-open children contesting dominant gender norms? I think gender-open children are contesting the gender binary, and dominant gender norms more generally, in a new way. Although I identified the space of children as a separate space to reflect upon, when thinking about how gender-open children might contest dominant gender norms it is clear that at least for their first years, if we require some sort of intention for contestation to happen, then it is the child’s parents and/or caregivers who contest these norms, not the child themselves. Perhaps a better way to articulate this inquiry into how gender-open children contest dominant gender norms would be this: What are gender-open children inviting us to do differently (in relation to dominant gender norms)?

Gender-open children invite us to rethink dominant ideas and practices about gender. Here I focus on one of them, the Gender Disclosure norm. According to this norm, we owe it to society to disclose our gender. Not doing so means failing to play by the rules. It’s like not respecting the turn taking in a conversation or randomly changing the topic: you are breaking norms that help coordinate social behavior. Not disclosing your gender prevents others from treating you appropriately, for they won’t know how. It causes confusion. Moreover, the expectation is not only that everyone discloses their gender, but also, that we do it correctly, unless we want to deceive others. Trans folks are often accused of deceiving, because they might signal a gender that is, according to dominant gender norms, not the correct one. For Talia Bettcher, the normalized practice is about disclosing your sex (or rather, your genital status, which plays a crucial role in determining one’s sex in our cultural practices), and this disclosure happens by marking your gender appropriately (“appropriately” here works under the assumption that there is a straightforward relationship between one’s sex and one’s gender). In her words, “Gender presentation literally signifies physical sex.”

Marilyn Frye writes that this compulsory disclosure comes with an associated urgent need “to know or be able to guess the sex of every single person with whom one has the slightest or most remote contact or interaction.” This need to know becomes intensely apparent when you care for a gender-open baby. I remember a friend expressing their frustration with exactly those words after repeatedly asking for the baby’s gender: “But I need to know.” This need might start fading as soon as you make the effort to pause and reflect on what you need that information for. In the case of babies, this pause should be enough to realize you actually don’t need that information.

There is a decision tree guide cartoon, “How to tell if a toy is for boys or girls: A guide.” It consists of only one question: “Do you operate the toy with your genitalia?” If yes, it’s not for kids. If no, it’s for either boys or girls. And, I would add, for those who are neither. Somehow, this simplicity ends up lost on us. We panic, as did the owner of a shoe store in my hometown. An old family friend, the store owner decided to give my baby a gift: a pair of sandals. The owner headed for the shelves but stopped dead in tracks: “But . . . I don’t know which color.” The owner stared at us, paralyzed in the middle of the shop, with shelves of sparkling pink shoes shining in the background, as if desperately trying to entice all of us away from our resistant attitude. My sibling came to the rescue, casual and effective: “What about yellow, white, red, green, orange, or black? There are many colors.” The shoe store owner came back to life from what had looked like a spell: “Oh, of course.” With contented expression, the owner handed us a pair of tiny white sandals. The shoes were for the feet—baby girls’ feet, baby boys’ feet, or gender-open-babies’ feet, and be they white, yellow, blue, or green, they all get pulled off, chewed on, and thrown out of the stroller for the fun of watching your parent pick them up, over and over again—but that’s a whole other story.

The shoe store owner was not an outlier in feeling at a loss for a moment. In societies like the one I’m writing from, there are two different scripts that track perceived sex/gender. “[I]n everything one does,” Frye writes, “one has two complete repertoires of behavior, one for interactions with women and one for interactions with men. Greeting, storytelling, order-giving and order-receiving, negotiating, gesturing deference or dominance, encouraging, challenging, asking for information: one does all of these things differently depending upon whether the relevant others are male of female.”

Gender-open children do not conform to the Gender Disclosure norm, providing an opportunity to reflect on it. And they invite us to embrace the dissonance of not knowing what script to follow, and to be creative. However, gender-open children are not necessarily promises of annihilation of the gender binary. When they grow up and start making decisions about their gender identity, they might contest the gender binary and other dominant gender norms, or they might end up endorsing (some or maybe even all of) them. If my child ends up embracing a binary gender identity, and one that has been traditionally associated with the way the hospital sexed their body when they were born, I will not take that as a defeat. Any outcome that stems from my child’s freedom will be a success and the very goal of this entire endeavor.

3. THOSE WHO PHILOSOPHIZE ABOUT GENDER-OPEN CHILDREN

As a feminist philosopher who has worked on sex and gender, raising a gender-open child is a tasty topic to philosophize about. Here I am, after all, writing this piece on gender-open children.

As I wrote above, my decision to raise my child gender open is not alien to the work I’ve done on sex and gender. While being related to my work seems to give my parenting decision a sort of academic halo, I want to make sure this halo doesn’t reach too far. I want to make sure that I keep my theorizing in check. Let me explain my concerns.

When philosophizing about gender-open children, I set a warning to myself: I should make sure my mental gymnastics take into account that I’m talking about a specific person (my child) and real people in general, people who, in this case, are not in a position to speak for themselves, but
who will be at some point. When theorizing about other people’s identities, I see that an ethical commitment is required. They are not tables, propositions, or hypothetical aliens. This ethical commitment is, interestingly, nothing extraordinary to ask for.

Having a research ethics is not, at least it is not intended to be, a limit on the sort of conclusions we can arrive at, given the facts. It’s rather about the way we proceed in our research, the values that guide our inquiry, and the goals we set for it. This is just the paraphernalia that accompanies any scientific inquiry. A research ethics for theorizing about other people’s identities invites us to reflect, as individual researchers and as a society, on why we are interested in researching the questions we do. This applies to science generally. We don’t do science on everything. Not all possible questions are investigated, not even all reasonable questions. Only those that bear on something we are worried about, something we are interested. For example, there are many scientific studies comparing cognitive capacities of humans and chimpanzees, but we are a lot less concerned with comparing cognitive lives of turtles and tortoises. The same happens in philosophy. It’s only the overly optimistic or, perhaps rather, vain philosophers who think that there is no actual limit to the questions that philosophy can explore. There are limits to philosophy: the limit is set by our interests, our concerns, our desires, and available resources. Every once in a while, we should wonder why we inquire into the questions we do (again, as individuals and as a society). This meta-inquiry can reveal a lot about the very issues we are investigating.

Take, for example, the so-called “science of sex differences”: Why does it never lose its sexiness? Why do we keep producing studies about sex differences? Asking this question can tell us more interesting things than perhaps any single study within it. Perhaps we are obsessed with sex differences for reasons that are independent of how many differences there actually are and what their origin is. Similar things can be said about studies on racial differences. There are those who see in this meta-question a threat to freedom, and a desperate attempt at hiding ugly truths. “Let science be science,” they say as a response. But, as feminist philosophers of science have been arguing for decades, science is not something that happens independently of (communities of) scientists who work within specific social contexts. Whatever the “science” in “let science be science” is, it is already inevitably shaped by interests, values, and fears of our society. So we are letting it be science, while also being mindful of the choices that shape it, typically hidden beyond awareness of the scientific community and the consumers of science. A healthy reflection on our interest in sex and race can go a long way in helping us understand those very notions, and provide us with some perspective on our inquiries. Why are we interested? What is our goal when pursuing those studies?

Interestingly, philosophical debates always include some self-reference questioning, for example, about the importance of the debate at issue. So philosophy is plagued with questions about the significance of the very philosophical inquiries we engage in, and about the goals and purposes of our inquiry. Thus, it is far from radical to ask philosophers, when engaging in philosophizing about other people’s identities, to take the time to reflect about the interests guiding their theorizing, and the purposes of their inquiry. This reflection can be humbling, as you might come to realize that what you thought was pure, neutral, and objective theorizing (whatever that means), stems from very specific commitments about what those identities must be; that it is not by natural law that only certain identities are up for questioning, but by our own contingent social norms. That what you thought was a purely innocent and definitely enriching new addition to the marketplace of ideas is actually another way of enforcing, if in new elaborate jargon, the same old ideas about how things should be. Would there be gender wars if our societies had different values about what gender should be? It’s not a war about the facts, it’s about values.

Summing up: I call for a research ethics when philosophizing, especially when philosophizing about other peoples’ identities. First, this does not mean putting a limit to an actual limitless philosophy. Philosophy has limits. It is not, and never was, free of our values and fears, and these set the limits. Second, a research ethics keeps us in touch with those values and fears, and this will enrich our inquiry and make us more responsible researchers. Thus, I propose the following ethical avowal:

when philosophizing about other people’s identities, I will have ethical principles guiding my inquiry, and I will use these principles as I reflect on my interest on the subject matter, the goals of my inquiry (e.g., Why am I interested in this? What do I want to attain with my research?), and the way I proceed in my research. This does not imply that my philosophizing will be spoiled or coerced towards specific positions. It means I will examine my subject matter not from an impossibly abstract and supposedly neutral armchair, but acknowledging and being transparent about the values that are guiding my inquiry in more or less implicit ways.

I commit to philosophize about gender-open children guided by this research ethics.

My child will grow up, and I hope they won’t have to see me engaged in philosophical debates about the authenticity of their gender identity.

4. CONCLUDING REMARKS

Both the theory and practice of gender-open parenting are complicated. Lots of awkwardness in playgrounds and shoe stores, but also lots of wonderful discoveries as you see yourself and others creating new ways of interacting and categorizing. And definitely a very attractive topic for discussion for those of us professionally and personally interested in gender. The well-being of real human individuals is at stake, sometimes clashing against the value of uprooting problematic social patterns. We can tread through it in a sensitive and reasonable manner, keeping in check our biases and self-promoting needs, and embracing the challenge to both our theories and our imagination.
ACKNOWLEDGMENT

Thank you, Nadya Vasilyeva, for this learning and loving adventure. I’m grateful to Loren Freeman for encouraging me to write this piece, and for making that process easier with her comments and suggestions.

NOTES


2. See Fausto-Sterling, Lamarre, and Coll, Sexing the Baby: Part 1—What Do We Really Know about Sex Differentiation in the First Three Years of Life? for a review of relevant research.


4. See, for example, the opinion expressed here by psychotherapist Fran Walfish: https://www.parents.com/parenting/should-you-raise-a-gender-neutral-baby/.


6. Jenkins, “Towards an Account of Gender Identity.”

7. Betchter, “Trans Identities and First-Person Authority.”


9. By having an open mind and utilizing hermeneutical resources to interpret the speaker alternative to the mainstream structurally prejudiced ones, this interlocutor would avoid committing what Dotson calls contributory injustice (Dotson, “A Cautionary Tale: On Limiting Epistemic Oppression”).


14. In relation to this paralysis when meeting a person of whom you cannot guess their sex, Frye writes: “Of course the paralysis does not last. One is rescued by one’s ingenuity and good will; one can invent a way to behave as one says ‘How do you do?’ to a human being. But the habitual ways are not for humans: they are one way for women and another for men” (The Politics of Reality, 20).


REFERENCES


Taking Children’s Autonomy Seriously as a Parent

Quill Rebecca Kukla
KENNEDY INSTITUTE OF ETHICS, GEORGETOWN UNIVERSITY

My research tends to be rather egocentric; I write about what I am experiencing and passionate about in the moment. Fifteen to twenty years ago, my research revolved around the culture of pregnancy and infant care. Now, with my eighteen-year-old child having freshly moved away from home to launch their adult life, I spend a lot of time thinking about the culture of parenting older children as they transition into adult people. I argued in Mass Hysteria (2005) that we labored under a damaging cultural myth that equated good mothering with an erasure of the boundaries between mother and infant. Lately I’ve been thinking about a different kind of agency- and boundary-obliterating parenting mythology. I think that we fundamentally fail to respect the basic autonomy rights that children have as people, and instead that we equate “good parenting” with severe violations of these autonomy rights. We accept and even approve of subjugating children to their parents’ will in ways that would be shocking if enacted against other sorts of people. Here I want to argue—in provisional, overly sweeping terms—for a radical reorientation of how we think about children’s autonomy and parents’ duties to enable and respect it. This essay has a bit of a manifesto-like feel to it, and I recognize that each of my claims requires more argument, and that I’ve left a great number of important details unaddressed. My goal is to display what I take to be something deeply morally wrong with how we picture parenting and its goals, and to offer up a different framework.

My moral starting point is that children are people. Not partial people, or on-their-way-to-being people, but people. Like many people, their competence and skills are limited, and they are dependent on others in various distinctive ways, but this is no compromise of their moral personhood whatsoever. And yet it is routine that we disregard the basic personhood and moral dignity of children in ways that would be clearly unacceptable for almost any other group. For instance, it is common for well-educated, progressive people to say, almost proudly, that they “don’t like children,” which we would simply never say about any other category of people. These same people would be horrified by someone claiming that they just “don’t like Hispanics” or “don’t like lesbians” or whatever it may be. Children are people; how could it be acceptable to dislike them based on their demographic group membership? Moreover, there is good evidence that children’s testimony is routinely discounted, and that they are treated as less reliable reporters than they are, without anyone feeling the need to give evidence for this discounting. We gaslight children by insisting that incomprehensible things such as Santa Claus are real and then lie directly when they question us, and think it’s cute when they believe our lies.
More directly to the point, we restrict where children can go and what they can do even when there is no evidence-based argument for doing so. We make decisions about which hobbies they will have, which friends they can have, which relatives they will hug. Basically, we violate all of the ground-level rules we normally recognize for how to treat people as self-determining agents with dignity.

As people, children’s autonomy rights—which for my purposes broadly include their rights to self-determination, privacy, dignity, and freedom from subjugation—are as strong and as central to their flourishing as anyone else’s. I want to take it as a starting point that, especially given our position of power and children’s vulnerability, one of our core moral tasks as a parent is to enhance and protect our children’s autonomy as much as possible. But in fact, cultural parenting norms push in almost the opposite direction. Not only do we routinely and grotesquely violate children’s self-determination and privacy—restricting their mobility, their choices of activity, their speech, their dress, and much more—but we associate “good parenting” with these kinds of violations, judging other parents who refrain from them as irresponsible or uncaring. What justifies these violations? I think that there are two undergirding ideologies here, both of which are deeply morally wrong.

The first is the idea that children are not full people with normal autonomy rights because they are less competent and more dependent on others for care and support than are (most) adults. But surely if we have learned anything from listening to disability theorists and bioethicists who think about competence, it is that independence is not a requirement for moral personhood; indeed, fluctuating dependency is a part of the human condition. The moral ideal is that we scaffold the agency of those who are dependent, and support those with less than full competence, so that they can express their autonomy as much as possible. Certainly, outside of the domain of childhood, we do not think that we have the right to subjugate, coerce, or gaslight those who are dependent on us or those who need help with competence. Our moral task, as Hilde Lindemann so beautifully puts it in Holding and Letting Go, is to hold others in personhood when they are vulnerable and dependent, not to co-opt their personhood.2

Parents have immense power over children, because children cannot afford basic necessities or protect their safety without us. Many parents think that it is obvious that with this power comes the right to restrict our children’s mobility and their choices, and to subjugate them to our will and vision. But quite the opposite seems right to me. Precisely because children are vulnerable and dependent on us in this way, we should be extra careful not to abuse our power. Just as a dissertation adviser needs to go especially out of her way to protect the autonomy of her student because of the power differential between them, the power imbalance gives parents extra reasons to be careful not to violate their children’s autonomy. Similarly, decent people think that it’s especially demeaning and violating to tell those who depend on government assistance programs how to spend their money or their time, or to subject them to special surveillance. Children’s economic dependence on their parents in no way licenses us using our economic power against them as a tool of lifestyle control. Our children’s dependence on us is something that ought to make us more concerned about protecting and enabling their fragile autonomy, privacy, and dignity, not less.

The second undergirding ideology is that it’s parents’ right, and indeed their responsibility, to create a specific kind of person. Indeed, it is deeply woven into our productivity culture that we treat children as products—as things that we craft—and that we measure our own success by how well we create what we set out to create, regardless of how we got there. Coercion, regulation, and surveillance are acceptable and even expected as tools for creating these products. I think this idea penetrates deeply into our imagination of what reproduction is all about. When we reproduce, we are not just creating a biological entity, but also reproducing families, communities, and generations. It makes sense that we want to create good people who live by values and make contributions that we admire. It is a small step from there to feeling like we have the right to demand that our child turns into a certain kind of person, or to take coercive and manipulative steps to make them be one. When it comes to parenting, the ends are almost always used to justify the means: We almost axiomatically treat a parenting method as “good” and worth emulating if it contributes to creating “successful” people (whatever our standards are for that).

For example, arguments against physically assaulting children (or “corporal punishment,” as it is euphemistically known) generally proceed without pause by arguing that assaulting them makes them turn out less well in the end. It seems to me, though, that the argument against corporal punishment is that assaulting people is wrong. It is completely unnecessary for me to know anything at all about the long-term outcomes from assaulting children in order for me to condemn the morality of the practice. This is an extreme case, but similarly, as a culture we push back (rather gently) against “tiger moms” and “helicopter parents” by arguing that they supposedly produce adults with insufficient grit or creativity or whatever, not by directly saying that controlling other people and invading their privacy is wrong. But children are not products; their quality is not a measure of our parenting success. They are people. We don’t get to use force or manipulation or autonomy restrictions to help determine what kind of people they will be, because doing that to other people is wrong, simpliciter.

In contrast to these two ideologies, I think that we ought to enable and protect children’s autonomy however we can, and to make incursions into it only when absolutely necessary, and then only in ways driven by good-quality reasons and evidence. We are, I think, morally obliged to parent by way of what I think of as a “minimal restriction” and a “minimum invasion” principle, according to which we curtail children’s self-determination and invade their privacy to the minimum extent possible. The two legitimate reasons why as parents we do sometimes need to compromise our children’s autonomy are: (1) when we are directly protecting them from a clear and present danger, because they are not able to keep themselves safe, and (2) when exercising their autonomy would directly and substantially diminish
the autonomy or the well-being of other people, notably including other people with whom they are living and must coordinate and cooperate (we parents included). Of course, when children are very young, these two conditions will kick in a lot, and there will have to be many incursions into their autonomy indeed. A three-year-old is routinely not in a good position to make safe choices about where to go and what to do, and they are typically terrible at cooperating, taking others’ needs into account, and making reasonable demands. All the same, my least-restriction and minimum-invasion principles apply to young children as well; it’s just that the “least” and the “minimum” will be quite high. Though we will often have to compromise their autonomy or privacy, we are not justifiably doing so just because it is more convenient for us, or because we want them to be a certain kind of person, or because we wish they had different desires than they have. For instance, the principles definitively rule out ever making a young child kiss or hug a relative (or anyone else) they don’t want to, regardless of whether it hurts the person’s feelings not to, and regardless of whether it is socially expected that they will. It rules out making them play any game, or have any friend, or pursue any hobby against their will. It rules out forced audience participation and forced birthday party attendance. It definitely rules out forced food consumption, unless there is a present medical danger that necessitates that they ingest something. These are all routine coercions that we enact upon the bodies of young children, which neither protect their safety in an evidence-based way nor protect the substantive autonomy or well-being of other people. That we think that a child will turn out better in the long run if they do these things cannot speak against moral principles as serious as those of respecting the autonomy and privacy of other people.

The situation becomes more interesting, I think, when we turn to older children—pre-teens and teens, who are able to go places on their own, and can form sophisticated goals, relationships, tastes, and preferences, and who understand the basic rules of how to stay safe, and the basic principles of what is and isn’t a viable plan (that is, they are unlikely to demand that you turn them into a dinosaur). At this point, it seems to me, coercively restricting, monitoring, and controlling children beyond what is needed to keep them safe and to ensure that they are not harming others is both extraordinarily common, and wrong. People should have basic rights of mobility and of self-determination when it comes to where they go; which avocational skills they develop; how they dress; who they spend their time with; what they play, watch, and read; what they touch and how; what they eat; and more. These rights are core components of people’s autonomy and privacy.

It is parents’ responsibility to teach their children early about how to stay safe, how to reason about risk, and generally how to make good choices, through discussion and example. But once a child has these competencies, it is a violation of their privacy and autonomy to monitor and restrict their movements and relations. For example, once children are competent at finding their way home and at using a phone to get help if needed, it is, I believe, none of parents’ business where they go or who they spend their time with, when they are not at school or meeting family responsibilities. Parents widely and unquestioningly assume that they have the right to this information, but I simply don’t understand what the basis of that right could be. Controlling and surveilling people’s motion and relationships is a deep form of subjugation, and children are people. Some parents will object that it would be awful to be completely cut off from their child’s life in this way. But this betrays the assumption that your child will talk to you about their life and their day only under coercive threat, and if that is true, then parenting has already derailed. Normal, free people who live together and are intimately connected voluntarily keep one another reasonably informed about their lives, but they don’t get to coercively control or monitor one another’s comings and goings.

Similarly, and I am sure contentiously, I insist that once a parent is reasonably sure that their child has a good understanding of sexual safety and consent and the importance of each, it is none of parents’ business when or with whom their child decides to have sex. Precisely because we are responsible for enabling and protecting their autonomy, we are responsible for making sure they have this understanding as soon and as thoroughly as possible. But once they do, our role in their sexual decision-making is over. People’s sexual lives and preferences are deeply private, and controlling or insisting upon surveilling someone’s sexual choices is seriously invasive, and children are people.

I have acknowledged that parents can impose restrictions on adolescent and preadolescent children if they have good reasons to think that these restrictions are required to keep them safe, or if the restrictions are needed so that no one else is seriously harmed or restricted. In contrast, one factor that does not count as a morally legitimate reason to impose a restriction is that it will make the parent feel better or more comfortable, albeit not for any evidence-based reason. For example, many parents do not allow their children—especially girls—to be out alone after dark, because of fears of kidnapping or rape. But street kidnapping is an extraordinarily rare crime, and stranger rape is rare as well; in the United States, girls are at dramatically less risk of assault on the street than they are at school or at home. But we’ve created a moral panic that has served the specific purpose of restricting the mobility of girls and women. Parents who have internalized this panic, when confronted with the actual statistics, will often tell me that they know I am right about the objective risks, but that they will just “feel more comfortable” if their daughter isn’t out alone, perhaps because they don’t want to “imagine” what might happen. It seems to me like a paradigmatic example of a morally unjustified autonomy violation to restrict and control another person’s movement through public space for the sake of one’s own (admittedly irrational) psychological comfort.

To a large extent, one person’s authority to restrict and surveil the basic mobility of another person always relies on implicit threats and fear. I think that by the time my child was in middle school, I had lost—or, better, divested myself of—the authority to tell them where to be, even if I had wanted to. If I had told them that they were grounded,
they would have just calmly disagreed. My child knew that I would never hurt them or stop supporting or protecting them even if they did something I did not want them to do, so there was no implicit threat to back up such a ruling on my part, and I am glad of that; this was an important sign of my child’s trust in me.

Of course we do, in fact, have a legitimate interest in our children turning out to be good, happy people who contribute to the world in interesting and worthwhile ways. Many people also have an interest in raising children who will share their culture, be part of their community, and carry on their traditions. We raise our children in specific communities, surrounded by specific ways of life, modeling specific values and engaging in specific traditions. In the normal course of events, children will grow up at least somewhat identified with and shaped by these experiences. It is not only morally permissible but generally laudatory to expose our children to our values, inviting them to participate in our ways of life, and using rational persuasion to get them to make choices we think they should make as best we can. These are ways of showing respect for them as people and valued community members.

But it is always possible that our children, given the freedom to do so, will reject part or all of their communities of origin and their values. (Of course, we run the risk that they will reject values and ways of life we care about eventually even if we do coerce them into conforming while they are children.) They may do this for reasons of personal taste, or because they have genuinely good reasons to critique these communities (for their homophobia, ethnocentrism, sex-shaming, or whatever it may be). This is just a risk we are stuck with accepting, as parents. One cannot morally impose an identity on someone by compulsion. In general, we find it morally abhorrent to keep people in a community and a way of life by force, and I see no reason why children should be an exception. We need to shape our children into people we are proud of, to the extent we can, via engagement, persuasion, collaboration, and presenting them with a way of life that is valuable and fulfilling, not by instituting a totalitarian regime in the home.

Children differ widely in temperament, needs, skills development, and risk reasoning. The minimum necessary restrictions and incursions required by one child in one context may be very different from those required for another child in another context. But none of that variation defeats the fundamental moral principle for which I am urging here: Since children are people, it is our moral duty to restrict their autonomy and invade their privacy as little as possible. Just as we’ve built it into law that children are owed an education in a “least restrictive environment,” regardless of their needs and abilities, they have a moral right to no less in the home. Neither their dependence on us, nor their nascent competence, nor our power over them, nor our desire to “produce” a certain kind of person can undercut that duty.

NOTES
2. Lindemann, Holding and Letting Go: The Social Practice of Personal Identities.
3. Notice that I am only talking about parents’ incursions into their children’s autonomy and privacy; my focus on this paper is entirely on parenting. It is, of course, also true that children, like all people, are subject to laws and restrictions imposed by the state. I could write a separate paper on whether our laws regulating and restricting what children do are just and legitimate, but that issue is not my topic here.
4. I do think there’s some complexity here if a child wants to consume something that is hateful, like a racist or misogynist game, because it is reasonable not to want actively hateful material in one’s home. I’m sure exactly how to negotiate the line here. I always erred on the side of not restricting or disallowing anything and instead being very vocal in my critiques and encouraging a lot of discussion about why I objected to the thing and why my child enjoyed the thing. But I feel like my tolerance limits could have been reached if my child had picked up awful enough material.
5. Certainly, fathers don’t get to play the creepy culturally sanctioned game of delaying their daughter’s dating altogether via threats of violence against potential dates.

REFERENCES

Parenting in Trauma
Melissa Burchard
UNIVERSITY OF NORTH CAROLINA, ASHEVILLE

The call for papers for this issue of the APA Newsletter on Feminism and Philosophy includes the hope that discussion on the many facets of parenting will be “important for many philosophers who have and who are thinking about having children.” I hope this too, and agree that it is likely that more discussion on parenting and childrearing is important for those engaged in these activities. However, I want to make a more ambitious claim than that—I want to say that serious discussion and engagement with issues of parenting and childrearing is important for philosophy, if we want to continue to assert that philosophy is important for understanding reality, the world we live in, and our experience. This is not a wholly original statement, of course, as a wealth of scholarly, philosophical engagement (including my own) with these topics over the last several decades shows. But my contention is perhaps more pointed than that the discussion is important: what I am saying is that for the sake of philosophy’s claims to be able to explain the world, philosophy needs to spend more time and effort thinking about the work of parenting, and understanding the development of children.

I have arrived at this belief through a very difficult experience of childrearing. About fifteen years ago, my partner and I adopted two children from foster care, and over about the first year and a half with them, we began to understand just how much we did not know about the kind of parenting that I would never hurt them or stop supporting or protecting them even if they did something I did not want them to do, so there was no implicit threat to back up such a ruling on my part, and I am glad of that; this was an important sign of my child’s trust in me.

Of course we do, in fact, have a legitimate interest in our children turning out to be good, happy people who contribute to the world in interesting and worthwhile ways. Many people also have an interest in raising children who will share their culture, be part of their community, and carry on their traditions. We raise our children in specific communities, surrounded by specific ways of life, modeling specific values and engaging in specific traditions. In the normal course of events, children will grow up at least somewhat identified with and shaped by these experiences. It is not only morally permissible but generally laudatory to expose our children to our values, inviting them to participate in our ways of life, and using rational persuasion to get them to make choices we think they should make as best we can. These are ways of showing respect for them as people and valued community members.

But it is always possible that our children, given the freedom to do so, will reject part or all of their communities of origin and their values. (Of course, we run the risk that they will reject values and ways of life we care about eventually even if we do coerce them into conforming while they are children.) They may do this for reasons of personal taste, or because they have genuinely good reasons to critique these communities (for their homophobia, ethnocentrism, sex-shaming, or whatever it may be). This is just a risk we are stuck with accepting, as parents. One cannot morally impose an identity on someone by compulsion. In general, we find it morally abhorrent to keep people in a community and a way of life by force, and I see no reason why children should be an exception. We need to shape our children into people we are proud of, to the extent we can, via engagement, persuasion, collaboration, and presenting them with a way of life that is valuable and fulfilling, not by instituting a totalitarian regime in the home.

Children differ widely in temperament, needs, skills development, and risk reasoning. The minimum necessary restrictions and incursions required by one child in one context may be very different from those required for another child in another context. But none of that variation defeats the fundamental moral principle for which I am urging here: Since children are people, it is our moral duty to restrict their autonomy and invade their privacy as little as possible. Just as we’ve built it into law that children are owed an education in a “least restrictive environment,” regardless of their needs and abilities, they have a moral right to no less in the home. Neither their dependence on us, nor their nascent competence, nor our power over them, nor our desire to “produce” a certain kind of person can undercut that duty.

NOTES
2. Lindemann, Holding and Letting Go: The Social Practice of Personal Identities.
3. Notice that I am only talking about parents’ incursions into their children’s autonomy and privacy; my focus on this paper is entirely on parenting. It is, of course, also true that children, like all people, are subject to laws and restrictions imposed by the state. I could write a separate paper on whether our laws regulating and restricting what children do are just and legitimate, but that issue is not my topic here.
4. I do think there’s some complexity here if a child wants to consume something that is hateful, like a racist or misogynist game, because it is reasonable to not want actively hateful material in one’s house. I’m not sure exactly how to negotiate the line here. I always erred on the side of not restricting or disallowing anything and instead being very vocal in my critiques and encouraging a lot of discussion about why I objected to the thing and why my child enjoyed the thing. But I feel like my tolerance limits could have been reached if my child had picked up awful enough material.
5. Certainly, fathers don’t get to play the creepy culturally sanctioned game of delaying their daughter’s dating altogether via threats of violence against potential dates.

REFERENCES
and abuse they had experienced in their birth home and in the foster placements they had been in subsequent to being removed from that birth home. Although we do not know with certainty what that experience was, we know with significant clarity what many of the effects are, and they are severe. My experience in parenting, then, is the specific experience of parenting children who have been traumatized in multiple ways over an extended period of time. They were traumatized through the neglect of birth parents who were drug addicted, thus often under the influence of those drugs (and the evidence is clear that neglect has significant effects on infants and children), as well as through domestic violence in their birth home and direct physical and sexual abuse. They were nine and four years old when they came to live with us, which means that their formative years were spent in trauma-producing circumstances, such that whatever sense of self they had at that point had been, as trauma therapist Judith Herman puts it, “formed in and through trauma.”

As a parent, such an experience is generally not one that is hoped for. Although we had had extensive training through our adoption agency, and we understood (intellectually, at least) that the likelihood was very high that any children coming through the foster care system would have some level of traumatic effects to work through, we were still at times overwhelmed by the level and kinds of difficulties inherent in the task of parenting children with severe developmental trauma, as it is coming to be called. As a parent, I needed to do everything I could to try to help my wonderful, beloved, undeservedly harmed children recover. As a philosopher, my sense of needing to understand drove me to immerse myself in the study of trauma and trauma therapies. What I realized I needed, as a philosopher-mother, was a philosophy of trauma; what I discovered, of course, was that there was very little to be had. Although quite a lot was to be found in cultural studies, literary studies, and psychology, of course, only a very few brave pioneers in philosophy, such as Susan Brison, had published work on trauma. Under the impulse of necessity, I began to cobble together what I could of a philosophy of trauma, especially with regard to the developmental trauma of children who have been traumatized under circumstances like my children had: the prolonged and chronic trauma of having to live with the abusers; the betrayal trauma of being harmed by those who have the responsibility of your care and are supposed to love you; the loss of being forcibly removed from the only home you have ever known, together with the contradictory relief of escaping from that abusive home. In hopes that it will prove useful to other parents who might find themselves in somewhat similar circumstances (which, of course, can be created in many ways; trauma is not only produced by abusive situations but also by, for example, divorce, accidents, natural disasters, and oppressions), and in hopes that philosophers more generally will begin to realize that we need to understand these things, I offer some of what I believe are implications of trauma for parenting and philosophical understanding.

**TRAUMA AFFECTS BASIC INTERACTIONS**

If we didn’t have other ways of knowing this, new developments and trends in neuroscience and brain imaging would still provide us with overwhelming evidence that trauma affects the development of the brain in infants and young children and the working of the brain in older children and adults. One of the major effects of this in infants and young children is that trauma changes their ability to form the basic attachments that are required for the creation of what is referred to as an internal model of safe relationship. The optimal kind of attachment ability, which is called secure attachment, is created through the experience of being attended to and cared for appropriately as an infant and very young child. In simple terms, infants get hungry, or wet, or frightened, so they cry. If all is going well, a familiar caretaker (usually a primary caretaker or one of a small set of caretakers) appears to comfort them and provide whatever it is that they need. In good circumstances, this happens over and over again, reliably, throughout the child’s formative years, and the result of this is that the child learns to trust, and to believe that their world is relatively safe. That is, this experience creates for the child an internal image of relationship in which their needs are reliably met, and they are consistently cared for appropriately, with stability and affection. This internal model then remains active throughout their lives (unless it is disrupted), allowing them to form new relationships in the expectation that each relationship will work in the same way, and with a sense of appropriate interactions that enables them to reject relationships which do not in fact live up to the expectation.

In ordinary development, they learn that they also should live up to such an expectation for others.

Traumatic experience during the formative years makes it impossible for infants and young children to develop and/or sustain such a model of interaction, because, obviously, it is not the interaction they are having. Neglect creates traumatic effects precisely for this reason: it means that no one is in fact responding to the needs of the infant or child, no one is consistently appearing to comfort and care for them. The development of the above described internal model is not possible, then, and the child develops an insecure model of attachment, which affects their ability to form relationships throughout their life if no intervention occurs. Severe neglect may lead to conditions in which a child fails to thrive (a technical term in pediatrics) in which case they suffer malnutrition and attendant problems, and which if extreme, leads to death.

Here, then, is an implication of trauma that philosophy needs to take seriously: the kinds and quality of interactions that we have with others is essential to the development of a sense of self. For most feminist philosophers, I take it that this will come as no surprise; we have been working for decades with the belief that humans are interdependent, relationally autonomous creatures rather than the separate and totally (almost ferociously) autonomous beings of liberal individualism. But here is a different kind of evidence that seems to me utterly undeniable: if infants and children do not get the proper kinds of interactions, they do not develop certain relational abilities, which may prevent them from forming satisfying, mutually trusting relationships throughout their lives. That is, one’s ability to function as a successfully (relationally) autonomous individual depends on one’s having the basic interactions of stable, consistent caretaking in one’s infancy and early childhood. Having a “good” sense of self depends, ordinarily, on having these experiences.
I do not mean to say that one absolutely cannot create a good sense of self if one was traumatized as a child, but the level of difficulty in doing so is very high, and it is a different project from that early developmental one. Our children, traumatized through neglect and through direct abuse, came to us with “bad” senses of self. Their senses of self were bad in that they believed they were bad children, having been told they were, but also in that their senses of self were incapable of doing the kinds of thing that a good sense of self can do. They could not form healthy, secure attachments, but rather sought to get their needs met mainly through deception and manipulation, as they had had to do in their birth home. They could not trust anyone because no one in their important relationship schemas had been consistently trustworthy. They could not be friends or accept ordinary friendship because they had been taught that they were for the use of others, and that others, in turn, were for their use. And it is important to realize that these were not casual lessons that could easily be set aside once they were presented with better alternatives; these were formative lessons, part of their identities, and as such very resistant to change, even if they could come to see that those changes would be for the better.

**TRAUMA AFFECTS ABILITY TO PLAY**

Given the above effects on interactions, it seems obvious that play would be affected as well, although I confess it wasn’t obvious to us and we were sometimes baffled, at first, by our children’s playing and lack of playing. Some games seemed perfectly ordinary. A perennial favorite, the “chase me” game, was one we played endlessly (or at least it seemed endless to me, as running was never my strong suit!). They knew many games already, and we taught them many of our favorite games, both indoor and outdoor: board games indoors, frisbee, badminton, ring toss, bocce, and croquet for outdoors.

We realized fairly quickly, though, that the instructions “go play x” were never sufficient, even if they knew the game. Either they did not want to play that, or it would last for about five minutes and then they would be back asking what they could do. We also began to observe that they did not engage in imaginative play—they would not, on their own, create a game or a scenario for playing in. Even with toys like Legos, which tend to enable imaginative play as kids put them together in a variety of ways and make up things to use them for, our children did not play the same way as others. If we said “make something up,” it was as though they just did not understand the project.

But that was odd because they seemed to have sufficient imagination in some ways. Ask them to talk about monsters, for example, and they had plenty of things to say, and plenty of ideas about how to fight them and how you needed to protect yourself from them. Gradually, through observing them, playing with them, and learning about traumatic effects, we came to understand a couple of things about post-trauma playing.

One thing we came to realize was that they did not like to play without our direct participation. One reason for this might have been that they wanted our attention, and that is certainly likely, as they did want our attention. But it is also probable that their lack of security and inability to trust made it difficult for them to play without our “help,” because they were not secure about whether they were doing it “right.” If we were there, guiding the interactions, then they could be relatively sure it was going ok, and that they were not going to do something wrong and get into trouble for it, be punished for it. That is, any adopted kids (if they were adopted as older than infants) are likely to have a fairly long period of adjustment in a new family, in which they have to learn how to “do things right” in that family; this includes learning how to play properly. They are likely to be anxious until they have figured out what playing properly is like.

For traumatized kids, this anxiety is exponentially greater because they have been in unstable circumstances in which punishments and outbursts of anger could not be either reliably predicted or understood (and certainly not avoided). One of the common effects of trauma, hyperarousal, refers to the survivor’s need to remain on “high alert” to watch for the signs of impending danger at all times. This survival strategy is necessary in order to allow the survivor to bring to bear any and all possible counter-tactics if danger seems imminent. For example, being able to detect early signs of an impending episode of violence makes it possible to try distracting or soothing tactics in hopes of preventing or mitigating the outburst. Consider, then, what it might be like to try to play while maintaining the constant awareness of the environment and especially the moods and emotional indicators of the parents or other caretakers/adults, the ones who are most likely to be the source of such danger. If “getting it right” means avoiding a beating, or being allowed to eat, then a significant amount of a child’s attention and mental and emotional energies must be directed at getting it right. In the absence of certainty (or as much as can be had) of getting it right, it may seem better to the child simply not to take a chance; that is, it may be better not to play at all.

So here is another implication for philosophy: even something like play cannot be taken for granted as an obvious, universal feature of experience; styles of play actually can tell us something about the people who engage in it. Again, this is not a totally new observation. María Lugones, for example, wrote about this in her well-known article on “world-traveling” and playfulness. What it is like to be playful is not simply the same for everyone; it depends on one’s socialization, one’s ability to feel comfortable and “at home” in one’s circumstances, one’s gender training, and, I would add, one’s experiences with trauma.

**TRAUMA AFFECTS SENSE OF REALITY**

One of the most dramatic realizations we had as we came to understand more about our children was that in some incredibly important ways they were not living in the same world, the same reality, that we were. In our world, parents are people who provide for the needs of children; they make sure that there is enough food (hopefully of healthy types), appropriate and clean clothing, and a home in which children are safe (except, of course, for the occurrence of ordinary pains or injuries). When anything bad happens, parents are not the ones who cause it, but
are the ones who provide comfort and remedy. Parents impose consequences for choices, but do not arbitrarily or unreasonably or unpredictably punish, or punish in hurtful or harmful ways.

This is perhaps an idealized picture, but again, we had had professional training and we had professional support in constructing our parenting and ourselves as parents, such that we were able to structure our parenting pretty intentionally right from the beginning. We were also older parents, and had taken a lot of time to think about what we believed parenting should be/do. And it is certainly not the case that we never made mistakes, or bad choices, with regard to parenting—I wish! But ultimately, in our world, parents were not people that children needed to be afraid of.

In our children’s reality, and the reality of so many children, parents are exactly the people to be afraid of, as well as the people who must be loved and depended on. When parents abuse children, they create a reality in which what children learn is that some people have the power to do to you whatever they want whenever they want; and you cannot escape it or prevent it (although you may learn to manipulate it or postpone it). What we had to learn to understand was that this was not simply their “picture” of the world; this was what was real, and they understood it to carry over into our home, the new home we had created for them. To put it differently, for children abused by their own parents, the monster in the room at night, under the bed, or in the closet, is absolutely, terrifyingly real, and it is also a person who has responsibility for their care. It is a person/monster, a creature with a dual nature, which may be loving at one moment and terrifying the next, inflicting pain and humiliation in the night and dropping them off at school the next morning, cheerfully waving at the teacher.

This also gives the world a dual nature, as abusers will also require that that part of the reality be kept secret. So abused children have to function in a secret reality as well as the “normal” one, in which no one knows (or admits to knowing) what is actually happening in the other. The requirement to repress that secret reality, coming from the abuser(s), is reinforced by the child’s own “internal necessities,” so to speak. That is, if a child is in an abusive situation they cannot escape from, in which they have to continue living with their abuser on a long-term basis, their own psychology will compel them to “compartmentalize” the abuse in ways that allow them to hide it, at least part of the time, from themselves as well as others. It is psychologically and emotionally extremely difficult for a child to live with the actual knowledge of ongoing abuse; Roland Summit, developer of the Child Sexual Abuse Accommodation Syndrome model, says that the most healthy reaction a child can have to ongoing abuse is to block their own knowledge of it, as that knowing would be too difficult to live with. In extreme cases, having to live in a dual reality compels dissociation; there is evidence that the majority of cases of multiple personality disorder stem from childhood sexual abuse at the hands of parents or other very close family members.

Learning to understand the traumatic double reality of abused children is a staggering task, as it challenges a parent’s own conceptions of the real and forces us to expand those conceptions in order to be able to accept and respond to our children’s pain, injuries, and needs. In some ways, the most difficult task of parenting traumatized children, especially those who have been traumatized in ways that are taken to be so forbidden that they are nearly unthinkable, is simply to accept that it has happened. So one implication of trauma for parents, and philosophers, is that it requires us to open ourselves to a reality that we do not want to admit; we do not want to admit that it is true, and we do not want to admit it into our lives. But the realities that abused children are experiencing must be admitted, must be validated, if anything about their lives is to change. If—when—the unimaginable does happen to children (or to adults), we must learn to believe in its reality. Perhaps we must say something even stronger: perhaps we must learn that the “unimaginable” is already the real, and it is our ideals about reality (that it is safe, that children are never hurt; that no one would do such things) that are, in fact, the not-real.

This discussion also has implications, I believe, for social and political (and moral) philosophy because it can help us articulate the wrongs of oppressions or oppressive actions. I take it to be uncontroversial that the abuse of children, in general and in particular by parents, is morally wrong, and that the wrongness of it is at least partially laid out in the foregoing discussion of how it requires the distortion of reality (only one feature of that wrongness, of course). If this is the case, we should be able to draw analogies with other forms of abusive control, such as the oppression and subjugation of specific groups of people under racism, sexism, heterosexism, etc. Consider, for example, the phenomenon of “double consciousness” articulated in the work of W.E.B. du Bois and used since then by many other critical theorists. Du Bois talks about the experience of coming to know oneself as black in a world ruled by anti-black racism as creating the sense of being forced to live a doubled life, with doubled thoughts and double duties, and presents this as harmful, as a “wrenching of the soul.” He names among the effects of this doubling “a painful self-consciousness, an almost morbid sense of personality, and a moral hesitancy which is fatal to self-confidence.” These effects are very much what we see in the traumatized behavior of many who were abused as children, usually in an even stronger form. What that suggests is that we are looking at similarly abusive phenomena in various forms of oppression. This could gain us some added force for the condemnation of such phenomena, and although it would be wonderful to be able to say that we don’t need any such reinforcement because we are already in the post-race(ist), post-sex(ist) era, it seems to me that we have far too much evidence to the contrary. Thus I welcome any arguments that enable a more emphatic condemnation of oppression and oppressive practices.

**TRAUMA AFFECTS UNDERSTANDING OF “THE GOOD LIFE”**

Although conceptions of “the good life” vary significantly from culture to culture, and even from person to person, what parents tend to hope for their children include things like finding ways of making life meaningful, having
enough at the very least (whatever that means), having relationships that are supportive and fulfilling, and being able and willing to act in ways that are morally responsible and permissible. This conception of a good life, however basic, may be challenging when children have been severely traumatized, especially if that trauma has affected or disrupted their developmental processes.

For one thing, a very common effect of trauma, at least in Western cultures in which time is understood as a linear path from distant past through present and into distant possible future, is referred to as the “foreshortening of the future.” What this means is that traumatized (Western) children’s picture of the future may be no more than months or at most a couple of years, and they will have difficulty imagining themselves surviving/living to adulthood. Consequently, conceptualizing and planning for a good life, which in this culture seems to depend on being able to imagine a future, will simply be something they are not able to do and may not even be able to grasp the meaning of.

More than that, as important recent work in the neurosciences has shown, trauma affects the development of the brain, its neural pathways, and the connections among the different brain areas. What this means, for example, is that a traumatized child may have a much more difficult time than others in developing abilities like recognition of the relationship between causes and effects, or what are often thought of as the “intuitive” fundamentals of logic (another reason, in my judgment, to reject the claim that Western logic is intuitive, but that is another paper). This can be incredibly frustrating for parents, as these abilities are often taken to be simply “natural” and thus not considered to be something that is at stake when things go wrong. It may mean that no matter how many times you patiently (or not so patiently) go over the idea that if they do X, consequence Y will follow, and they do not want to have consequence Y, they will still do X and not be expecting or looking for the consequence. It may mean that they have a much more difficult time learning certain basics at school, both in terms of content they need to master, but also in terms of behavior: no matter how we struggled to help our children with homework, for example, once it was actually done, they were just as likely to leave it in their backpacks as to turn it in on time, because they didn’t really see that as part of the exercise, and didn’t recognize bad grades at the end of the semester as actually connected to anything they did with their homework months before. They simply couldn’t see that it mattered.

These are frustrating problems for both parents and children, but the effects of inability to plan, to decide on a good and reasonable goal and then plan out the steps needed to get to the goal (and how to achieve each of those steps), seems to me only to become a bigger problem as children begin to grow to adulthood. As adults, we need to be able to manage our lives, and being able to plan and move through our plans in appropriate steps seems fundamental to that management. Living independently was a goal that was not difficult to “sell” to our children (they could hardly wait), but the idea that you have to have a plan to make it work continues to be much harder. Being able to get and keep a job is basic to this kind of planning, but for some who have been traumatized, there may be nothing straightforward or easy about doing so. Keeping a job requires recognizing that there are direct consequences, usually negative, for not showing for a shift, or for blowing up at the boss, or for walking out because some fellow worker was giving you bullshit.

Of course, beyond recognizing that there will be consequences, keeping the job requires that you be able to care about the consequences as well. Again, for some who have been traumatized, this may be difficult to do. Being traumatized in circumstances of abuse, for instance, means that you have been violated and humiliated, terrorized and made to feel utterly helpless, utterly unable to assert yourself as an effective agent in any way at all. This experience may make it seem paramount to a survivor that they work to assert control in all circumstances that they can, even when doing so is actually going to work against their own interests in either the short or long term. That experience of helplessness can make it so that being in control of a situation, or of what consequences one will suffer, may seem more important than suffering any kind of consequences. This is one explanation of why some traumatized persons will work to “sabotage” their own interests. For example, when our children were still young, we learned that it was better to keep secret from them certain possibilities for good activities. We stopped telling them when a big trip was planned because their insecurities drove them to feel so anxious about whether they would do something bad and “lose” the opportunity to participate that they would inevitably go ahead and do the bad thing, just so that they could control when it was going to happen. They lost that opportunity, but they relieved their anxiety about it, and that seemed to be more important. Again, that was partly about the problems of the future: any goal more than a few days out was too far away to be as important as a present anxiety.

What implications does this discussion have for philosophy? One is to remind philosophers that we tend to take for granted a certain kind of normative experience of the world, and forget to include in our analyses those experiences that are not normative, but that are statistically significant, as experiences of trauma clearly are. Any parents whose children are non-normative in almost any way know, for example, how our system of education is constructed to work best for those with a certain range of abilities and life experience, and how that system has very little room or help for “others,” often in spite of the efforts of caring teachers. In fact, in our roles as teachers, philosophers need to remember that the tendency to take a normative experience for granted affects our own ability to teach. As much as any teachers, we need to become much more aware of the needs of diverse learning styles, and to begin recognizing our responsibility to build into our classrooms more of the tools of, for example, trauma-informed pedagogies.

On a larger scale, if philosophy were taking seriously the experiences of the seriously traumatized, it would be able to render much more helpful analyses of, for example, the behaviors of other groups of “others.” Having had many such discussions with people unfamiliar with our children’s
experiences, I know that it is easy for many people to dismiss their difficulties with platitudes such as “that’s a problem that every kid has to struggle with” or even harsher “observations” such as “maybe they just need to learn not to be so lazy.” That makes me think about other groups to whom such “observations” have traditionally been attached, groups that tend to be marginalized on the basis of class, race, ethnicity, gender, and/or dis/ability status, etc. But more understanding of trauma is leading experts in the field to theorize that oppressions, for example, are traumatizing. Is that, then, what we are often seeing when we look at members of marginalized groups and their struggles to “manage” their lives, to “live up to” the norms of the dominant culture? It makes far more sense to me to believe that oppressed groups struggle with the effects of trauma than that any group of people can truly be characterized as simply “stupid” or “lazy.” Philosophy could be applying its strengths here and helping instead of allowing injustices to continue.

TRAUMA MAY AFFECT DESIRE
I noted in the last section that one effect of trauma may be that it can create in a survivor a desire to be in control that overrides any better kind of judgment about what to do in some circumstances. That is not the only way in which trauma can affect desire, and may not be the most problematic. I have discussed elsewhere at some length the ways that desire can be affected but will repeat some of it here because it is so important for parents to realize that depending on what kinds of trauma children have suffered, their desires may become warped in ways that can deeply affect their relationships with adults as well as with other children.

I am speaking, of course, especially of sexual abuse of children. As noted above, when children are abused during their formative years, their developmental work is affected by trauma such that they form a sense of self through that trauma and its effects. One thing this may mean for sexually abused children, especially those abused over a prolonged time and/or by caretakers or close family members, is that they can come to understand sex as a “normal” part of relationship in general. If their basic relationships are sexualized, they may simply grow to understand that sex is how relationships are. If this is the case, they may then expect sex in relationships in general, and may attempt to sexualize any relationships that they enter into, with either other children or with adults.

If such a child is adopted, there is every likelihood that they would attempt to reestablish patterns of relation that they are familiar with in their new family situation; this is often referred to as “acting out sexually” and sometimes diagnosed as “sexual reactivity.” But what new parents need to understand is that they have to be able to respond to it not with disgust or revulsion but with understanding and tactics of support and correction. Normatively trained adults tend to find it extremely disturbing, or even downright unbelievable, that children, including very young children of three or four, are capable of having sexual feelings and expressing them toward others in direct and unmistakable ways. What those children will need from their new parents, however, is acceptance and commitment to the work that will be required to help them move from inappropriate sexual expression to control over their very real sexual feelings and urges, in order to grow into a more appropriate stance toward those desires.

It is, presumably, obvious how important this is to the life of an abused child. If inappropriate sexual desires and activity are ignored and allowed to continue, the child may never learn to moderate those desires on their own, and will be likely to be frustrated by failed relationships throughout their life. Not understanding that sex is not a part of every kind of close relationship, they may attempt to impose sex on others (although it is also important to realize that not everyone who is sexually abused as a child abuses others as an adult), or drive others to reject them because of their tendency to inappropriately sexualize relationships. They are also more likely than others to be sexually victimized, as their “air” of sexual awareness tends to draw the attention of sexual predators, and they may not see their danger until it is too late.

One implication for philosophy regards the so-called liberal sexual paradigm, in which the tendency is to assume that anything done between consenting adults is fine, and which in some articulations, like that of Thomas Nagel, claims that “bad sex is generally better than none at all.” Again, because mainstream philosophy tends to assume a certain kind of normative experience, including a level of autonomy that may actually be unusual, the philosophy of sex has tended to assume both that consent is simply, obviously possible, and that once it has been “given” that no other worrisome questions remain to be asked. From my perspective, the fact of childhood sexual abuse and its effects means that consent in matters of sex is probably often nowhere near as free or as meaningful as it is taken to be. For those survivors who are helped to successfully recover from much of their experiences, it may be the case that they are capable of free and meaningful consent. But some do not recover, or do not recover fully or in every aspect, and a forceful demand for consent (which, let’s not kid ourselves, is too often how “consent” is manufactured) may easily defeat their autonomy. Further, given that we know that childhood sexual abuse is among the most underreported of crimes, we must also recognize that many survivors may have had little to no support or help in recovering at all, such that as adults they are still to some greater or lesser extent affected by the evils of that experience. Among these may well be that they have little of what it takes to stand up for oneself in the face of forceful, or “impassioned,” requests/demands for consent. We can take these implications of trauma seriously to indicate the need to revise, or at least complicate and problematize, the liberal model of sexual ethics in favor of one that requires a heightened responsibility to take care that our sexual actions do no harm, instead of just assuming that they cannot because we do not wish to believe that they do, as such a belief would require constraining them. The work of people like Catherine MacKinnon, one of the first to point out the insufficiency of consent for ethical sex, and more recently Rebecca Kukla, who is developing a much more comprehensive model of consent within a framework of broader sexual communication, is definitely moving in the right direction.
**FINAL WORDS**

One final implication I will mention here is that feminist philosophers and others have been urging for philosophy as a field for a long time now, and that is the importance of taking personal lives and personal experiences seriously, as well as learning how to properly value the lives and experiences of non-normative “others.” I do not believe that I would be able to articulate these reasons for philosophy to take trauma seriously if I had not had so much personal experience with it. My own experiences with trauma started me down my non-traditional philosophical path, but living with and through the experiences of my children made it impossible for me to ignore these implications for philosophy. How could I claim to understand the world if I did not understand the world my children were literally living in and held captive by? How could I claim to know the truth about who they were, and who I am in relation to them, if I did not know something at least of the truth of their experience? How could I think that I understood what is the good life for my children when I did not understand the extent to which the options for life had shrunk for them due to their trauma? Although I realize I have made far too many mistakes and have so much more to learn, I have at least begun to see how as a philosopher I cannot ignore marginalized experiences and still claim to be anything like an authority on the real, on knowing, on the good. Philosophy, and philosophers, need to learn how to believe in the realities of “others” in order to more properly understand reality in general, and to be able to do our own proper work.

**NOTES**


3. Herman, *Trauma and Recovery*, 96.


6. See, for example, Herman, *Trauma and Recovery*; Bowlby, *A Secure Base*.

7. Block, et al., “Failure to Thrive as a Manifestation of Child Neglect.”


10. Summit, quoted in Corwin, “An Interview with Roland Summit.”

11. Herman, *Trauma and Recovery*, 120–26; for a compelling example, see Fraser, *My Father’s House*.


14. Ibid.

15. Herman, *Trauma and Recovery*, 47.

16. Of course, another huge reason why traumatized children have trouble with schoolwork is simply that the effects of trauma, including hyperarousal, keep their attention focused on trying to figure out where the next threat is coming from. This can leave little attention or energy for paying attention to actual schoolwork, which obviously is much less important from their perspective than the need to be aware of where trouble is going to come from next.
On Muddling Through

Amy Allen
PENNSYLVANIA STATE UNIVERSITY

When Lauren Freeman approached me to write about my experiences combining parenthood and philosophy for this newsletter, I hesitated. Her invitation asked me specifically to talk about succeeding in academic philosophy while being a mother of four. I was reluctant in part because I worried that there's no way to write a piece like this without sounding like someone who thinks of herself as a success, either as a parent or a philosopher or both—which not only makes me sound like a jerk, but also seems like tempting fate. I changed my mind only when Lauren told me the horror stories she had heard from other senior women in the field: women who claimed, as part of an invited panel on "succeeding as a mother in the field" at a conference for women in philosophy, that there was no way to combine motherhood with a successful career in academia, that they regretted having children before they had tenure, or even at all. Lauren persuaded me that it is important to do so to tell a more positive story, to help others see what is possible, so I (still rather tentatively) agreed.

My misgivings also stemmed from the fact that I am all too aware of the very many privileges that I had and still have, all of which enabled me to combine motherhood with philosophy to whatever degree of success I have attained in either domain. I was fortunate enough to land a tenure-track job relatively quickly—though, it must be said, not right away, and my first child was already almost three years old when that happened—and when I did, the job came with a scandalously low teaching load and a generous parental leave policy. While I was on the tenure track, I also had a very supportive female associate dean (shout out to Lenore Grenoble, wherever she is!). She helped me at a crucial pre-tenure moment, when I was pregnant with my third child, to take my maternity leave, pre-tenure research leave, and an external fellowship back to back. All of these advantages gave me two of the most precious gifts any aspiring academic with young children could ask for: flexibility and time. Without them, who knows how things might have turned out, or how I would feel now about my choices?

So I'm well aware of how lucky I've been. I also know that I'm far from unique. In my department at Penn State, I'm one of three women faculty with four children each; all three of us have tenure, one is an endowed chair, and one is department head. Although our department may be an outlier in this respect, I mention this in order to emphasize that there are other success stories out there as well.

Still, that's not to say that it has been easy. My oldest child was born two weeks after I defended my dissertation. I started my first job—a visiting assistant professorship at a small liberal arts college where I was replacing a faculty member who was going on parental leave—when he was three months old. I recognized, but didn't have much time to appreciate, the irony of the circumstances. We needed the income, and especially the health insurance (my oldest was born with a heart condition that required extensive treatment), so I took the job and didn't look back. My second child was born when I was on the first year of the tenure track, the third in the middle of my pre-tenure period. My youngest was born several years after I got tenure, just as I was finishing my second book. Parenting young children is a challenge in the best of circumstances; doing it while in the midst of an academic job search and then later while working toward tenure is incredibly stressful, even overwhelming at times. My husband and I certainly upped the degree of difficulty by deciding to have four children—and yes, that was a conscious decision—as we still remind ourselves whenever we feel overwhelmed with juggling the demands of career and family.

I suspect it is obvious to most readers of this newsletter that the factors that make it so difficult to combine parenthood with academic work are structural. Some of those are pretty generic, having to do with the lack of support for paid parental leave and the lack of access to high quality, affordable childcare in the United States. Others are more specific to academia, most notably, the fact that we have to run the gauntlet of the tenure track during our prime childbearing years. To be sure, academia also offers one tremendous structural benefit over other professions: a greater degree of autonomy over one's work schedule. (I think this is a general benefit of academic work, but it is perhaps especially useful for those who have significant caregiving responsibilities, including parents of young children.) Still, I think it is fair to say that the deck is stacked against would-be academic parents, structurally speaking, and there is no such thing as an individual solution to a structural problem. Unless and until we get widespread structural change, the best individuals can hope for is to muddle through.

As a result, I would never presume to give advice to those who are hoping to combine parenting with an academic career. The most I can say is a bit about the things that I did that helped me to muddle through. And I would hope that readers will refrain from judging too harshly individual decisions made in an effort to live a good life under conditions of structural oppression.

We got as much daycare as we could afford. We were lucky to have access to very high-quality daycare, but it was expensive enough (especially for infants) that in the beginning “as much daycare as we could afford” meant three days a week, and even that was a stretch financially. In those first few years, my children went to daycare on my teaching days, and the other days I “worked from home,” which usually meant that I got up at 5 a.m. to grade or do class prep and then caught up on email during their naptime. There were definitely times when it wasn’t clear that it made financial sense for us to spend so much money on daycare, given that, at the time, the cost represented a very high percentage of my husband’s take-home income. Still, although he is a terrific father and very involved parent, my husband never had any interest in staying home full time, so we pressed on. Now, twenty years later, the payoff of that strategy is clear: having invested the time and energy early on, we both enjoy satisfying and successful careers.
Relying so much on daycare required me to get over my working-mother guilt. I’m still not sure where that guilt came from—perhaps from having been a teenager in the 1980s, at the height of the post-second-wave anti-feminist backlash—but I had definitely internalized the idea that women should feel bad about putting their kids in daycare so that they could focus on their careers. (Bonus side effect: reflecting on this and similar experiences helped to shape my own views about the importance of feminist ideology critique, which was useful for my research). I grappled with that guilt a lot in the first few years. I even felt insecure enough in my choices that I stopped speaking to my former college roommate after she (unthinkingly, I suspect) made a comment about how it is selfish for women to put their children in full-time daycare. (This was probably sometime in the early 2000s.) Whenever a female student of mine would say something like “I don’t care which parent stays home with the kids, but one of them definitely should”—the “postfeminist” version of my college roommate’s more traditional judgment—I bristled. (This came up more than once in my women’s studies classes throughout the early 2000s, in the context of discussions of socialist feminist critiques of the gender division of paid and unpaid labor.)

Looking back, I’m not sure exactly how or even precisely when I got over the guilt, but I did. Perhaps it was because it became increasingly clear how good daycare was for my kids—they had caring and experienced teachers, they learned a lot and made friends, they developed extremely healthy immune systems (as I said, we were fortunate to have access to high-quality childcare, which I know is not everyone’s experience). Perhaps it was because I got more invested in my work as the years went on. Probably I was just too busy juggling work and family to notice that I felt guilty anymore.

Having my first child so early in my academic career forced me, I think, to become ruthlessly efficient with my time and realistic about my standards. If I had only two hours left before I had to leave to pick my kids up from daycare, then I finished that abstract in two hours. Generally speaking, I didn’t continue to work in the evenings after getting home from my on-campus office. This was not so much out of a principled decision to concentrate on family time in the evenings, but rather because I was so exhausted after dinner, baths, and climbing Mount Bedtime (as my friend Johanna Meehan aptly puts it) that by the time that was finished there was no way I could summon the energy and focus required for intellectual work. Sometimes I’d get up early to try to work for an hour or two before the kids were awake; occasionally, this strategy would backfire because they would hear me up moving around and just wake up earlier, but often enough it worked. Still, this meant that my work hours were relatively constrained, so I had to learn very early on how to be as efficient and productive as possible within the time that I had. I think that skill has served me well even as my kids have gotten older and the demands on my time have shifted. I also long ago gave up on the ideal of perfection, either as a philosopher or as a parent. I’ve always been more of a finisher than a perfectionist—and I think that academia ruthlessly punishes perfectionists, whether they have children or not—but still I had to learn to let go of some of my unrealistic standards and expectations. Of course, that’s not to say that I don’t take pride in my work nor is it to say that I don’t strive to have good relationships with my kids (a task that is shifting now that they are starting to leave home). It is just to say that I learned that the best way to alleviate the more or less constant feeling of failure experienced by many working mothers was to let go of impossibly high standards in both domains. After all, if there is no such thing as an individual solution to a structural problem, then it makes no sense to berate oneself for failing to do the impossible.

As I said, I offer these reflections not as advice but rather as a report of my own experience, as I understand it. I realize that some of it may no longer be relevant—at any rate, I really hope that working mother guilt is a thing of the past—and lots of it is unique to me. Still, when people ask me—and it is mostly women who do this, usually younger women who are just starting to make their way in the profession—how do you do it? This is more or less what I tell them.

Parenting, Feminism, and Academic Life: My Happy Story

Samantha Brennan
UNIVERSITY OF GUELPH

I’m Dean of the College of Arts and Professor of Philosophy at the University of Guelph in Canada. In recent years, I co-founded and co-edited Feminist Philosophy Quarterly, an online, open access journal. Last year, I ended my three-year term as a member of the Canadian Philosophical Association executive (vice president, president, and past president.) I ride my bike long distances, I go back country canoe camping, and I race small sailboats. I also have three adult children. It amuses me that it’s the last bit that makes people say “wow.” I often get asked, “How did you do it?” and I try to tell people that I didn’t do it alone. I’m not a single parent. If I was, then the “wow” would be justified. I did it as part of a team and no other members of the team ever got the “Wow. How do you do you do it?” reaction. I think that says something about our expectations of mothers.

I’ve been reluctant to tell my happy story of combining an academic career with parenthood, largely because I know most people, especially most women, haven’t had it so easy, and I hate for my story to be heard as “if I can do it, so can you.” But I have been struck, especially in comparison to women friends who became parents while practicing medicine or law, how much family friendlier my career was than theirs. I’ve also been struck by some pretty big differences between the shape of my life and that of my peers who struggle to find the balance between family life and an academic career. Don’t get me wrong. There is a lot about parenting that I’ve found challenging, but very little of that relates to my job.

My kids, all adults now, are 21, 23, and 27. My daughter Mallory, the eldest, was born while I was in grad school, and my two sons were born not post tenure, but after I’d written and published enough to make tenure promotion pretty
clear. The decision to have a child in grad school made sense to me then. I was at home, with a SSHRC doctoral fellowship, writing my thesis. I still think now, as I did then, that I had a lot of time on my hands. As someone who worked almost full-time while going to school full-time as an undergrad, grad school felt positively relaxing timewise. What has changed is that I think I was lucky when it came to the "baby in grad school" decision. I had an easy, healthy, happy baby. She slept so much that I occasionally woke her up to take a break from my thesis. She didn't cry much and loved being read aloud to no matter what the content, so I used to read journal articles on moral philosophy aloud to her. There may or may not be therapy bills later. She attended her first feminist philosophy retreat at six weeks of age, at Sandra Barty's Michigan cottage to hang with the other feminist grad students. Mallory attended her first APA at nine months, and provided a lot of stress relief for worn-out grad students on the job market. In a way, having a child early set the pace. Starting at three months, she went to work every day with my partner and came home at lunch. I started work when they left and wrote furiously until noon. I wrote a lot of my thesis on that schedule. I’ve always had a good sense of being at work—kids are in daycare, school, etc.—and good sense of being off. My new challenge will be seeing if I can stop work leaking more into my life as I move into academic administration at the same time as the kids all leave the nest. Wish me luck!

I hear lots of women talk about how hard it is to be a mother and a professor but mostly the stuff that strikes me as hard isn’t intrinsic to either parenthood or academic careers, it’s about trying to do that in a country (the United States) with no parental leave, inadequate daycare, far away from extended family, usually being married to partners who don’t do their share, with unrealistic standards for how much work parenting requires. I think I’ve lucked out on almost all of these fronts.

Here are ten thoughts I have about what worked well for me and why. YMMV, as they say.

First, it’s significant that my partner in parenting had less workforce attachment than I did and do. He left his job to follow me from grad school to my first academic position and took advantage of the spousal tuition benefit to return to university. When he first went to work after graduating he worked mostly 9–5, Monday to Friday. I don’t agree with everything Rhona Mahoney has to say, but when I read *Kidding Ourselves: Breadwinning, Babies And Bargaining Power* (Basic Books, 1996) it made sense of certain facts about my life. It turns out I had taken her advice directed at women who want an equal share of household and family work, “educate up, marry down.” “Down” is harsh and it’s not true in all respects, but in terms of income, hours worked, and education it’s been I who have had more.

Second, I had a family in a country with real parental leave and good daycare. You can do a lot with paid parental leave and good daycare. It goes a long way. The university daycare and preschool were pretty important to me in the early years. It wasn’t cheap and there was a waiting list. We got in because my partner went back to school and university undergraduates got priority access to daycare.

We were also eligible for provincial daycare subsidies because between my student loan debt and Jeff being a student, our income was low enough to qualify in the early years. I could go visit my children for lunch if I had the time and we commuted to campus together. My university also had good policies around this. The tenure clock stopped automatically for my leave and though I didn’t need the extra time, I appreciated having it.

Third, I wasn’t the first woman faculty member in my department to have children. Indeed, Kathleen Okruhlik was chair when I was hired and dean shortly after, and she also had children. This was a switch for me. Dalhousie University, my first philosophy department, when I was an undergrad, had just one woman on faculty, and she was married but didn’t have children. My second department, University of Illinois at Chicago, where I went to grad school, had six women on faculty, but none of them had children until Dorothy Grover adopted her son in my final year. I think it made a difference not being first at Western. Even with this, though, the students didn’t always react well to seeing me on campus with small children. The funniest comment I’ve ever received on a teaching evaluation was “Professor Brennan cares more about her children than she does about us.” Right. A senior colleague once stopped by while I was having lunch in my office with my daughter and said he had no idea how I did it. His wife could have never managed being a professor while taking care of the house and the kids. (I didn’t say, though I thought it, “that’s because she’s married to you.”)

Fourth, you might have already noticed that I use the word “parent” rather than “mother” as I tried to sidestep a lot of gendered baggage that came with having children and living in a house. I used to think there would come a point in my life when I’d care about curtains and cutlery, but that hasn’t yet arrived. I’m still not a big fan of cooking—see Sam dislikes cooking and she’s not alone. There’s a lot I didn’t do and still don’t do that lots of people think of as part of parenting for women. I’m glad I got through before the years of Pinterest and the mommy bloggers.

Fifth, my family is super supportive. My parents moved to the city where I was a professor to live with me and help with the kids when we were about to put a third child into daycare. We looked at the daycare bill. We looked at my mother’s current paycheck. We looked at how good she was with the kids and how much she missed them when she was back in Nova Scotia. We looked at the daycare bill. We looked too at how exhausted and overworked we were. It just made sense all round. For more than twenty years we’ve lived side-by-side, in connected living spaces, helping each other out. “Help” is, of course, an understatement. My mother worked full-time keeping house, providing full-time care, watching the kids before and after school. She was paid for this work and went from full time to part time, from focusing mainly on the kids to mainly on the house, and just retired altogether last year. (True confession: She still does the laundry. However, she has grandkids who shovel snow and lift heavy things. We’re still a team.)

My father and my father-in-law did all the driving. At one point we’d send out Sunday night emails with chauffeuring
assignments for each grandfather. They loved it. The kids loved it too. Grandfathers, it turned out, will go for drive-thru fries after school. While they drove, I got to commute by bike.

Jeff’s sister also moved just a few blocks away when the kids were young, and she was the aunt to whom the kids ran away when they’d had enough of their parents and their in-city grandparents.

Sixth, we’ve always thrown a lot of money at childcare. My kids spent a lot of time on campus in various camps and sporting clubs. Often it was pricey but it felt like money well spent. Why camps when we had the grandparents? We wanted to keep everyone happy for the long haul. We used to joke that it was easy to take care of any two of the three children together, but the three all at once were too much. During the summer one child at a time went to camp usually. Lots of the camps weren’t really designed for working parents. Often, they ended at 2:30. That’s where the grandfathers came in.

Seventh, the doors of our house were wide open. My Facebook description says that I am joyful about living as part of a large, extended family with porous boundaries and long tentacles. We’ve had a lot of friends, in addition to the extended family, who’ve played a significant role in the lives of our children.

As a young adult in my twenties, in grad school, I knew I wanted children in my life but I wasn’t, yet, committed to the idea of becoming a parent. I was an idealist and I imagined, as I think many young people do, that my generation would do things differently. I thought of cooperative parenting, of communal living, and of alternative family arrangements. Instead, those friends having children also got married, and parenthood looked like this incredibly private, intimate thing. There was no easy access to the children of other people, and it seemed as if what I wanted was a more open model of parenting. I’d need to be the parent, inviting others in, rather than the other way round. In addition to our three biological children we’ve also opened our home through foster parenting, for a short period of time, to other children as well. During the teenage years, there was usually an extra child living with us. Often, they were friends of my gay son who had been thrown out of their own homes for a while or permanently. I was sad to find out how much this still happens.

Eighth, I know other parents did stuff that I didn’t do even beyond the “mothering” point above. Now I see young academic couples who on principled grounds don’t use daycare or after-school care. I can’t imagine what that would be like. Deans tell stories about academic couples who say they need to end work at 4 p.m. every day so both parents (both!) can pick up the children. Both? After-school care? My partner started dubbing these extremely intensive parenting practises—no screens, only organic food, no daycare—as “artisanal parenting.” These are the same young parents who only drink craft beer and listen to music on vinyl. He asks, “Do they knit and carve all their kids’ toys too?” Maybe.

Nine, we travelled as a family on sabbaticals, and that is something I definitely recommend as both a great research experience and family bonding experience. Our kids had gotten to the ages where they spent more time with friends than with one another. Moving to a place where we were all new and learning our way around together made us feel more connected as a family. Our children became friends with one another again.

Tenth, babies and toddlers were easy compared to teenagers but not for the reasons most people think. In many ways, I loved having a house full of teenagers. So energetic, so earnest. But it was work. Babies and toddlers have pretty simple needs. I also think lots of different people can meet those needs. My kids, when young, loved being with lots of different adult caregivers. And even if it was me providing the care I could have my mind on other things. You can walk a baby in a stroller and think about philosophy. You can read toddler books and think about something else. Teenagers wanted and needed my actual full-on attention. One morning I was broken by a teenage child sitting at the edge of my bed, “It’s five o’clock. You sometimes get up at 5 so I figured it was okay to wake you. I haven’t been able to sleep. I’ve been thinking. Why is the world so shitty? Why is life worth living?” We got up and made waffles and drank orange juice and talked about some hard stuff. They all wanted to talk about friendship and about relationship problems. And about ethical problems and about politics. I spent all night at the ER once with a high school-aged kid, not mine, who had taken too much of something and wasn’t breathing properly. He wouldn’t tell me his last name and my kids claimed to not know it. They just went to school together. I was giving a talk at the Canadian Society for Practical Ethics the next afternoon. I did it on just a few hours of sleep.

I actually wish there had been parental leave for parents of teenagers. I needed it.

With teenagers it wasn’t the kind of parenting you can pay someone else to do. You’re on, a lot of the time, and it’s messy and complicated and I don’t think I thought too much about how messy and complicated life is when I decided to have children.

Oh, they also talked my parents’ ears off.

I know this is not your usual woman-academic-with-kids story. My family is more easily recognizable to other immigrant families with its multigenerational lifestyle. My parents were able to move from the east coast to Ontario in part because there was no other family in Nova Scotia. They’d moved from England to Canada in the 1960s and all of their extended family is in England or Australia. Last fall, my mother and I (my dad died a few years ago) celebrated fifty years in Canada. It helps too that my parents weren’t wealthy. The deal where my mum took care of my kids was good all round. We all benefited. They got a lot more time with the grandkids. We got time to ride bikes and sail boats. And I had, and still have, a happy work-life balance as a philosopher.
Bridging the Divide: Thoughts on Parenting as a Grad Student

Carol Gray  
UNIVERSITY OF CONNECTICUT

I am a feminist, an activist, an attorney, a graduate student, a wife, and a mother. There are a few other labels I could add, but the point is, like others who are parents and graduate students, I constantly feel torn between these two roles. There is never enough time to do either role as I would like to do them, and I regularly feel guilty for falling short in both arenas. But parenting has been one of the most enriching experiences of my life, and I believe it can fit well with graduate work most of the time. Here I will offer some stories and tips on how to navigate and sometimes bridge parenting and graduate work.

FINDING THE TIME

Though grad school demands one’s full attention, one cannot put parenting on hold. When I started my PhD in political science, my son, Cameron, was nine years old. He is now fifteen years old. His childhood has passed while I’ve been in grad school. Perhaps the greatest challenge for parenting in graduate school is finding enough time to be the kind of parent I want to be and to complete the volumes of work required for graduate school. Before I started grad school, I would read aloud with my son before bed, taking turns reading as we plodded our way through *Oliver Twist, Little Women, Little Men*, and others. Grad school started when we were in the middle of *David Copperfield* and we never finished the book. I gradually shifted to singing Cameron a song before bed instead of reading because that 45–60 minutes of reading meant one hour less sleep. I just couldn’t risk getting less sleep because of the 15–20 hours of driving I was doing every week commuting to school. Before grad school, I tried to arrange my adjunct teaching schedule such that I could be home when my son got home from school so I could hear about his day. But grad school meant I was away so much more . . . missing so much more of his life. When I was at school, I’d try to find a break right after the time when the school bus dropped Cameron at home with his father so I could call and hear about school over the phone. I learned about many important events by voicemail (“Mommy, I qualified for the State Geography Bee today!”). Phones have been my lifeline for parenting. When I couldn’t attend his parent/teacher conference in person, I worked out with my husband to have me on speakerphone as I called in from school. When a doctor’s appointment was during a seminar I was taking, I explained to my professor in advance that I would need to step out of the seminar at the time of the appointment after my husband texted me that he was in the doctor’s office. And there I was again on speakerphone asking the many questions I would want to ask that my husband, who is quite reserved, would not want to ask.

One of the hardest things is when there’s just no way around the conflict. The class I taught as a GA was on the night of my son’s band concert. When I’d arrive late for soccer games, I dreaded the question Cameron inevitably asked after the game: “Were you there when X happened?” [sometimes a goal or great play at the beginning of the game]. When I’d have to admit I didn’t get there until after that, his facial expression would drop in disappointment and I would sink with guilt.

Harder still was the guilt of losing my patience because I was just so tired and stressed with the workload. I had the good fortune of having a mother who almost never raised her voice with me. I fell short so many times with Cameron in this regard, regularly apologizing afterwards and recommitting to doing better next time. But doing better was usually contingent on getting at least 7–8 hours sleep, a rarity for me in grad school. I struggled against chronic sleep deprivation because of school, my almost 1.5-hour commute each direction, and trying to keep up with parenting responsibilities. Fortunately, my husband and I share parenting duties. He drives Cameron to soccer practices and theater rehearsals; I drive him to piano and oboe lessons and orchestra rehearsals. Sometimes I am so tired that I take a nap during the oboe lesson or I have to pull over to sleep when driving to or from school. I keep one of our inflatable camping bedrolls in my drawer at my office and nap at school sometimes to be able to drive home safely.

The teenage years made the effects of sleep deprivation even harder. My politics have always been to teach my son to question authority. During the teenage years, that authority was usually me. My husband was always the easy-going one, leaving the discipline to me. Cameron inherited my strong will and tenacity. His determination and litigious debate style of arguing (. . . I wonder where he got that from . . .) was a formidable test for my sometimes threadbare nerves.

As I’m entering the job search phase of my graduate years, parenthood has meant that I can’t apply for any jobs that are more than an hour or so from our home. Our son is in a good high school, so our lives are very much rooted to the school system, the soccer team, the local theater program, the orchestra where he plays the oboe, etc., etc. Uprooting teens in high school isn’t advisable. I decided long ago that I may need to be jobless for a year or two after I finish my PhD until my son goes away to college and I can then be on the national job market.

Though juggling grad school and parenting has been challenging, I have found some strategies to be helpful. Here are my tips and the stories behind them.

BRING YOUR CHILD TO SCHOOL

To try to minimize the alienation of having my school world so separate from my son’s world, I’ve looked for ways to expose him to my school world. I look for events at the university that would be good family entertainment. A couple times during the summer, my husband, son, and I travel more than an hour each way to attend a musical at my university. With my student ID, I get two student tickets, so the entertainment is both fun and affordable. When there is a snow day at my son’s school, I bring him to school with me. For classes that I’m a student in, I speak to the professor in advance to make sure they are ok with me.
bringing my son. They have all been very accommodating. This has meant that when I’m at home talking about a class, I can say, “Cameron, that’s the class with Ernie that you went to, remember?” A couple times I have taken Cameron to classes I’m teaching, particularly if it’s a fun activity. I create a jeopardy game to help my students prepare for quizzes, and my son sometimes comes to be the score keeper. Cameron likes to meet my students and my students seem to like to meet him.

**SCHEDULE CLASSES AROUND FAMILY ACTIVITIES**

I made a point of scheduling classes that would maximize my family time. Thankfully, the professor who oversees graduate students at my university has been very accommodating, always scheduling my teaching for classes that meet one night a week. This meant I only missed one evening with my family and was still home in time to see Cameron before he went to sleep. In terms of classes I took as a student, I sometimes opted not to take classes I would have enjoyed because they met at 6:30 p.m. and would have meant I didn’t see my son at all that day. I also managed to avoid Friday classes several semesters, which meant I was available to volunteer as a chaperone for my son’s school field trips.

**MAKE IT UP TO THEM**

It is inevitable that grad school will mean that during crunch times, I just don’t see much of my son and husband. I may miss my son’s bedtime routine because I have a paper due at 10 p.m. and even that 15 minutes at bedtime is more than I can spare. The summer I was taking my two comprehensive exams, I planned a five-day vacation with my family at the beginning of the summer, then studied for the rest of the summer and went to a hotel to take my exams so that I could be away from family distractions. Leaving them for those exam days seemed far better than failing and having to go through more grueling months of studying. Thankfully, my study plan worked and I passed both exams the first time. I told my advisor at the end of that summer that the following summer I had to make it up to my family. And I did. We spent weeks driving across country visiting national parks. We listened to many books on tape from *Sherlock Holmes* to *The Adventures of Origami Yoda*, and all of the *Harry Potter* books.

Our national park odyssey meant that I made little progress on my dissertation that summer, but it was one of the best things I’ve done in my life. When I hear my son talk to friends about his big national park summer, I know I made the right choice. Creating those childhood memories will hopefully help him forgive me for all the times I wasn’t there or wasn’t as patient as I should have been. One thing I have accepted is that it will probably take me longer to complete my PhD because I’m trying to be a good parent at the same time, and that takes time. Fortunately, I’m still expecting to be able to finish my PhD within the years of funding provided by a fellowship that I’ve received. I’m also fortunate that I’m on this journey with my husband so we are not dependent on just my income.

**TRY TO KEEP UP FAMILY TRADITIONS; BUT CUT YOURSELF SLACK WHEN YOU CAN’T**

We’ve tried to keep our regular family traditions despite the intensity of graduate school. Every year I take my son trick-or-treating and help support him in creating his own Halloween costume. We plan a birthday party at some new location each year (e.g., at a hiking trail, a science museum, a corn maze, etc.). We have been pretty good about keeping Friday night as our “family movie night,” rotating who in the family gets to pick the movie of the week. (One professor in my department suggests that all grad students and professors take one day off from work each week to be able to come back fresh, rejuvenated, and more productive.) But sometimes we just haven’t had the energy and time to keep up with family traditions. One year, we didn’t get a Christmas tree because it was so late in the season and we were going away soon after Christmas anyway. We all realized we were just too tired, and it was too much work, so we would be happier not adding that “one more thing.”

**FIGURE OUT WAYS TO BE TOGETHER**

Children, like parents, often just want to find time and ways to be together. Sometimes I’ve been able to do this while doing things I had to get done for school. When I had to drive the one hour to school to return library books that were due, I invited Cameron to come and we’d go out to pick up some food afterwards. Sometimes we’d go to the science museum and I would bring my articles along to read while he learned and explored. At least we were together. Now that Cameron is in high school, he has a lot of homework like me, so we plan homework outings where we study at the local bookstore that doubles as a library.

One of the ways I’ve been able to maximize time with my family has been by doing my homework in the car while commuting. I get some books in Kindle form and I email myself PDFs of articles. Then I use the voice-over feature of my iPhone to have a monotonous mechanical woman’s voice read my assignments to me. With courses that are electives, I sometimes request the syllabus before registering to see how many of the books I can get in Kindle format so I can know if I will be able to complete reading while commuting.

To an outsider it might seem that my roles as parent and grad student are compartmentalized. I usually don’t talk in my graduate seminars about my role as a mother. I usually don’t talk at my son’s soccer games about what I’m learning about Fanon and decolonialism or how to rethink feminism in a context that is not dominated by white, Western thought. My hour and 15-minute commute from Massachusetts to the University of Connecticut is my transition between these two worlds. But in truth, I’m almost always living in both worlds at the same time. And therein lies the challenge as well as the richness of parenting in grad school.
**Anthropologists from Mars**

Leigh Viner  
**INDIANA UNIVERSITY SOUTHEAST**

It’s a Sunday evening and students start appearing at our door, bearing cheese, crackers, grapes, beer, books tucked under arms, bringing friends, some nervous, some excited, some cool and confident. They greet me and chatter with each other as we unwrap the snacks and pour drinks. Everyone finds a spot in the living room on couches or cross-legged on the floor, and we start to discuss the text we’ve chosen that month. My son comes in, listens for a while, chats with the students, grabs some cheese and crackers, then leaves. One evening at the end of Reading Group, he invites everyone to his room and we have a round of Rock Band with our new colleague, the Kant scholar, providing a karaoke rendition of Radiohead’s Creep. It was hilarious and wonderful, the essence of community.

One summer, a student loaned me his Sex and the City DVD collection. I binge-watched the whole series during break, my twelve-year-old son seemingly not paying attention on the couch with his laptop. Years later, when he was in college, he rewatched the series with his partner and housemates, and I discovered that he knew and loved every character, their relationships, friendships, and story arcs. When asked by one of his professors to program the university’s international film festival, he quietly and without fanfare chose only films directed by women.

My son is grown up now, and living in New York City with his partner. But over the years, I described parenting him as feeling like an anthropologist from Mars, with him as my assistant. Philosophers, like artists, always seem to be simultaneously inside and outside of society’s conventions, inside and outside of our own experiences. And my good fortune at finding a full-time position at a regional campus of a state university, with its emphasis on teaching and service, allowed me to provide a simple but stable life for us, with plenty of free time to take naps and walks, watch movies, play music, and talk about life together.

The philosophers that I read, teach, and write about are the ones whose ideas fuel and shape my life and imagination, and are therefore the ones who have shaped my parenting as well. Too many to list, but generally, they’re the philosophers who value art, play, freedom, dignity, and openness. I was trained in the history of philosophy, with a focus on the ancient Greeks. Then, through teaching, I explored and grew, following my curiosity and the needs of the program. But it was my son and my students who led me to feminism. Like all women, I’ve had to deal with injustices and indignities over the years, which I usually tried to ignore or rise above. But, coming to a new understanding of those experiences, seeing them as they were reflected back to me through the eyes of people I love and care for, and who love and respect me, made it necessary to confront those issues differently, more directly, more deeply, and more honestly, both for my sake and for theirs. So the simplest and most powerful truth is we learn at least as much from our children and from our students as they do from us, if we’re doing it right.

**Raised in Philosophy**

Eli Kukla  
**MCGILL UNIVERSITY**

I am a “pure-bred” philosophers’ kid; both of my parents are philosophers. I am also a second-generation philosophers’ kid, as my grandfather is also a philosopher. So, growing up, I was even more surrounded by philosophy than most philosophers’ kids, and my immersion in the discipline has affected seemingly every aspect of my life and personality. My philosophical upbringing is seen on the surface level of my personality. It shows in my sense of humor, my rhythm of speech, my style of argumentation, and in the ways in which I ask questions. It manifests all the way down to the fundamental way in which I think, reason, consider evidence, and evaluate claims. Being raised by philosophers has also shaped how I move around the world and my relationship with academia and college campuses in general.

We who have been inducted into philosophical norms and habits (I say “we” because of my social and intellectual integration into the field, though not my academic one, as not only am I not a philosophy PhD or grad student, but I’m a linguistics major at McGill) have such a distinctive manner of speaking and communicating. Our senses of humor, our verbal ticks, our communicative cues: it’s all so unique to philosophy and makes us instantly recognizable. You may even be seeing these communicative cues while reading this, what with my long digression explaining the context of my use of the first person when talking about philosophers.

The very rhythm and flow of how I speak was shaped from the beginning by philosophers, and it’s a rhythm and flow of speech that I have personally found to be ubiquitous among—and exclusive to—philosophers. We preface our speech to provide specific contexts to what we’re saying. We have a distinctive pattern of changing the speed at which we speak, speaking very quickly for certain parts of our sentences and then pausing to carefully consider which word needs to come next. We digress to try to clarify and to justify specific seemingly minor aspects of what we’re saying. When trying to convince someone of something, we continue to give reasons for our arguments long after they’ve already agreed, just to show that we have every counterargument and varying aspect of the issue covered.

Speaking “Philosopher” almost feels like speaking a specific dialect. For example, philosophers often have a distinctive way of having friendly arguments which contrasts with how we have non-friendly arguments. The friendly argument does not come from a place of hostility or anger, but is instead used as a kind of marker that the person you’re arguing with is smart and interesting and has things worth saying. Non-philosophers, however, often can’t tell the difference between these two kinds of arguments, and therefore confuse the friendly argument with a non-friendly argument. This means that people who don’t speak the “Philosopher” dialect can sometimes perceive philosophers as overly confrontational or critical, similar to how non-New Yorkers hear the New York dialect and conclude that New Yorkers are rude and harsh.
These are all things that as philosophers, we may never realize we do until it’s pointed out to us, but once we recognize the pattern of the “Philosopher” dialect, it makes others with a philosophy background instantly identifiable. I know that for me, growing up surrounded almost exclusively by philosophers, it took me until around grade school to realize that no other people, neither my peers nor my teachers, talked like that, and that I had a distinctly unique or odd cadence to my speech, from their perspective.

The most paradigmatic example of our dialect that I’ve ever heard was a quote from my grandfather, also a philosopher, who is currently living in an eldercare home in Toronto. He suffers from very serious Parkinson’s, so any kind of communication is extremely difficult. We asked him, not really expecting an answer, “So, how do you like the nurses here? Are they doing a good job?” to which he, speaking slowly and laboriously and fighting through the Parkinson’s, managed to say, “I haven’t interacted with them enough to formulate a proper inductive base yet.”

Not infrequently, I’ll be faced with one of those boring conversation-starter questions by friends or something. You know the type, someone will disinterestedly say something like “if you could be any animal, what animal would you be?” to which I, in the way that only philosophers do, would respond with something like “well, I don’t really have an epistemic framework from which to answer your question,”—this would usually be met with a chorus of groans and eye rolls—“Do I get the animal’s brain and thoughts? In which case I don’t care what animal I am because whatever animal I choose wouldn’t have the capabilities to appreciate its being that animal . . .” and I’d usually go on from there.

These patterns and rhythms of speech, argumentation, humor, and questioning have all been things that were imparted to me by virtue of growing up around philosophers. I’ve used the dialect metaphor a couple times, not only because of my personal interest in language and linguistics, but because I think it’s a really apt comparison. Like imparting a dialect, teaching me to speak and reason in this way wasn’t a conscious choice of my parents, but rather an inevitable consequence of my growing up around them. However, there were other ways in which my parents chose specifically to use philosophical principles and thinking to parent me. My mother in particular was very careful not to prescribe rules or assumptions without proper and convincing evidence. So there was never an attempt to control my behavior with reasons or excuses like “because I said so” or “that’s just the way it is.”

Most people, at some point in their lives, reach a point where they question many of their assumptions about life. People question religion, sexuality, monogamy, traditional family structures, gender, et cetera. But for me, being raised by philosophers, I never started out with any of these assumptions to begin with. I never reached a point where I questioned religion, because I wasn’t given any convincing arguments for the existence of a God in the first place. I never had to question the value of monogamy because I was never raised to assume that it was the only way for a relationship to be, even though it’s the norm. I was exposed to all sorts of non-traditional family structures, various sexualities, gender identities, and beliefs and left to decide for myself their value. It was never assumed (by my mother, anyway) that I was straight or that I was cis, and she never treated me as though those were a given. (I discovered for myself over time that, at least as of writing this, I am not straight, but I do happen to be cis.)

This parenting method of teaching me not to accept things without valid reasons did give me a fierce disdain for unquestioned authority, much to the dismay of many of my teachers throughout grade school. I would particularly anger certain teachers by my unflinching refusal to participate in the morning Pledge of Allegiance during elementary and middle school. Multiple times, I was told to stand up to say the pledge and would be met with flustered disbelief or confusion when I would argue back about the racism and neo-colonialism of dogmatic American Exceptionalism, or my many reasons for not wanting to perform support for the United States or American interests.

Of course, my argumentation got more sophisticated over time, and by eighth grade I was much better able to express why I refused to participate in the Pledge of Allegiance than I was during fourth grade (the first time I had to say it). However, although I didn’t have the vocabulary or sophistication yet to explain why I had problems with the pledge, I knew that there was something about it that I found weird, unnecessary, and creepy. I knew this because of how my philosopher parents raised me to never accept something at face value, and to question unsubstantiated authority and assumptions.

Socially, being a philosopher’s child gave me an interesting and unique community and set of relationships all over the world. I remember a while ago I was trapped unexpectedly in Toronto due to a snow storm, and philosophers from the University of Toronto took me in on short notice late at night. The experience got me thinking about how many connections I could have through philosophy all over the world. There are very few cities where I could be trapped and have no connections to anyone. Philosophy, as a field, is just the right size to enable this. Larger fields like English or biology don’t have the same kind of intimacy as philosophy. Philosophers tend to know each other, either personally or by their work, and seem to be quite well connected over social media. With a smaller and less well-represented field, there is that same interconnectedness and intimacy as in philosophy, but you wouldn’t be able to count on connections in every city in the same way.

Growing up accompanying my mom while she gave various talks or went to conferences in so many cities, I’ve come to view college campuses, and even more specifically philosophy departments, as little centers of my community all around the world. I’ve visited campuses in Bogotá, Johannesburg, Hong Kong, Beirut, Paris, and more, and in each of these places I’ve found some sense of home and belonging in them. I like to think of my relationship to campuses and philosophy departments sort of like Chinatowns. Almost everywhere seems to have one, and no matter how foreign or unfamiliar a place may feel,
Chinatowns all tend to have a similar feel and sense of welcoming familiarity. Being raised a philosophers’ kid has allowed me to have this feeling of a familiar community enclave almost anywhere I go.

Being a philosophers’ kid also certainly helped me easily integrate into university life, and it helped me absolutely crush my first formal philosophy class I ever took, Social and Political Philosophy. But for me, being a philosophers’ kid is more than that; it’s a sense of self. Philosophy to me is more than a career or a department; philosophy is an identity.

Children, Parenting, and the Nature of Work

Lauren Freeman
UNIVERSITY OF LOUISVILLE

This morning at breakfast, in the context of absolutely nothing, my five-year-old daughter looked at me, incredulous, and said, “Wait, you’re a teacher and a mother? How do you do everything? That’s so much work!”

I just stared. I was dumbfounded at how her question and comment so clearly hit upon the truth. I was both surprised and grateful that she recognized that. I hadn’t been complaining about a deadline that I’d missed this week, or how behind I was in preparing for my classes and grading, nor had I so much as mentioned all of the seemingly insurmountable tasks I’d set out for myself that day. In fact, it was in a very rare moment of silence at our breakfast table that for some reason unknown to me, she came to that realization.

It made me think about a somewhat similar episode from several months earlier. I was invited to be a community visitor at my son’s tiny elementary school. I’d visited the previous year, with great success, and I agreed to return to do some philosophy with the children. It is a classical school, a rare place where I can ask a room of five- to eight-year-olds, “Who is considered to be the first classical school after all; every Friday they spend half an afternoon drawing upon their own four-month study of Ancient Greece), I followed up with “and who was his most famous student?” “PLATO!!!” they all yelled, smiles abound. After making a silly joke distinguishing Plato from Play-doh (they found it to be funny), I continued to ask who Plato’s most famous student was and all I could hear, in joyous unison, was, “ARISTOTLE.”

This led to our discussion of the good life. Within that context, we talked about the value of community and the nature and value of work. In order to do this, we read a children’s classic, Leo Lionni’s 1967 book, Frederick.

The protagonist is Frederick, an idiosyncratic little field mouse. All through autumn, while his friends and family are busy collecting food and supplies to last them through the cold winter, Fredrick is off by himself, doing his own thing—to some, doing nothing. He collects rays of the sun for when it’s cold; colors, for when it’s dreary; and words, for when conversation runs out.

Predictably, by the end of winter, all of the food and supplies that the field mice had collected in autumn are gone. They start to wither. But Frederick warms them by summoning rays of the sun in their imaginations (“or is it magic?”), brightens their days by reminding them of the glorious colors of nature, and recites verse, invigorating their minds, getting them through until spring.

Frederick, we learn by the end, is a poet.

After I finished reading the book, I asked the children a series of questions about the nature and value of work, in order to test their intuitions about these ideas. As adults reading this book in a capitalist society, we tend to think that “work” is what you get paid to do, what you can “cash in on” in terms of tangible or monetary rewards. The questions I was asking were meant to problematize for the children the nature of valuable work by bringing them to think about what kind of work is valuable in itself (and not just for some monetary or extrinsic gain).

To my shocking delight, the majority of them had profoundly feminist intuitions.

They all believed, from the outset and without any convincing from me, that Frederick’s work—collecting rays of the sun, and colors, and words—was inherently valuable. Moreover, when I played the devil’s advocate and asked them to convince me that Frederick’s work was valuable (I was essentially asking them to construct an argument to this effect), as if it was the most obvious thing in the world, they told me that it’s absolutely necessary to collect rays of the sun, for how else will the mice stay warm in the cold winter? They also told me that it’s entirely necessary to collect colors, for we need them to brighten our lives in the grey of winter (they looked out the window). And, finally, they pointed out to me what was obvious to them (this is a classical school after all; every Friday they spend half an hour learning and reciting poetry together), that poetry is an important part of the good life and valuable work. It seemed almost absurd to them that I was belaboring this point. To them, it was all commonplace.

Their argument about work, if I may be so bold as to reconstruct it, was that care work is work and that it’s inherently valuable. Frederick was caring for his friends and his community; it was Frederick who got the field mice through the winter by invigorating their imaginations. And the children recognized this kind of work as being of supreme value.

I then brought the discussion back to their own lives. I asked them what work they see around them in their families and communities that they think is valuable. And again, to my delight and surprise, so many of them remarked upon how
their parents do the really hard work of raising them and caring for them and their families. What struck me is that children really and truly valued the care work that is done for them and only secondarily mentioned the “real jobs” that their parents have.

The other day I was teaching a class in my feminist bioethics course and we were talking about vulnerability and dependency. We were talking about the various ways in which feminist bioethicists draw attention to the ways in which the work of caregiving in our society, that primarily and not accidentally tends to fall on the shoulders of women, is neither rewarded, nor recognized, nor valued.

But then I thought of my daughter’s comments at breakfast, and the intuitions and strongly held beliefs of the children in my son’s school, and the deep value they see in the care work that surrounds them, and, like Frederick, I thought that maybe the rays of light to be found in the future will sustain us.

BOOK REVIEWS

Learning from My Daughter: The Value and Care of Disabled Minds


Reviewed by Lisa Tessman
BINGHAMTON UNIVERSITY

Eva Feder Kittay’s Learning from My Daughter is, in my view, her best work to date. As I read, I felt I was being guided through a varied terrain—some of it familiar, some unfamiliar—by someone who is much wiser than I am. Kittay’s experience—of loving and caring for both a daughter with severe cognitive and physical disabilities and a son who is not disabled—motivates and informs the book in crucial ways, while her philosophical mastery allows her to think through and convey with clarity the insights that this experience yields. The book made me realize that, as someone who has not raised a disabled child, I have much to learn from Kittay’s experience. But I also realized that having had the experience of raising any child—and thus the experience of loving exactly the child I have, and loving her exactly the way she is—I can understand and assent to pretty much everything Kittay proposes in her book, for it all follows in a sort of obvious and irrefutable way from such love.

The book is divided into three parts. I found Part I to be the most profound—it is truly on what matters. Part II is a superb example of powerfully argued analytic philosophy. Part III is a significant and innovative contribution to care ethics.

For readers like myself who accept and appreciate some of the basic premises of Harry Frankfurt’s work (in, for instance, “The Importance of What We Care About” and The Reasons of Love), Part I of Kittay’s book can be read as a tremendous deepening of the Frankfurtian position. For Kittay, as for Frankfurt, what really matters depends on what matters to us, which is to say what we love and care about. The value of a person’s life is not to be found in any of its objective properties; rather, it is through valuing that we confer value. The life lived by a person with disabilities matters—in exactly the same way that an able person’s life matters—because they value their own life, and also because their life matters to those who love them. It really is this simple. Valuing her daughter, Sesha, has given Kittay a different, more expansive, answer to the question of what matters than that offered by traditional philosophical accounts that center narrowly on Reason.

Sesha’s life is a good life—but saying this requires a reexamination of the concept of the normal, because given the tendency to conflate the good life and the normal life, it might seem that a life that is not normal also cannot be good. When “the” good life is based on either a statistical norm or an ideal modelled by those who “exceed” the statistical norm in some way, people like Sesha appear to be living lives of lesser value. One might think that to resist the conflation of the good life with the normal life, disability advocates would want to abandon norms altogether and adopt, perhaps, an entirely subjective account of what a good life is. Kittay argues against this, instead examining why a desire for the normal persists, and finding that the desire for the normal is “a desire to have one’s own worth and that of one’s children confirmed” (33). The subjective experience of what constitutes a good life remains central in Kittay’s account—and so what Sesha cares about is what matters the most to the question of what makes her life good—but because value is also constructed socially, it is important for what we each value to also be valued by others, or put differently, for others to recognize the value of what we value. Within her family, Kittay created a “new normal,” but because recognition from those beyond the family is also vital, Kittay argues that, socially, we need to create “more capacious norms . . . that embrace more varieties of flourishing and that generate their own source of desirability” (41).

Part II of the book takes up a set of questions about whether or not one is morally permitted (or required) to use reproductive technology to select either for or against disabling traits. While Kittay affirms that there can be good reasons to select against an impaired fetus in certain cases—making it permissible—they are not the reasons cited in the most common arguments; and thus, one of her main points is that there are problematic ways to approach the topic, found in some arguments for the claim that it is obligatory to select against disability. The idealizing methodologies behind these arguments tend to trigger the wrong intuitions, which then appear to confirm the assumption that the lives of people with disabilities are less good than the lives of people without disabilities. This sends a message that is both wrong and harmful to people with disabilities—that their lives are of lesser value and that they are not welcome in the world. Disability theorists have developed the “expressivist objection” not only to this type of argument, but also to any argument that supports the permissibility of selecting against disability. While Kittay endorses the expressivist objection when it
is applied to the kind of argument she has critiqued, she insists that not all reasons for engaging in prenatal testing and selection express the view that the lives of people with disabilities are of lesser value. In particular, she believes that a woman who is carrying a fetus may permissibly select against disabling traits if she decides that, while the life of a child with disabilities is as valuable as any other human life, when she takes into consideration whether she is capable of caring for a child with disabilities, whether she is willing to subordinate the rest of what she cares about, including her work and her other commitments, to the needs of a child with disabilities, whether it is the right time in her life to undertake this kind of care, how it will affect her other children, and so on, there may be good (all things considered) reasons to avoid carrying a fetus with disabling traits to term. Of course, disability is stigmatized, and not enough resources are directed to supporting people with disabilities and those who care for them—and it is in the context of these injustices that a pregnant woman must make her decision about whether to raise a child who is likely to have disabilities. This should be remedied. But even if it were, there are additional burdens to consider, including the emotional difficulty of caring for someone who is extremely vulnerable and dependent. In keeping with a feminist commitment to reproductive choice, a woman must be able to choose—without moral disapprobation—whether or not she accepts the burdens of any particular pregnancy. Nevertheless, Kittay reports on an important disagreement with her non-disabled son who gives several arguments in support of the expressivist objection. There is one point that he raises from his perspective as the sibling of a person with disabilities: a parent who selects against an impaired fetus runs the risk of communicating to any of their existing children that they are not loved unconditionally—they, too, would have been aborted had they had impairments. Kittay argues that loving unconditionally—whatever child one ends up with—is compatible with aiming to select against disabling traits (as long as one does so for the sorts of good reasons named), for all attempts at selection must be made with full awareness that our control is limited: whom we love, and whom we care for, is very much a matter of luck. An existing child is indeed to be loved unconditionally; a fetus, on the other hand, is not yet in relation until someone accepts the role of caring for them in a dependency relation.

While everyone is dependent at some points in their life, people with disabilities as severe as Sesha’s are utterly dependent for their entire lives. Caring for Sesha has thus given Kittay a view of dependency writ large and has led her to deeply appreciate the complexities of caring. Kittay’s analysis of the dependency relationship, and her development of care ethics in Part III of the book, builds on the insights of Part I by putting the subjective experience of what matters—which Kittay terms a person’s “CARES”—at the center of care ethics. Recognizing the fact of human dependency focuses us on the importance of care, for “when we cannot attend to our CARES on our own, we are dependent on others” (152). In Kittay’s account of what care is (which is a normative, and not just descriptive, account, namely, it presents what care ought to be), to care for someone is to promote their flourishing, while at the same time being attentive to the flourishing of other parties, such as the carer, as well as those outside of the relationship. Care is to be given to those who cannot flourish without it. Importantly, the flourishing to be promoted is “flourishing as endorsed (implicitly or explicitly) by the one cared for,” and the carer promotes such flourishing by attending to the “genuine needs” and “legitimate wants” of the cared-for, where genuine needs must have “both an objective and a subjective basis” and legitimate wants are limited to those that can be satisfied without harming others (139).

To care for someone, then, I cannot just do what I think is good for them; I must do what they themselves take to be good for them; but differently, I must care about their CARES. This is what makes it clear that the “completion of care”—namely, its reception as care—is necessary (though not sufficient) for something to count as care in the normative sense. If the purported “care” is not something that is at least eventually welcomed or “taken up” by the intended recipient, then it becomes evident that it was not something subjectively valued and did not contribute to their flourishing. For recipients of care that are not subjects, one just has to observe what seems to make them flourish—such as when one waters a plant and sees that this keeps it from wilting. But when the recipient of care is a subject, there must be some form of subjective endorsement of the care as care (where what counts as endorsement varies depending on what sort of agency the subject is capable of exercising). This means that care is not always possible—such as in the case of an intended recipient who refuses to accept any care. It also means that luck plays a large role in whether or not one can successfully care, because there are some unavoidable epistemic limitations to intuiting or predicting what will end up counting as care. When the recipient of care is a non-verbal subject, it can be hard to know what they do and do not endorse. Furthermore, “it may be questionable whether the endorsement is something that we ought to respect when it appears to go contrary to the person’s best interests. A resistance to an action intended as care may be due to a failure to comprehend the nature of the care” (200). In clear-cut cases, one should act to promote well-being even if the recipient does not welcome the care (though they still might eventually appreciate it in retrospect); for instance, a child, or an adult with cognitive impairments, might not understand that taking medicine will actually promote their well-being long term. But what Kittay highlights is that many cases are not clear-cut and it is difficult for someone who intends to care to avoid mistakes: How seriously should one take the wishes of someone who is limited in what they can understand (either from immature or impaired cognition or judgment)? Should their wishes be overridden whenever they seem to conflict with some objective list of what is necessary for flourishing? The danger here is of paternalism that overreaches, or of substituting one’s own subjective sense of what the good life is for that of the recipient.

What I find most illuminating about Kittay’s newest contribution to care ethics is the way that it dissolves what I had previously understood to be a conflict between the value of autonomy and the value of care. According to my prior view, respecting someone’s autonomy would frequently come into conflict with caring for them. I had
in mind, for instance, a non-disabled teenager who does not yet reliably act on sound judgments about what is good for them, due perhaps to their inexperience, or to impulsiveness, or to their susceptibility to the influence of others. Then, I thought, a parent must choose between either failing to respect the teenager's autonomy by not permitting them to make their own decisions, or failing to care adequately for them by not protecting them from their own bad decisions. But at least, I thought, by violating their autonomy a parent could adequately care for them. Of course, such scenarios are possible—the teenager might, years later, engage in the “completion of care” by coming to acknowledge how the parent’s decision contributed to their flourishing. But Kittay alerted me to a third, and very likely, possibility: the parent’s version of flourishing and the child’s might genuinely differ, and not because the child’s version is deficient in any way; the parent might act to promote the version of flourishing that they themselves endorse, instead of the version of flourishing that their teenage child endorses (now or ever). Thus, the parent might intend to care but instead undermine flourishing as endorsed by the one who is to be cared for. Not only has the parent failed to respect the child, they have also failed to care.

At the core of this insight is the dependence of value on valuing: what is valuable in life depends on what we care about, and we do not all have the same CARES. Perhaps the parent of a child who in some clearly manifest way falls “far from the tree”—to use Andrew Solomon’s (2012) metaphor—is forced to recognize this when they confront their child’s difference from the “normal,” though as Kittay points out the news of one’s child’s difference is not always welcome and may be experienced as “a profound shock . . . an upheaval of everything” (27). But for a parent as loving as Kittay clearly is, the shock wears off and the flourishing of the child—flourishing that is endorsed by the child in whatever way is possible for them, and that may be quite different from what flourishing is for the parent—is what comes to matter, thus making care, in Kittay’s fully normative sense, possible. Parents whose children appear more similar to themselves may be slower to grasp this basic point because the child’s similarity to them may support the parents’ illusion that they can substitute what they themselves value for what their child values. But if they never grasp it, they will miss out both on the joy of the completion of care—its reception as care, which signals successful caring—and on the anguish of becoming aware of their own unwitting failures of care.

REFERENCES

Overcoming Epistemic Injustice: Social and Psychological Perspectives
Edited by Benjamin R. Sherman and Stacey Goguen (New York, Rowman and Littlefield, 2019).

Reviewed by Claire A. Lockard
LOYOLA UNIVERSITY CHICAGO

In a situation of oppression, epistemic relations are screwed up.

– José Medina

Although editors Benjamin R. Sherman and Stacey Goguen do not frame Overcoming Epistemic Injustice as a response to José Medina’s characterization of the state of many epistemic communities, I read the book as offering a set of responses to the screwed up nature of epistemic relations. Some contributors to the edited volume detail the ways in which epistemic relations are even more screwed up—and difficult to remedy—than we might at first think. Others develop analyses that offer finer-grained detail about the specific mechanisms by which epistemic relations are screwed up. And others offer strategies for epistemic agents to begin repairing these screwed up epistemic relations.

In their introduction to the book, Sherman and Goguen cite two concerns that, in their view, are discussed in epistemic injustice literature, but do not always come into focus as sharply as they should. First, they claim that discussions in epistemic injustice literature sometimes focus on uncovering and analyzing conceptual and metaphysical problems at the expense of focusing on what, in practice, we might do to ameliorate these problems (8). Although these conceptual and metaphysical problems are not left aside in the volume, Sherman and Goguen explain that “we were motivated by a more direct interest: we want to do better, and we solicited chapters that offer suggestions on how” (8). Second—and following from this desire to “do better”—each chapter in the volume deals, “at some level or another, with empirical research, to better understand what goes wrong, and how to improve (8).” “Empirical” is fairly broadly construed; Sherman and Goguen explain that “authors were invited to tie their ethical and epistemic arguments to empirical work and case studies” (11). Underlying both of these motivating concerns is the assumption that epistemic injustice can be, to some degree, overcome—an assumption that is reflected upon and problematized by various contributors throughout the book.

Overcoming Epistemic Injustice is organized into four parts, each of which explores problems of epistemic injustice in different contexts: in our own psychological tendencies, in healthcare systems, in legal and carceral institutions, and in education and sports. The contributors address a broad array of topics, including but not limited to whether individual epistemic agents can ever adequately assess and correct our own biases and prejudices; the utility of Miranda Fricker’s account of epistemic virtue for resisting
epistemic injustice; the intersections between epistemic injustice literatures and conversations in psychology, phenomenology, critical race theory, and care ethics; and the limitations of epistemic injustice for theorizing the extent to which our systems of knowledge rely on the existence and perpetuation of ignorance, unknowing, and forgetting.  

In my review, I follow the book’s two central organizing principles in order to tease out elements of the volume that perform this “doing better” and to identify places where the volume might have gone further in its aims. The book’s authors offer wide-ranging analyses and examples of epistemic injustice, with many different approaches to diagnosing and treating the problems they explore. In my view, Overcoming Epistemic Injustice offers its most generative conceptual and practical resources when contributors problematize the volume’s own explicit aims of overcoming, in favor of lingering with the unjust ways of knowing, forgetting, and learning that persist even—and perhaps especially—when we work to uncover them.

“DOING BETTER” AS A FRAMING DEVICE

After finishing the book—and particularly in light of the contributors’ accounts of epistemic injustices in so many different contexts and cases—I wondered about the framing put forth in the introduction. Can epistemic injustice be overcome? Many scholars of epistemic injustice, both within and beyond this volume, answer in the negative, while refusing to let us rest easy with this pessimism.  

Who is overcoming epistemic injustice: privileged knowers who perpetuate epistemically unjust knowledge practices, or marginalized knowers who must navigate and resist these practices? How can or should conversations about these types of overcoming interact? Contributors take different approaches to who is overcoming—for some, the overcoming agent is the privileged knower who learns to recognize and ameliorate credibility deficits, help fill hermeneutical lacunae, and generally work toward epistemic justice within their own knowledge practices and those of their communities. For others, the overcoming agent is the victim of epistemic injustice, who develops strategies and tools for resisting the harms perpetuated by unjust knowledge practices. These different approaches to the notion of “overcoming” are perhaps part of a broader—and productive—tension in epistemic injustice literature between the need to resist the kind of optimism that (re)produces unjust epistemic norms and practices, and the need to recognize opportunities for exposing and ameliorating these very norms and practices. And there is, of course, an important distinction to be made between overcoming an injustice and getting better at resisting it. Still, I worry that the book’s move toward improvement and progress, both in its framing and in some of its content, risks recentering privileged subjectivities and offers a comforting narrative of progress at the expense of exploring themes of epistemic resistance in more detail, or analyzing structures that actively work to preserve epistemic injustice.

In her contribution to the volume, “The Episteme, Epistemic Injustice, and the Limits of White Sensibility,” Lissa Skitolsky explores precisely this worry. In her view, contrary to the claims made by many scholars of epistemic injustice, “we do not need to perfect our own power of moral perception” (206) in order to resist epistemic injustice; in fact, unjust knowledge systems construct subjects as limited knowers who cannot hope to fully-recognize the ways in which our sensibilities are formed by them (206). It might be that testimonial injustice runs far deeper than much of the literature on epistemic injustice has acknowledged—power relations and social oppression are already operating at the level of our perception and social imagination (211). Thus, for Skitolsky, self-reflection and self-correction are far more fraught tools for correcting epistemic injustice than we tend to assume. She contends that “the problem of epistemic injustice is not essentially related to a lack of knowledge but rather to an affective disposition that is reinforced by our perception of what is ‘real’” (212). Rather than focusing on our own progress toward testimonial justice, people with structural privilege should work to get better at sensing their/our own epistemic, interpretive, and perceptual limits (212). To adopt Skitolsky’s framing, my concern is that questions about the impossibility of overcoming epistemic injustice—and the injustices that we risk by assuming that we can—are underexplored in this volume.

There are, however, some notable exceptions. Indeed, one difficulty of reviewing this volume is the impossibility of generalizing about its contents—this might also be one of the volume’s strengths. In her afterward to the book, Miranda Fricker, whose work grounds a great many of the contributors’ analyses, wonders about the overcoming of epistemic injustice. In her view, “an in-principle pessimism is in order as regards this broader ideal of epistemic justice, for only pessimism will keep us safely alert to the ordinary injustices that, I believe, are bound to reoccur and reinvent themselves in new contexts and new mutated guises” (304). Some contributors to Overcoming Epistemic Injustice remind readers that even if we want to claim, perhaps contra Skitolsky and Fricker, that epistemic justice is possible within our current episteme, there may be some specific epistemological contexts that foreshadow the possibility of overcoming. For instance, in their contribution to the volume, “Carceral Medicine and Prison Abolition: Trust and Truth-Telling in Correctional Healthcare,” Andrea Pitts contends that carceral systems distort the capacity for trust between incarcerated individuals and their healthcare providers (226)—these “patterns of distrust found in correctional healthcare settings cannot be alleviated by merely addressing the interpersonal dynamics between patients and providers” (231). Instead, Pitts argues for a prison abolitionist model in order to address these patterns and to develop new strategies for broader resistance to carceral systems. While I do worry about the collection’s emphasis on overcoming and getting better, I appreciate the moments in the text where overcoming and improving are rethought and reframed as opportunities for transformation and reimagination.

A PLURALISTIC APPROACH TO THE EMPIRICAL

In addition to providing strategies for ameliorating epistemic injustice, the contributions to Overcoming Epistemic Injustice also offer empirical evidence to support the claims and arguments being made. The volume takes quite a broad approach to what it means to offer empirical evidence to support the claims and arguments being made. The volume takes quite a broad approach to what it means to offer empirical evidence to support the claims and arguments being made.
that emerge when philosophers are drawn to this data at the expense of engaging with literature that centers people of color’s experiences with racism. She cites work by Patricia Hill Collins, George Yancy, Nancy McHugh, Charles Mills, and María Lugones as offering these conceptual resources—some long before implicit bias literature emerged. Davidson worries that the privileging of implicit bias research over accounts like those offered by the theorists above functions as a way to avoid engaging with testimony that people of color have long been offering about the experience and operation of racism, in favor of using a particular kind of scientific evidence to think about white people’s racial biases (274). Reading Dotson and Davidson’s work together provides an example of an application of Dotson’s account of epistemic exclusion; furthermore, it illustrates a commitment to addressing epistemic injustices within the field of epistemic injustice research itself.

The methodological pluralism of the volume could have been even further enhanced by what might be called citational pluralism. Although contributors work from a variety of different philosophical traditions, many essays take Miranda Fricker’s Epistemic Injustice as their theoretical starting point. Although Fricker’s work has been highly influential in thinking about epistemic injustice and oppression and many authors frame their analyses using additional resources that complement and complicate Fricker’s account, Fricker’s dominance in the volume is striking. Sherman and Goguen point out in their introduction that there is a long history of work on questions about the social, ethical, and political stakes of epistemology, and that this can be found in a variety of philosophical traditions (1). While Fricker’s importance in framing contemporary discussions of epistemic injustice should not be overlooked, the reliance on her work already sets contributors along particular paths, at the expense of engaging additional (and in some cases, marginalized) resources.

This repeated citation of Fricker is not the result or fault of any individual author; instead, it is a reflection of the broader epistemic injustice literature. Veronica Ivy [Rachel McKinnon] reminds us of what is at stake when we repeat what Sara Ahmed calls “citational paths” that begin with Fricker (and no earlier):

while Fricker is widely acknowledged as the first to name and circumscribe epistemic injustice in these terms—and there’s been remarkably wide uptake of her view—there’s a long history in black feminist thought, and other feminists of color, that should be seen as also working on issues of epistemic injustice.13

My comments here are not intended to be critical of Fricker’s work specifically, nor are they an indictment of this volume’s editors or contributors; rather, they are an invitation for those of us writing on questions of epistemic injustice to engage with the broader (and earlier) literatures on the topic. I have in mind, for example, Mariana Ortega’s 2006 essay “Being Lovingly, Knowingly, Ignorant: White Feminism and Women of Color.” Ortega worries about the
active ignorance of white feminists, despite women of color—like Audre Lorde and María Lugones—pointing out the whiteness of many feminists’ conceptual frameworks, questions, knowledge claims, and methodologies. I also have in mind Kristie Dotson’s “A Cautionary Tale: On Limiting Epistemic Oppression”—here Dotson cites women of color who discussed “the problem of biased hermeneutical resources” long before the terms “epistemic injustice” or “hermeneutical injustice” were coined, including Patricia Hill Collins, Gayatri Spivak, and Patricia Williams (44).15

QUESTIONS OPENED BY OVERCOMING EPISTEMIC OPPRESSION

Where in our lives and communities do we find epistemic and hermeneutic injustices? How do testimonial and hermeneutic injustices operate in our healthcare, educational, and carceral systems? How can individual knowers work against our own biases in order to work toward testimonial justice? How do epistemically unjust systems interact with the epistemically unjust behaviors or attitudes of individuals? What does (or should) it mean to “overcome” epistemic injustice? Why might the very assumption that privileged knowers can perform this overcoming be, itself, fertile ground for epistemic injustice to reassert itself?

These are only a few of the questions posed and explored by the contributors to Overcoming Epistemic Injustice. While the book’s orientation toward overcoming is one that, in my view, ought to be problematized (and indeed, it is problematized at several important points by several different authors within the volume itself), its uses of case studies and examples make it a generative resource for those seeking on-the-ground accounts of identifying and beginning to address, as Medina puts it, screwed up epistemic relations. These questions are far from settled by the book, and this opens space for further philosophical investigation of the ways in which epistemic oppression, violence, and injustice structure our lives—and how these injustices might be rendered otherwise.

NOTES

5. See McWilliams’s contribution, "Can Epistemic Virtues Help Combat Epistemologies of Ignorance?"
8. See Skitolsky’s contribution.
9. See, for example, Medina's The Epistemology of Resistance, or Kristie Dotson’s "A Cautionary Tale: On Limiting Epistemic Oppression."
10. There are many additional approaches to the empirical that I am unable to summarize in detail here. I have cited a few here that, in my view, illustrate the wide range of approaches taken by the volume’s contributors.
15. See also Ortega’s bibliography for “Being Lovingly, Knowingly Ignorant,” as well as Veronica Ivy [Rachel McKinnon]’s endnote 7 in “Epistemic Injustice" for additional examples of women-of-color feminists exploring questions of epistemic injustice and ignorance.

BIBLIOGRAPHY


Full Surrogacy Now: Feminism Against Family


Reviewed by Jina Fast

NOTRE DAME UNIVERSITY OF MARYLAND

Sophie Lewis in her book Full Surrogacy Now: Feminism Against Family doesn’t seek to enter the feminist debate on surrogacy, but rather to completely reset the terms of it. Over the past fifty years, the ethics of surrogacy have often focused on two aspects: the rights of couples who desire access to reproductive technologies to extend genetic lineage and construct family on the basis of shared biology, and the economic and bodily rights of the surrogates who make possible said reproduction for the intended parents. Often these rights are posed as being in an inherent tension, which considerations in ethics are meant to mitigate or solve as best they can. Lewis, by contrast, working from a queer, cyborg, feminist perspective and within a reproductive justice framework, rejects the possibility of an ethical surrogacy practice within a cissexist, heterosexist, capitalist, and anti-black racist global system. Yet, she argues not for its abolition, but for its radical expansion to build solidarity between paid and unpaid gestators. Practically, radical expansion, Lewis maintains, would better approximate access to reproductive
justice for surrogates through making concrete the practice of “surrogates running surrogacy,” which is rhetoric often deployed by fertility clinics without concrete realization. However, the desired end of surrogates working without bosses is not the limit of Lewis’s argument; rather, she sees this as a step in the ultimate project of dismantling capitalism and its alienating effects on the many forms of labor marginalized groups perform within this economic model. For Lewis, the abolishment of capitalism would eradicate the very meaning of the word “surrogate” as the boundaries between the original or intended parent and the surrogate or substitute parent are made fluid (145).

Importantly, while Lewis argues that the practice of surrogacy requires radical transformation, she is quick to reject the radical feminist approach steeped in savior politics. Surrogates, for one, do not empirically benefit economically, legally, mentally, or physically from the banning of the practice of surrogacy. Rather, the effects, unintended or not, tend to further obscure the work that surrogates do and compromises their lives as they are subjected separation from their families, forced travel across borders, inductions, and c-sections in order to fulfill a market demand amid restrictions on their economic opportunities. Centered here in Lewis’s concerns are the people who work as surrogates, but her argument quickly expands as surrogacy functions primarily as a metaphor for the way kinship under capitalism is structured by possession.

The strengths of Lewis’s text are many and include (1) her analysis and rhetorical disruption of the language surrounding surrogacy, which has failed to be seriously considered in other contemporary works on the ethics of surrogacy, (2) the novel application of a queer, cyborg feminist theory to the issue of surrogacy, and (3) the extent to which she enables thinking anew about heterosexist and patriarchal conceptions of the family freed from a capitalist and neo-liberalist logic. Regarding the first, Lewis takes as her target arguably the most famous surrogate center in the world, Dr. Nayna Patel’s Akanksha Infertility Clinic in Anand, Gujarat, India. Patel’s work has been previously highlighted in both popular coverage and scholarly work. For example, she has been interviewed on the English BBC World program HardTALK, featured on a Vice News segment, and centered in the documentaries Google Baby (2009) and House of Surrogates (2013). In terms of scholarship, Lewis relies on the extensive ethnographic work of scholar Amrite Pande reading her interviews with the surrogates who live and work at Patel’s clinic with the language Patel uses to describe the work of surrogacy and herself as the lead clinician. Here, Lewis reveals the many contradictions and obscurities linguistically performed that have semiotic and concrete effects. For example, Dr. Patel frames herself as choosing the work of surrogacy in order to facilitate “women helping women,” but her clinic is for-profit. Thus, the work at Akanksha is not charitable, but is oriented toward the goal of creating profit for the owners of the clinic, of which Patel is one. Yet, by using this language repeatedly and having others endorse her through this rhetoric, Patel is able set her clinic up as different from other surrogacy clinics that are “womb-farms” (91). Furthermore, through framing in pseudo-feminist language the surrogate process as “women helping women” she is able to “ritualistically unburden her customers’ doubts about the ethics of the exchange” (91). One does not have to be a Surrogate Exclusionary Radical Feminist to have doubts about surrogacy, as the racialized, classist, and nation-state of origin global politics can be quite obvious. Paying for surrogacy is expensive, thus one’s socio-economic class, which is further tied to racialized identity, one’s nation-state of origin, gender, and even religious and ethnic identity (for example, in India many of the surrogates hired by Hindu couples are Muslim). Yet, the language deployed here does even more work by obscuring surrogates’ labor as work by reframing it as “helping” and reinforcing genetic off-spring as a gift beyond quantification. That both of these notions are patently false is both beyond and undoubtedly the point. The surrogate is paid, though arguably not fairly, and there is a market for surrogacy and other forms of reproductive technologies, which shows the value of genetic offspring is monetarily quantified within a capitalist economic order.

The second notable strength of Lewis’s work is her use of queer theory and cyborg feminism to show that radical feminists’ efforts to thwart surrogacy is theoretically incoherent and practically disastrous. Queer theor(ies) are especially suited to support Lewis’s argument that surrogacy, like pregnancy and sex work, can be experienced in multiple ways. It is far from, as radical feminist groups like FINRRAGE claim, always experienced as oppressive and constituted by a loss, namely, loss of the child once delivered. Here, Lewis also extends a common critique of the white feminist framework of Margaret Atwood’s famous text The Handmaid’s Tale (1985), which serves the anti-surrogate pseudo-feminist message through rendering all surrogacy as enforced and indicative of oppressive gender politics. Yet, as is the case in most components of human life, there are multiple meanings possible at once constituting the same act or experience, thus universal and universalizing claims lead to misunderstanding and harmful effects. For many surrogates, pregnancy and childbirth are literal work for which they are paid, and the fetuses gestated are never imagined to be theirs to care for in the way that they care for the children in their families. Pande’s scholarly work complicates this notion a bit in that surrogates interviewed often imagine some kind of relationship with the child postpartum, though not a relationship akin to “parent” or “mother,” but a relationship that will not materialize, nonetheless. But this doesn’t necessarily hurt Lewis’s argument in that the notions of “mother” and “parent” themselves are not naturally determined relationships and contain within them contradictions and multiple meanings. For one, a child might have multiple mothers, or the role of “the mother” might be split between multiple people. Thus, on Lewis’s reading the surrogate exclusionary radical feminist (SERF) argument reinforces at least three problematic dictums.

First, it reinforces the idea that surrogacy is not “work” and thus erases the labor that constitutes gestating and birthing. Inevitably, this reinforces part of Dr. Patel’s marketing concept, namely, that surrogacy is a “different kind of work” and a “labor of love.” Second, SERFs re-naturalize parenthood as a cis-hetero project that entails cisgender
women gestating “their own” fetuses. Finally, the common SERF refrain reinforces a white savior politics, and SERFS themselves as white saviors, as it sets up brown and black surrogates of the global south as victims unable to act or make choices for themselves and their families. Lewis’s work, by contrast, provides space for a consideration of the voices of surrogates themselves, who do not understand themselves as oppressed or as victims in need of being saved.

For solutions to these problems, Lewis reads theorists like Donna Haraway (1991) and Maggie Nelson (2015) together and argues that we first must accept that authorship can only ever be co-authorship with others and even with technologies. One never makes themselves or undertakes a project solely on their own. Rather, it requires being made by and with others as we learn from others, co-operate with others, and recognize others. Understanding pregnancy and the care of children in this manner would, Lewis asserts, remake in revolutionary fashion our collective understanding of responsibility, family, parenthood, and even gender. Pregnancy here is understood as always being a form of alienation (128) whether chosen, accidental, or surrogate; but it is an alienation that is simultaneously an intimacy. One is alienated from one’s body as one becomes ever closer to the goings on of the body; one becomes closer to one’s projects as one is split from oneself quite literally in becoming multiple. It is, to put it succinctly, an inherently queer experience. But it is a queer experience that can be aided by and the pain lessened through technological advancement. This last point, that pain can be lessened through advancements in reproductive technologies and science, Lewis notes, is especially prescient as we consider the discrepancies in care to which white pregnant people and black pregnant people have access and the effects of such discrepancies that manifest well beyond the experience of pregnancy.

Finally, while Lewis provides further evidence that the family itself and the work that goes into producing “family” can be varied and multiple, she simultaneously argues for the destruction of the family. The family, for Lewis, is the space where capitalism, neo-liberalism, heterosexism, and patriarchy intersect necessarily to maintain the oppression and alienation of marginalized groups. Queer reformation strategies that focus on remaking the family through laws protecting queer unions, childbearing, and kin-making will not be enough in the end to transform the oppressive politics of capitalism as long as the family remains tied to notions of privacy and ownership. For Lewis, the only way to end exploitative surrogacy is to reimagine these ideas of privacy and ownership, which themselves can be abolished only concomitantly with the abolition of capitalism.

While I thoroughly enjoy Lewis’s work in *Full Surrogacy Now* and her arguments are sound, her considerations at times fall short in their practical application to the work of peripartum gestation and care. As revealing as it is to follow the author in her deconstructive analysis of the performative language surrounding surrogacy, as well her positive argument that frames all work as gestational, it doesn’t follow that there are multiple people who can do (some of) the work of literal gestation of a fetus and postpartum breastfeeding. Pointedly, the collapse of the “can be” or “ought” and the “is” here is somewhat troubling. To be fair, and with full disclosure in mind, perhaps part of my hesitancy to accept that the work of gestating fetuses can be, and thus is, like any other work is due to the fact that I have recently participated in the gestational work of pregnancy and am still engaged in the work of breastfeeding. It doesn’t feel like work that could be done by another, and while support makes each of these forms of work easier, there is a materiality to the embodied experience of peripartum work that is arguably missing here.

Nevertheless, Lewis has a way out. Through the construction of her argument, she can hypothetically respond to any practical or material concern by stating that the objector simply is unable to imagine the world—and thus surrogacy as practice—freed from the constraints of capitalism, neo-liberalism, and heteropatriarchy. For example, Lewis may counter that my concern mentioned here simply reveals me to be like her father attached to the notion of shared genes as the source of family. I, she may note, am unable to conceive of the family, parenthood, care work, and so forth outside of the oppressive ideological and structural constraints that contemporarily bind and obscure the possibilities of imagination. And to be reflective, perhaps this is true, in the same way that it may be impossible to conceive the being of womanhood outside of patriarchy, or what blackness is outside of the constraints of white supremacy. The fact that projects exist with the intention of imagining womanhood outside of the male gaze, and blackness outside of the white gaze, suggests that surrogacy, parenthood, and gestation in general can be reimagined and remade as well.

**REFERENCES**


### Reproductive Justice: An Introduction


Reviewed by Kathryn Lafferty-Danner

**UNIVERSITY OF LOUISVILLE**

Reproductive Justice: An Introduction, by Loretta J. Ross and Rickie Solinger, provides a comprehensive and compelling argument for using a reproductive justice framework to address issues that incorporate reproductive policies but extend beyond the “pro-choice/pro-life” dichotomy that is typical in American cultural discourse. Beginning in the 1990s, women of color and low-income women began to push back on the use of the word “choice” within white feminist discourse, demonstrating the serious limitations to the “choice” framework for many marginalized communities.
Also, the “choice” framework, as it has been used, focused nearly entirely on preventing motherhood and less on the full spectrum of reproductive health care services and the intersecting issues that surround reproduction. Combining their scholarly and activist backgrounds, Loretta J. Ross and Rickie Solinger work to create a useful introduction to the topic for scholar and activist audiences alike, crafting a complex, comprehensive, and intricate theoretical framework with a practical application. Beginning in 1994, twelve women of color formed an alliance called Black Women on Health Care Reform and began vocalizing the myriad ways in which white feminism dominated discussions in the pro-choice movement and in 2003, SisterSong Women of Color Reproductive Health Collective held a national conference where they sought to define reproductive justice. Reproductive Justice: An Introduction builds off of the work set out in the first book about reproductive justice, Undivided Rights: Women of Color Organize for Reproductive Justice (2004), written by Jael Silliman, Marlene Gerber Fried, Loretta Ross, and Elena Gutiérrez.

Ross and Solinger define reproductive justice as having three primary principles: “(1) the right not to have a child; (2) the right to have a child; and (3) the right to parent children in safe and healthy environments” (9, emphasis original). The term “reproductive justice” comes from combining reproductive rights and social justice, creating a coherent and wide-ranging term that encompasses a variety of social justice concerns that intersect with reproductive rights issues. With this definition, Ross and Solinger extend the very limiting discussion of pro-choice/pro-life political and social discourse well beyond the common conversation that focuses on individual or personal choices, and demonstrate the ways in which reproductive policies and decisions dictate how whole communities are impacted by these policies. Reproductive justice focuses on communities having the resources to not only have an abortion and access to contraception, but also for individuals to have jobs where they are paid a living wage, live in environments that are free of racism, and enact laws and policies that take into consideration how some populations face multiple types of oppression simultaneously.

Reproductive justice incorporates an intersectional perspective and focuses its theoretical lens on fundamental human rights that highlight and, at times, resist policies that are based on gender, racial, and class prejudices. Furthermore, a reproductive justice lens not only focuses on individual experiences of those who are the most marginalized, but on the ways in which state strategies and policies affect communities over time. This lens situates reproductive issues within both an individual context as well as within a larger social and historical context where they occur. Ross and Solinger’s examples illustrate how reproductive policies have affected whole communities, such as the racialized practice of sterilization and the impact the Hyde Amendment has had on lower-income women.

In the first chapter, Ross and Solinger detail the foundational theory of reproductive justice, showing the ways in which it utilizes and expands the human rights framework, and demonstrate how fundamental human rights are not separate from reproductive health care. Building off of a broadly defined human rights framework, Ross and Solinger argue that the right to bodily self-determination is perhaps “the most foundational human right” (56). Reproductive justice differs from reproductive health and reproductive rights in that it calls for us to identify how reproductive oppression is the result of multiple, interlocking oppressions and is intertwined with the global human rights system as a legal framework. They argue that healthcare, including reproductive healthcare, is not a “commodity for purchase” (17) but is a human right that is foundational to bodily self-determination.

The second chapter describes the challenges and possibilities for building on current reproductive justice work and highlights some of the ways in which reproductive justice can help build an inclusive, unified human rights movement. Ross and Solinger claim that personal, firsthand storytelling within reproductive justice acts as a powerful tool, providing insight into an individual’s unique situation and that can help us to understand how others think as well as make decisions. They use the metaphor of “shifting lenses” (59) to describe the ways in which no single story can describe everyone’s experience and discuss the idea of “polyvocality” (59), or using many voices with different perspectives that come together to make a unified movement for human rights activism and theory. Historically, narratives have been used to give voice to those who are often silenced by racial or gendered oppression, such as women speaking out about rape in the 1970s that started a global movement to end violence against women. Reproductive justice advocates the use of narratives as a “key strategy for making social justice claims to change the world” (60) by building powerful coalitions through storytelling. There is also a useful discussion of the differences between reproductive rights, reproductive health, and reproductive justice, noting how reproductive justice calls for a holistic and integrated analysis that pushes against the structural conditions within which individual bodies, labor, sexuality, and reproduction are regulated (68–73). The reproductive justice framework centers on the power dynamics that are at play and seeks to work across social justice issues, demonstrating the capacity for reproductive justice to be thought of as not only focusing on reproductive concerns, but also on other issues that intersect with reproductive issues, creating a multi-racial, multi-issue, and polyvocal movement.

Ross and Solinger move into discussions of contraception and access to reproductive health in the third chapter, providing further distinctions between reproductive justice and other reproductive frameworks. They discuss how Roe v. Wade, often hailed as one of the largest wins for feminists, is couched in privacy claims and is considered a “negative right,” the right to be left alone by the government and for offices not to interfere in personal healthcare decisions. Reproductive justice seeks to reframe access to reproductive health services as a positive right, a right that includes not only choice in medical decision-making, but that also extends this positive right to reproductive services outward, to networks of support that provide all women with the ability to exercise their right to an abortion.
Ross and Solinger take on the legal aspects of abortion as well and note the ways in which legislation leaves out the social environments from which these discussions stem. For instance, legal aspects of abortion fail to recognize how economically burdensome obtaining an abortion in the United States can be for low-income individuals. Instead of insisting on the Court creating rights for humans, the authors contend that human rights are natural and inherited, simply based on the fact that we are all human. To demonstrate this claim, Ross and Solinger explain that reproductive human rights start with the “acknowledgement that a person has an inherent human right to control her own body and then seeks to use the political process to express this right and the judicial process to protect this right” (128). In other words, a reproductive human rights framework challenges the belief that rights begin with the judicial system and argues that social justice movements influence laws first. Instead of rooting the reproductive framework in the law, their position is that the most effective way to build reproductive autonomy is through the use of grassroots organizing, coalitions with other social justice-oriented organizations, and forming alliances across race and class lines. The Hyde Amendment, which prohibits public funding for abortion and was passed shortly after Roe, is one example they use to demonstrate the ways in which politics has more influence in health policy than actual health care and medical science. In addition, Ross and Solinger argue that accessibility is key to high standards of health care and that accessibility comes in the forms of physical accessibility, non-discriminatory services, economic accessibility, and the accessibility of information about those services. They demonstrate the ways in which these issues that are discussed within reproductive justice do not only focus on how it impacts the individual—although personal storytelling is one aspect—but on how these issues impact communities as a whole.

The fourth chapter focuses specifically on the right of individuals to parent, particularly the right to parent their children in safe and healthy environments. Ross and Solinger demonstrate the ways in which giving birth has become less of a private affair and that motherhood has become more of a public concern. For instance, legal aspects of abortion fail to recognize how women of color and low-income women centering other social justice movements. Reproductive justice, built on the foundation of human rights, argues that social justice movements influence laws first. Instead of rooting the reproductive framework in the law, their position is that the most effective way to build reproductive autonomy is through the use of grassroots organizing, coalitions with other social justice-oriented organizations, and forming alliances across race and class lines. The Hyde Amendment, which prohibits public funding for abortion and was passed shortly after Roe, is one example they use to demonstrate the ways in which politics has more influence in health policy than actual health care and medical science. In addition, Ross and Solinger argue that accessibility is key to high standards of health care and that accessibility comes in the forms of physical accessibility, non-discriminatory services, economic accessibility, and the accessibility of information about those services. They demonstrate the ways in which these issues that are discussed within reproductive justice do not only focus on how it impacts the individual—although personal storytelling is one aspect—but on how these issues impact communities as a whole.

The epilogue to this text is useful to activists who want to get a closer look into some of the ways in which reproductive justice can be implemented within grassroots organizations. Ross and Solinger blend historical examples, legal rulings, personal stories, and a thorough explanation of the reproductive justice framework to alleviate that overwhelming feeling. However, Ross and Solinger further explain that poor women are often seen as being poor decision-makers and, therefore, are denied dignity and shown to not be deserving of the same sexual and biological opportunities as others who are not poor. In this chapter, they focus heavily on five ways in which the reproductive justice framework intersects with other social issues, discussing immigration, incarceration, housing and gentrification, education, and environmental issues as all being related to reproductive justice. The authors do an excellent job explaining the ways in which reproductive justice is not a single issue; it extends out into countless other social issues and cannot be considered separate from other large, structural concerns in society.

This book is written for both activists and scholars. For activists, the book provides strategies and practices “to build a unified, radical, and inclusive movement” (58) with clear examples from successful organizations and practical uses for implementing a reproductive justice framework into community organizing. For scholars, this introduction is a useful resource in discovering ways to incorporate the reproductive justice framework into research and scholarship, appealing to both new and veteran scholars in a variety of fields, from public health to history. As a framework, reproductive justice is a tremendously expansive term that encompasses so many social justice issues to adequately cover in an introductory text; at times, it can become overwhelming for the reader. However, Ross and Solinger blend historical examples, legal rulings, personal stories, and a thorough explanation of the reproductive justice framework to alleviate that overwhelming feeling. They ask their audience to expand on the framework and incorporate it into research, community organizations, or other social justice movements. Reproductive justice, built by women of color and low-income women and centering
these marginalized voices, has become a new and useful way to envision radical and important changes to how we address the current issues we face as a society and how we can further human rights protections for all individuals as well as their communities.

ANNOUNCEMENTS

NEW APA BLOG! WHAT IT’S LIKE TO BE AN HISTORICALLY BLACK COLLEGE OR UNIVERSITY (HBCU) FACULTY MEMBER

Keep an eye out for this new APA Blog!

The purpose of the APA Blog on What it’s like to be an HBCU faculty member is to spotlight faculty members in our profession who work at historically black colleges and universities, which are minority serving institutions. The blog also aims to introduce and familiarize faculty members at majority-serving institutions with the distinct perspectives, experiences, and, sometimes, limitations, that are unique to students and faculty at historically black colleges and universities. If you have a recommendation for a faculty member at an HBCU, or other minority-serving institution, who deserves recognition, please feel free to nominate them!

https://blog.apaonline.org/2020/02/03/what-its-like-to-be-an-hbcu-faculty-member-brandon-hogan/

FAB 2020

FAB 2020 is the 13th World Congress of the International Network on Feminist Approaches to Bioethics. Due to the COVID-19 pandemic, FAB 2020 will be moving to a virtual platform.

Our World Congress brings together academics and practitioners from diverse backgrounds to explore and develop feminist perspectives on ethical issues relating to health. It is a vibrant and welcoming gathering that offers significant opportunities to share ideas, learn from others, participate in constructive debates and engage with a supportive network.

The FAB 2020 congress theme is “Feminist Perspectives on Solidarity and Autonomy.” We anticipate a diverse array of stimulating presentations and discussions on this theme. There will also be opportunities to discuss feminist investigations in other areas of bioethics.

For more information, visit the FAB website at https://www.fabnet.org/.

CONTRIBUTORS

Saray Ayala-López works as an assistant professor of philosophy at California State University, Sacramento. Their current work applies conceptual tools from different areas of philosophy (philosophy of mind and language, and epistemology), together with empirical insights from cognitive psychology, to understand morally interesting questions. They are especially interested in explanations (structural explanations are their favorite) and the absence of them, sex and gender in science, conversational dynamics and the many things we can do with words, cognitive externalism, conceptual ethics, and several questions within social ontology and epistemology.

Amy Allen is liberal arts professor of philosophy and women’s, gender, and sexuality studies and head of the philosophy department at Penn State. She is the author of four books, including, most recently, The End of Progress: Decolonizing the Normative Foundations of Critical Theory (Columbia University Press, 2016) and Critical Theory between Klein and Lacan: A Dialogue (co-authored with Mari Ruil) (Bloomsbury, 2019). Her current research focuses on critical theory and psychoanalysis.

Samantha Brennan is dean of the College of Arts and professor of philosophy at the University of Guelph. She’s a founder and past editor of Feminist Philosophy Quarterly. She works in normative ethics and feminist philosophy, with a special interest in family justice and children’s rights.

Melissa Burchard is professor and chair of philosophy at University of North Carolina, Asheville. Her teaching and research interests include interpersonal violence, trauma theory generally and especially intersections between trauma and mothering, representations of violence in children’s literature, and ethics generally. Recently published articles include “Non-Standard Witness: Lessons from Working with/for Traumatized Children” in Testimony and Trauma: Engaging Common Ground (Brill, 2019); “Not a Matter of Will: A Narrative and Cross-Cultural Exploration of Maternal Ambivalences” co-authored with Keya Maitra, in Lived Experiences of Women in Academia: Metaphors, Manifestoes and Memoir (Routledge, 2018); and “Abandoning Certainty in Favor of Moral Imagination: Shifting from Rule-Based Decision-Making to Caring,” Philosophy in the Contemporary World (Fall 2016). She has published a book with Routledge titled Philosophical Reflections on Mothering in Trauma (2018). Her current project is examining representations of trauma in children’s and young adult literature and popular culture.

Jina Fast is an assistant professor of philosophy at Notre Dame University of Maryland. She is currently at work on a book on Black Caribbean and American existentialism and phenomenology and has been published in the Journal of Critical Race Inquiry and Atlantis: Critical Studies in Gender, Culture, & Social Justice. As a feminist epistemologist and critical race theorist, her work centers on theories produced by and through the experiences and work of marginalized folks across disciplines.
Kathryn Lafferty-Danner is a PhD candidate in comparative humanities at the University of Louisville, where she also received her MA in bioethics and medical humanities. Her research takes an interdisciplinary approach to consider abortion in American culture and specifically, how medical humanities can be used to reduce abortion stigma. She is also an abortion access hotline volunteer with Kentucky Health Justice Network and teaches a range of English and humanities-based courses.

Carol J. Gray is a Crandall Cordero Fellow and PhD candidate at the University of Connecticut, where she studies political science, race and gender studies, and human rights. Her first career was as a public defender following a Prettyman Fellowship at Georgetown University Law Center working in their Criminal Justice Clinic. She was also a National Association for Public Interest Law Fellow at the Georgia Resource Center that represents prisoners on Georgia’s death row in their postconviction appeals. Gray’s dissertation is based on an oral history she conducted in Egypt of one of the country’s leading human rights organizations. She was in Cairo (and in Tahrir Square) during Egypt’s Arab Spring. Gray moved to Egypt with her family as a Rotary International Ambassadorial Scholar to study International Human Rights Law at American University in Cairo. From 2013 to 2014, Gray was a Fulbright Scholar at Concordia University in Montreal, Canada.

Serene J. Khader is professor of philosophy at the CUNY Graduate Center, Brooklyn College and holds the Jay Newman Chair in Philosophy of Culture at Brooklyn College. She is the author of Decolonizing Universalism: A Transnational Feminist Ethic and Adaptive Preferences and Women’s Empowerment, as well as a number of articles in moral and political philosophy. She also co-edited, with Ann Garry and Alison Stone, The Routledge Companion to Feminist Philosophy.

Quill Rebecca Kukla is professor of philosophy and senior research scholar in the Kennedy Institute of Ethics at Georgetown University, and for 2020, a Humboldt Scholar at Leibniz Universität Hannover. They are the editor-in-chief of the Kennedy Institute of Ethics Journal. Their books include Mass Hysteria: Medicine, Culture, and Mothers’ Bodies, and, with Mark Lance, ‘Yo!’ and ‘Lo!’: The Pragmatic Topography of the Space of Reasons. Their latest book, City Living: How Urban Spaces and Urban Dwellers Make One Another, is forthcoming in 2020. They are also a single parent, and a competitive boxer and powerlifter.

Eli Kukla is an undergraduate at McGill University studying linguistics, gender studies and social justice, and urban geography. Currently working for his bachelor’s degree, he also has an associate degree from George Washington University. He is co-authoring a paper, “Moving Through Real and Virtual Spaces,” about the epistemology of space and movement and presented a version of this paper at the 2018 Philosophy of the City conference at Universidad de la Salle in Bogotá.


Claire A. Lockard is a PhD candidate at Loyola University Chicago. She is writing a dissertation about the call for interpretive charity in academic philosophy. She uses feminist philosophy, critical phenomenology, and epistemic injustice literature to explore some of the harmful disciplinary structures that are often reinforced when philosophers call on one another to perform charitable interpretations of problematic texts. She has also published essays on white confessions of racism (2016) and on diversifying the discipline (2017), both of which can be found in Feminist Philosophy Quarterly.

Joseph A. Stramondo is an assistant professor of philosophy and associate director of the Institute for Ethics and Public Affairs at San Diego State University. His writing focuses on the intersection of bioethics and philosophy of disability and can be found in journals like The Hastings Center Report, The Kennedy Institute of Ethics Journal, The International Journal of Feminist Approaches to Bioethics, Science and Engineering Ethics, and Social Philosophy Today.

Lisa Tessman is a professor of philosophy at Binghamton University in New York. She has published in ethics, feminist philosophy, and related areas. Her work focuses on understanding how morality is constructed and how real human beings experience morality, especially under difficult conditions. Her most recent book, When Doing the Right Thing Is Impossible, discusses situations in which no matter what you do, you'll commit some moral wrongdoing. She currently is studying how unavoidable moral failures gives rise to what is called “moral distress” among healthcare practitioners, and the phenomenon known as “moral injury” in a military context. Her previous books include Moral Failure: On the Impossible Demands of Morality (2015) and Burdened Virtues: Virtue Ethics for Liberatory Struggles (2005).

Leigh Viner earned her PhD in philosophy at Duquesne University writing on the Ancient Greek Stoics. She is an assistant professor of philosophy and coordinator of gender studies at Indiana University Southeast, teaching a wide variety of courses, including aesthetics, Asian philosophy, philosophy of gender and sexuality, and, recently, a seminar on Michel Foucault. She is also a musician in the experimental music scene in her hometown, Louisville, Kentucky.
FROM THE EDITORS
Carlos Alberto Sánchez and Lori Gallegos de Castillo
SAN JOSE STATE UNIVERSITY

This issue of the newsletter begins with the winner of the 2019 APA Prize in Latin American Thought. In his award-winning essay, Noell Birondo seeks to reconsider the relationship between Aristotelian and Aztec ethics. Particularly, Birondo seeks to redeem Aristotle from the post-Conquest appropriations of Spanish interpreters. He argues that “a consideration of the actual historical collision of these two radically distinct belief systems, Christian and Aztec, reveals the possibility—even in the early modern period—of a helpfully ‘dialogical’ Aristotelianism, one that strains to understand, from within, the perspective of alien others.”

Our second essay is an ambitious and thought-provoking piece by Andrew C. Soto. Its aim is both to set the tone for and propose the epistemological contours of a “Chicano/a philosophy.” In “Chicano/a Philosophy: Rupturing Gringo Anti-Chicano/a Paradigms and Philosophies,” Soto argues that a Chicano/a philosophy will not get off the ground if its grounding epistemology is reflective of or imitates the grounding paradigms of the West, what he calls “gringo paradigms.” The reason why this starting point will not do is that those epistemologies “are upheld by a violent gringo legal system that a priori and necessarily constructs the Chicano/a as a permanent threat to its civilized and rational institutions.” What is needed is a “complete rejection of integrationist and assimilationist logics, anti-Chicano/a gringo ethics, and philosophies centered on the need for gringo recognition.” A Chicano/a philosophy, furthermore, “must seek to build on a Chicano/a logics that centers Chicano/as as creators of reason, knowledge, history, and philosophy.” Soto’s grounding intuition is that non-white, non-Anglo philosophers in the US seeking to articulate their own philosophical voice sabotage themselves by demanding recognition from a system that continues to marginalize them and oppress them. Soto’s article should evoke much discussion.

We continue with two book reviews. The first, by Manuel Chavez, considers Roberto D. Hernández’s Coloniality of the U-S///Mexico Border: Power, Violence, and the Decolonial Imperative. Chavez reads Hernandez’s text as a philosophical exploration of immigration and violence that “offers . . . a philosophically provocative way to rethink the hegemonic discourses about the politics of the border.” According to Chavez, Hernández’s text allows us to rethink contemporary instances of anti-immigrant violence (such as the 1984 McDonald’s shooting as well as the 2019 Walmart shooting) as extensions of colonial oppression. Ultimately, writes Chavez, Hernández “provides a political and moral ‘philosophy born of colonial struggle’ that can help us imagine and create resistant possibilities together.”

The second book review, by Roy Ben-Shai, is of Carlos Pereda’s Lessons in Exile. Ben-Shai’s review is less an academic review and more of a persuasive invitation to read Pereda’s text as offering a phenomenology of exile or, as Ben-Shai says, a “paradigm for experience in general.” Pereda’s Lessons in Exile is about exile, specifically, about Pereda’s own exile in Mexico (he’s originally from Uruguay), something that allows this small philosophical text to open a phenomenological window into the phenomenon, allowing, moreover, for one to have an experience of exile along with Pereda. As Ben-Shai puts it, “I did not anticipate, in fact, just how far Pereda’s book would end up [teaching me the] kind of lessons exile has to teach.”

CALL FOR SUBMISSIONS
The APA Newsletter on Hispanic/Latino Issues in Philosophy is accepting contributions for the FALL 2020 issue. Our readers are encouraged to submit original work on any topic related to Hispanic/Latino thought, broadly construed. We publish original, scholarly treatments, as well as reflections, book reviews, and interviews. Please prepare articles for anonymous review.

All submissions should be accompanied by a short biographical summary of the author. Electronic submissions are preferred. All submissions should be limited to 5,000 words (twenty double-spaced pages) and must follow the APA guidelines for gender-neutral language and The Chicago Manual of Style formatting. All articles submitted to the newsletter undergo anonymous review by members of the Committee on Hispanics.

BOOK REVIEWS
Book reviews in any area of Hispanic/Latino philosophy, broadly construed, are welcome. Submissions should be accompanied by a short biographical summary of the author. Book reviews may be short (500 words) or long (1,500 words). Electronic submissions are preferred.
DEADLINES
Deadline for spring issue is November 15. Authors should expect a decision by January 15. Deadline for the fall issue is April 15. Authors should expect a decision by June 15.

Please send all articles, book reviews, queries, comments, or suggestions electronically to the editor, Carlos Alberto Sánchez, at carlos.sanchez@sjsu.edu, or by post: Department of Philosophy, San Jose State University, One Washington Sq., San Jose, CA 95192-0096.

FORMATTING GUIDELINES

ARTICLES
The Virtues of Mestizaje: Lessons from Las Casas on Aztec Human Sacrifice
Noell Birondo
WICHITA STATE UNIVERSITY
Winner, 2019 APA Essay Prize in Latin American Thought

1. INTRODUCTION
Western imperialism has received many different types of moral-political justifications, but one of the most historically influential justifications appeals to an allegedly universal form of human nature. In the early modern period this traditional conception of human nature—based on a Western archetype, e.g., Spanish, Dutch, British, French, German—opens up a logical space for considering the inhabitants of previously unknown lands as having a “less-than-human” nature. This appeal to human nature originally found its inspiration in the philosophy of Aristotle, whose ethical theory pervaded the work of European philosophers at the outset of the early modern period and the modern age of empire. Indeed some Spanish writers—most famously, Juan Ginés de Sepúlveda (b. 1494)—explicitly appealed to Aristotle’s moral-political philosophy in order to justify the conquest of the Americas in the early sixteenth century, for instance, to justify war against the Aztecs and other indigenous peoples. At the time of European arrival, the Aztec civilization was easily the greatest in Mesoamerica—and yet the Europeans generally considered the Aztec people to be “barbaric,” i.e., less-than-fully-human.

Despite Aristotle’s association with the history of Western imperialism, the past forty years in moral philosophy have seen an explosion of interest in Aristotle’s ethics, especially the idea that the virtues are indispensable to a good human life. Today, proponents of an Aristotelian ethics can insist that Aristotle’s appeal to human nature can easily allow for—and even celebrate—the wide variety of lifestyles found in different cultural-historical contexts, that it can allow for a more flexible conception of the ways in which human nature is realized in different cultures and historical moments. Several philosophers have even developed accounts of previously overlooked virtues that people will need under conditions of oppression or social marginalization, conditions that are often the result of intercultural imperialism. These recent developments flow naturally from an Aristotelian orientation, and such developments should lead us to consider, further, whether the assumptions that enabled Western imperialism might linger enough today to influence contemporary conceptions of the virtues—for instance, unreflective assumptions about European cultural supremacy and American exceptionalism.

My main hypothesis here is that such unreflective and deep-seated cultural prejudices have shaped the Western development of Aristotelian ethics in various ways—as already illustrated in Sepúlveda’s appeal to Aristotle’s “natural slaves”—and that such prejudices partially explain the felt need for an extra-ethical foundation for the virtues, one provided by a universal and morally determinative form of human nature. An acknowledgement of the actual world-historical development of Aristotelian ethics would therefore be a first, but crucial step towards developing a more modest, intercultural version of a contemporary neo-Aristotelian ethics—an approach that aims precisely, in its openness and epistemological humility, to supersede any form of imperialism. Such cultural prejudices can obscure a more plausible and open-ended version—an intercultural and self-consciously “mestizo” version—of a plausible neo-Aristotelian ethics.

What I argue in this paper, much more specifically, is that a consideration of the actual historical collision of these two radically distinct belief systems, Christian and Aztec, reveals the possibility—even in the early modern period—of a helpfully “dialogical” Aristotelianism, one that strives to understand, from within, the perspective of alien others. This dialogical Aristotelianism disavows an “epistemology of ignorance”—it disavows the need not to know, the motivation not to learn, something that is arguably essential to Eurocentrism. A dialogical Aristotelianism strongly suggests that a philosophical version of “mestizaje” can enrich the best philosophical accounts of the virtues we have, both now and in future research on moral character (I will return to what, in my view, this type of philosophical “admixture” will fruitfully include in §5 below).

2. THE AZTECS AS ALIEN OTHERS
In order to illustrate this dialogical version of an Aristotelian ethics, I will discuss two of the central arguments deployed by Bartolomé de Las Casas (b. 1484) in defense of Aztec human sacrifice. This defense was originally delivered in front of the Council of the Indies, a tribunal convened in 1550 by Charles I of Spain—Charles V of the Holy Roman Empire—in order to determine the fate of the native inhabitants of the Indies (our Americas). The question
before the Council was whether waging war against the native inhabitants of the Indies was morally justified in order to convert them to the Christian faith. This question seemed urgent given the apparently barbaric nature of the Aztecs and other indigenous peoples—something most notably demonstrated by the religiously sanctioned practice of human sacrifice and the equally morbid practice (or so it was believed) of consuming the flesh of the sacrificial victims. Despite these apparently barbaric practices—which genuinely horrified sixteenth-century Europeans—Las Casas defends the rationality of the Aztec way of life.

The discussion here should not, of course, be thought to question the gruesome nature of Aztec human sacrifice. One recent historian, drawing on authoritative sources, offers this lurid description:

In a typical ritual . . . the helpless individual was confronted with the sight of the great sacrificial stone, stained with blood, which also matted the hair of the magnificently adorned priests. Seized by these gory apparitions, the victim was stretched backwards over the stone altar, each limb extended by a priest so that the back was arched and the chest stretched taut and raised high toward the heavens. A fifth priest struck open the chest with an obsidian knife, excised the heart with knife and hands and raised the fertile offering to the heavens, displaying to the gods the sacrificial fruit.  

Las Casas addresses the question of whether it would be just to wage war against the Aztecs, in the name of Christianity, in order to end this practice and to spare the lives of the innocent victims. The answer he gives is “No.”

Las Casas’s defiant approach to these issues already shows in his response to a different Spanish pretext for war. According to this different justification, war against the indigenous peoples is justified because they are guilty of killing Christians and therefore guilty of thwarting the spread of Christianity. Las Casas provides a sharp response. It highlights the contemporary relevance of thinking through his arguments—for instance, their relevance in evaluating past and present US policy toward indigenous peoples and their descendants. In response to this initial pretext for war—that war is justified because the Indians kill Christians and prevent the spread of the Gospel—Las Casas responds that although the Indians have indeed killed Christians, they have not killed them qua Christians. Rather, the Indians kill Christians qua perpetrators of violence, theft, rape, torture, and murder. This insightful distinction is a distinction of which any Aristotelian can be justly proud. Its contemporary relevance should be obvious.

3. ON ARISTOTELIAN ENDOXA

Overall, Las Casas argues that the Aztec way of life “cannot be excused in the sight of God” (that the Mexica are not objectively correct about the propriety of human sacrifice) but that it “can completely be excused in the sight of men.” This means that no one can justifiably blame the Aztecs for their violent religious practices—but certainly not the Spaniards. Thus the following line of inquiry, with which Las Casas opens his discussion of human sacrifice, is certainly intended to sting. Las Casas says that, “It would not be right to make war on them for this reason.” This is because it is difficult to absorb in a short time the truth proclaimed to them. . . . Why will they believe such a proud, greedy, cruel, and rapacious nation? Why will they give up the religion of their ancestors, unanimously approved for so many centuries and supported by the authority of their teachers. . . .?

In this passage Las Casas gestures toward the first of his two main arguments here, which is that the Aztecs are committing what he calls a “probable” error. In explaining the nature of probable error, as he sees it, Las Casas makes direct reference to Aristotle’s Topics Book I. Las Casas insists that “[as the Philosopher says, that is said to be probable which is approved by all men, either the majority of wise men or by those whose wisdom has the greatest following.” What is this reference to Aristotle?

At the outset of the Topics, one of his logical treatises, Aristotle distinguishes between two types of deductions or “syllogisms.” He calls the first type a “demonstration,” and he calls the second type a “dialectical” deduction. Aristotle explains the distinction in the very passage Las Casas cites:

Now a deduction (sullogismos) is an argument in which, certain things being laid down, something other than these necessarily comes about through them. It is a demonstration, when the premises from which the deduction starts are true and primitive, or are such that our knowledge of them has originally come through premises which are primitive and true; and it is a dialectical deduction, if it reasons from reputable opinions (ex endoxô̱n).

Aristotle goes on to explain what he means by saying that in the case of dialectical deductions the premises are reputable opinions—the Greek word here is endoxa (sometimes also translated as “probable assumptions”). Regarding such endoxa Aristotle says that, “those opinions are reputable which are [i] accepted by everyone or [ii] by the majority or [iii] by the wise—i.e., by all, or by the majority, or by the most notable and reputable of them.” This means that dialectical deductions will differ from demonstrations. For demonstrations begin from premises that are (or are derived from premises that are) “true and primitive.” That is, demonstrations begin from premises which, like each of the first principles in an Aristotelian science, “should command belief in and by itself.”

By contrast, a “dialectical” deduction will proceed from endoxa—it will proceed from those reputable opinions or modest human starting points which, as Aristotle says in the Nicomachean Ethics, are the only appropriate starting points in ethics. In a practical subject like ethics, the appropriate starting points are the ethical opinions that are accepted by everyone, or by the majority, or by the wise. We must start from things that are evident to us. Our starting points in ethics will certainly never have the epistemological firmness of premises that are “true and primitive.” Rather,
our starting points in ethics can only amount to the best ethical judgments that we and our society have managed to arrive at so far—the ethical judgments that seem most evident to us. This will be a subject that, as a historical matter, merits our ongoing ethical reflection.20

Las Casas’s explicit recognition of Aristotle’s ethical methodology—along with the way he utilizes this methodology in defense of the Aztec way of life, for instance by castigating the ethical outlook of his fellow Spaniards—indicates that Las Casas interprets Aristotle’s ethics in terms of what I have characterized elsewhere as an “internal” validation of the virtues of character.21 An internal validation of the virtues of character disavows any “external” appeal to a universal and morally determinative form of human nature from which one could derive a specific conception of the virtues of character. The general form of this different type of validation—an external validation of the virtues—manifests itself most obviously in interpretations of Aristotle that appeal to an alleged “metaphysical biology” or other form of natural teleology, usually culminating in a naturalistic conception of well-being or flourishing.

By contrast, Las Casas seems to recognize that Aristotle’s ethical project can be understood as significantly more modest than that. Indeed, Las Casas seems to follow Aristotle down this different philosophical path. This more modest Aristotelianism would certainly explain the sharp contrast between the charitable hermeneutical understanding deployed by Las Casas (even with respect to human sacrifice) and the quite different apology for conquest deployed by Sepúlveda—i.e., an apology for conquest that exhibits a form of willful hermeneutical ignorance, especially, but not only, in its appeal to natural slavery.22

4. THE LIMITATIONS OF THE NATURAL LIGHT

Las Casas also argues that it is not easy to convince even rational people to abandon their cultural heritage in a short amount of time, especially given only the resources provided by “natural light of reason”—that is, without the further epistemological resources that Las Casas believes are provided by “faith, grace, and doctrine.” Waging war on the Aztecs would therefore be unjustified, because “it is difficult to absorb in a short time the truth proclaimed to them.” Here Las Casas emphasizes that the “natural light of reason” displays epistemological limitations—that in the absence of divine revelation, natural reason seems to provide justificatory reasons in favor of human sacrifice.23

In what follows I want to mention three possible strategies for supporting this second line of defense. Las Casas employs the first two strategies in the Defense. His avoiding the third one must have been determined by facts on the ground.

First, Las Casas appeals to biblical and historical precedents of human sacrifice that seem to illustrate its consistency with natural reason. He cites biblical episodes apparently indicating that God sometimes requires (or permits) human sacrifice. He also cites episodes of human sacrifice within Western civilizations: for instance among the Greeks, Romans, and even “our own Spaniards.”24

Second, Las Casas argues that natural reason seems even to require sacrificing humans to God. He proceeds by first establishing four principles (mostly by appeal to theological and philosophical authorities): (1) No nation is so barbarous that it does not have at least some confused knowledge of God; (2) People are led by natural inclination to worship God according to their capacities and in their own ways; (3) There is no better way to worship God than by sacrifice, which is the principle act of latria [adoration]; (4) Offering sacrifice to the true God, or to the one who is thought to be God, comes from the natural law, whereas the things to be offered to God are a matter of human law and positive legislation.25 From these principles Las Casas derives the conclusion of the natural light of reason (given that no earthly thing is more valuable than human life). He writes:

Therefore nature itself dictates and teaches those who do not have faith, grace, or doctrine, who live within the limitations of the light of nature, that, in spite of every contrary positive law, they ought to sacrifice victims to the true God or to the false god who is thought to be true, so that by offering a supremely precious thing they might be more grateful for the many favors they have received.26

A similar conclusion might also be reached by direct appeal to Christianity, as follows.

Third, Las Casas might have emphasized—something that he does indeed mention—that Christianity itself essentially involves human sacrifice.27 Hence the activity of human sacrifice cannot, by itself, be any sign of barbarism and cannot be contrary to the natural light of reason. The charitable view would be that the Aztecs are only partially mistaken here (in absence of divine revelation), since the sacrificial debt has already been paid in the person of Jesus Christ. Moreover, as I myself would emphasize, if one takes seriously the Roman Catholic doctrine of transubstantiation—that the bread and the wine of the Eucharist are not mere representations of the body and the blood of Christ, but that they literally are the body and the blood of Christ—then Christianity also involves a form of cannibalism.

What each of these strategies demonstrates is the possibility of a radical form of hermeneutical charity even regarding the allegedly barbarous practices of the Aztec people. Gustavo Gutiérrez nicely summarizes this in his magisterial study of Las Casas. Gutiérrez writes:

By attending to the customs, lifestyles, and religious freedom of the Indians, [Las Casas] created the necessary conditions for a dialogue to be conducted in respect for both parties. In this manner of dialogue, reason, not undue pressure, makes possible an integral presentation of the gospel message: now that message is offered—without prejudice to the values of the one proclaiming it—for the free acceptance of each hearer.
Such a dialogue will respect the rational freedom of both parties. It will also involve, not only the giving of reasons, but also the taking of them:

If evangelization is a dialogue, it will not exist without an effort to understand the position of one's interlocutor from within, in such a way that one may sense the vital thrust of these positions and grasp their internal logic. Neither will it be possible unless one is ready to give as well as to receive.\(^5\)

This passage characterizes the dialogical approach to ethics that I am urging (but without any appeal to the supernatural). Although we should certainly be wary, in intercultural contexts, of any appeal to "evangelization," Las Casas's radical hermeneutical charity advances the discussion here.\(^9\) Las Casas demonstrates the central virtue involved in a philosophical version of mestizaje: a radical hermeneutical charity that constitutes a distinctive form of epistemic justice. This epistemic virtue disavows an epistemology of ignorance by recognizing and—where appropriate—encouraging philosophical admixture. This philosophical admixture will occur, in my view, in at least the following two ways. First, it will occur across spatio-cultural geography and between different philosophical, cultural, and academic communities. This is a kind of cross-pollination—something that seems to be more often lauded than practiced. Second, it will occur across world-historical time, as a result of one's own historical (i.e., "genetic") philosophical inheritance, an inheritance that shapes one's overall philosophical outlook, one's framework of thought. This is a kind of dialogue with the past.\(^30\) In the final section I gesture toward a more rounded view of each of these.

5. EPISTEMIC JUSTICE IN ARISTOTELIAN ETHICS

In the *Nicomachean Ethics*, Aristotle restricts the audience of his ethical lectures on the grounds that those who engage in moral philosophy must have been well brought up or brought up in good habits.\(^31\) What is less frequently noticed is that this requirement—to have appropriate ethical starting points and to have a character sufficiently well formed that one is not swayed by, for instance, unruly desires\(^32\)—is also one that applies to Aristotle himself, and to Aristotelian moral philosophers in general, since they are also engaged in the practice of moral philosophy. But in the aftermath of the Spanish conquest of Tenochtitlán and other parts of the Americas, Aristotelian moral philosophy did not generally embrace the dialogical approach advocated by Las Casas. I believe we need to trace the history of the damage done to moral philosophy in the long historical interim.\(^33\)

By way of analogy, consider an episode of European barbarism, recorded from the perspective of Aztec witnesses immediately before the fall of Mexico. In this episode, Spanish soldiers block the exits during a festive religious gathering, allowing the soldiers to massacre the participants. This gruesome episode seems to be taking inspiration from something in Homer and giving inspiration to something in George R. R. Martin—except that this actual historical episode involves gross violations of human dignity:

And when they had closed them off... they then entered the temple courtyard to slay them... they surrounded those who danced whereupon they went among the drums. Then they struck the arms of the one who beat the drums; they severed both his hands, and afterwards struck his neck, [so that] his neck [and head] flew off, falling far away... Of some, they struck the belly, and their entrails streamed forth. And when one in vain would run, he would only drag his entrails like something raw, as he tried to flee. . . .

And the blood of the chieftains ran like water; it spread out slippery, and a foul odor rose from the blood. And the entrails lay as if dragged out. And the Spaniards walked everywhere, searching the tribal temples; they went making thrusts everywhere in case someone were hidden there. Everywhere they went, ransacking every tribal temple they hunted.\(^34\)

Ultimately it is unclear whether Anglo-American moral philosophy has displayed an understanding of cultural others that has been much better than the understanding displayed in this historical episode. Whether intentionally or not, mainstream Anglo-American philosophy has been remarkably effective at securing its borders against what many of its practitioners consider to be alien influences. This includes influences from other cultures and from demographics other than the dominantly situated demographic in the profession; from other academic disciplines (for instance history, sociology, and anthropology, although this is improving in some quarters); and from philosophical methodologies other than the methodologies developed within Anglo-American philosophy in the early- and mid-twentieth century and still insisted upon by some philosophers today as the defining mark of any genuine philosophy. Indeed, some philosophers seem to be eerily at home with the history of Western imperialism. This is a history that such philosophers seem to think can be neatly left in the past, in such a way that they—and their favored research projects—can continue to benefit from centuries of past injustice.

Obviously, I cannot fully develop these suggestions here.\(^35\) Instead of doing so, I will emphasize something that I think is utterly crucial for developing a plausible neo-Aristotelian ethics informed by an intercultural perspective. This is a radical form of cultural self-scrutiny, especially a scrutiny of the ethical and epistemic prejudices that are embedded within our social-historical framework of thought—a framework of thought that is, of course, usually taken for granted. It ceases to be taken for granted—or can do so—when it comes into contact with radically alternative frameworks, ones that are culturally or historically distant from our own current location. To put the point differently: contemporary moral philosophers need to pay greater attention to history in at least two senses. We need a better understanding of the history and the historicity of philosophy, an understanding of the former that is not willfully inaccurate and that disavows the arrogance of knowing only one's own philosophical tradition.\(^5\) We also need a better appreciation of our current place in history.
and our cultural particularity—a critical understanding of the framework of thought that can, of course, seem inevitable to us. This would be a form of neo-Aristotelian ethics that takes seriously those genealogical approaches that still remain very much against-the-current in contemporary moral philosophy. It would also be a form of Aristotelian ethics that, in better appreciating our current (globalized, multicultural, postcolonial/neo-colonial) place in history, strains to embody the virtues of epistemic justice.

ACKNOWLEDGEMENTS

Research on this paper was supported by an URCA grant (2019) from the Office of Research at Wichita State University and is gratefully acknowledged here. Earlier versions of the paper were presented at the Pacific Division meeting of the American Philosophical Association (2019), at the University of Texas at El Paso for the Moral and Political Philosophy at the Border Conference (2019), and at WSU. Thanks are due to the participants and organizers on each of these occasions, especially Alejandro Santana, Caroline Arruda, Lauren Viramontes, and my colleagues at WSU. Special thanks are due to Carlos Pereda, Guillermo Hurtado, Fanny del Río, Helen Daly, Sebastian Purcell, Andrea Sullivant-Jake, and Dante Johnson for their valuable encouragement during the paper’s development; to Robert Eli Sanchez for sharing a chapter of his now published collection on Latin American and Latinx philosophy; to the students in my class on the Invention of Latin American and Philosophy: Eurocentrism as an Epistemology of Ignorance. On the earliest debates concerning the general capacity of the indigenous peoples of the Americas, see Hanke, All Mankind Is One: A Study of the Disputation Between Bartolomé de Las Casas and Juan Ginés de Sepúlveda in 1550 on the Intellectual and Religious Capacity of the American Indians, chap. 1.


3. Alcoff (“Philosophy and Philosophical Practice,” 402) argues that the Eurocentrism involved here essentially involves an epistemology of ignorance: “Such a construction of barbarian identity removes any motivation to learn other ways or creeds. The claim that those designated are inferior and inadequate thinkers is not justified by a study and evaluation of different practices, customs, forms of religiosity, institutions, beliefs, and the like, but simply by the observation that a group is not-Christian or not-rational or not-self.” She argues that Las Casas recognizes his own perspective as a perspective, and hence that he can “see the Other as having a substantive difference, and not simply as a ‘not-self’” (405). Cf. Gutiérrez, Las Casas: In Search of the Poor of Jesus Christ, passim, and 188–189, quoted below; Beuchot (*The History of Philosophy in Colonial Mexico*, 26–36). See also Pagden, The Fall of Natural Man, chap. 4, on the move to considering the native peoples of the Americas to be “nature’s children.”


5. Philosophers who are explicitly indebted to Aristotle here include Tessman, Burdened Virtues, and Fricker, Epistemic Injustice. Recent discussions of epistemic injustice are all ultimately indebted to the revival of Aristotelian ethics in the latter part of the twentieth century. In future work I aim to urge the importance of this genealogical fact for the future development of any plausible neo-Aristotelian ethics: Such an ethics must embody epistemic justice.


8. Julio Covarrubias’s recent case for “letting go” of mestizaje rightfully emphasizes concerns about epistemic settler erasure and the logic of elimination that threatens indigenous communities (Covarrubias, “Letting Go of Mestizaje: Settler Colonialism and Latin American/Latinx Philosophy,” §3). But the virtues of mestizaje are directed at the dominantly situated paradigm in philosophy, a Eurocentric paradigm which, as Alcoff says, apparently cannot “play well with others” (“Philosophy and Philosophical Practice,” 401); see also Pappas, “The Latino Character of American Pragmatism,” in the observations in William James and John Dewey of what North American philosophy and culture can learn from Latin America. Hence it is not true in this context that “to speak of mestizaje is to speak: . . . of a kind of cultural genocide that reproduces settler erasures” (Covarrubias, “Letting Go of Mestizaje,” 6). What is good for the dominantly situated gander is not necessarily good for the marginedly situated goose: cf. Nicomachean Ethics (NE) II.6, 1105a35–1106b7.

9. To speak of the Aztecs here is perfectly appropriate, in spite of the fact that the conquest of Tenochtitlán (1521), the Aztec capital, antedates the debate between Las Casas and Sepúlveda (1550). In the minds of sixteenth-century Europeans, nothing compared to what the Spaniards witnessed at Tenochtitlán. Here I am following the lead of Anthony Pagden, who notes that “The most famous of the Amerindian cannibals were, of course, the Aztecs,” who “sacrificial bounties of human sacrifice were assumed to have been followed by orgiastic feasts on the flesh of the victims” (*The Fall of Natural Man*, 83). The example of Aztec human sacrifice has been paradigmatic for late twentieth-century and philosophy interested in intercultural understanding. See, e.g., Taylor, “Understanding the Other: A Gadamerian View on Conceptual Schemes”; Williams, *Morality: An Introduction to Ethics*, 24–26.

10. See, e.g., Pagden, *The Fall of Natural Man*, 80–90.

11. Dodds Pennock, *Bonds of Blood: Gender, Lifecycle and Sacrifice in Aztec Culture,* 21; cf. Florentine Codex 2.2.

12. On the contemporary relevance of the Valladolid debate, see the excellent recent treatment in Santana, “*The Indian Problem*: Conquest and the Valladolid Debate”; see also the magisterial discussion of Las Casas in Gutiérrez, Las Casas: *In Search of the Poor of Jesus Christ*. The best short book on the debate in English—which encompasses both its prelude and its aftermath—remains Hanke, *Aristotle and the American Indians.* See also the more detailed discussion in Hanke, All Mankind Is One.


14. For details, see Las Casas, *The Devastation of the Indies: A Brief Account*. It has been long recognized that for rhetorical and political purposes (this was quite common) Las Casas engages in certain exaggerations of the devastation he documents, especially regarding magnitude (e.g., number of deaths). For contemporary discussions that significantly temper the “Black Legend” of unparalleled Spanish brutality, see Greer, Mignolo, and Quilliam, *Rereading the Black Legend: The Discourses of Religious and Racial Difference in the Renaissance Empires.*


16. Ibid., 220–21.


18. Ibid., 100b20–23; cf. NE VI.11, 1145b2–7.

19. NE I.4, 1095a30–b4. In Aristotelian science we presumably start from what is “better known to us,” proceed to what is “better known by nature,” and construct “demonstrations” of the completed science.

20. On Aristotle’s method in ethics, Kraut (*How to Justify Ethical Propositions: Aristotle’s Method*) provides a helpful overview, noting on Aristotle’s behalf that when we engage in ethical inquiry, “it is reasonable to throw into the mixture of opinions
that we take seriously not only the theories of those who have spent their lives studying the subject, but also the common moral consciousness, not only of our time and place, but of other times and places as well" (80). See further references is n. 21, below, and the discussions of neo-Aristotelian ethics cited there.

21. An internal validation contrasts with an external validation, the latter of which I have characterized elsewhere as follows: "An external validation of the virtues of character is an attempt to demonstrate that possession of the virtues of character is necessary in order to secure some good, or to avoid some harm, which is the good in question, or the harm, is recognizable as such independently of the particular evaluative outlook provided by possession of the virtues themselves. The validation will thus rely on resources that are 'external' to the particular evaluative outlook to be validated" (Birondo, "Aristotle and the Virtues of Will Power," 85; "Virtue and Prejudice: Giving and Taking Reasons," 191). An internal validation of the virtues need not be philosophically trivial, as I attempt to illustrate in "Patriotism and Character: Some Aristotelian Observations.

22. Cf. Pohlhaus, "Relational Knowing and Epistemic Injustice"; Fricker, "Epistemic Injustice and the Preservation of Ignorance."


25. La Casas, In Defense of the Indians, chap. 35.

26. Ibid., 234.

27. Ibid., 239, cf. Pagden, The Fall of Natural Man, 227, n. 198.

28. Gutiérrez, Las Casas: In Search of the Poor of Jesus Christ, 188–189, my emphasis). Gutiérrez finds a similar hermeneutical charity in the work, much later, of Sor Juana Inés de la Cruz (b. 1648); see Gutiérrez, Las Casas, 252, n. 69. On the importance of being as well as giving reasons in intercultural contexts, see Birondo, "Virtue and Prejudice," to which the current paper is a kind of late addendum.

29. This seems to be the context in which to understand Alcoff’s claim that in contrast to a Cartesian form of self-understanding, Las Casas is “groping toward a different self-understanding, in which one’s own inclinations are analyzed in relation to their social context” (“Philosophy and Philosophical Practice,” 405). She immediately adds something that could be helpful for contemporary philosophers: “Within this approach, dialogic models of philosophical thought, especially those that can span cultures and belief systems, are non-negotiable necessities for the development of understanding.” Castro (Another Face of Empire: Bartolomé de Las Casas, Indigenous Rights, and Ecclesiastical Imperialism) and von Vancan (The Color of Citizenship: Race, Modernity, and Latin American/Hispanic Political Thought) reach rather harsher verdicts on Las Casas’s evangelism. But neither author seems to me adequately to address Gutiérrez’s painstaking case for the claim that a “single idea” governs Las Casas’s Apología: “respect for the Indians’ religious customs” (Gutiérrez, Las Casas: In Search of the Poor of Jesus Christ, 174).

30. Beuchot ("The Study of Philosophy’s History in Mexico as a Foundation for Doing Mexican Philosophy") helpfully argues for the type of anti-presentism that I mention here. He argues that contemporary Mexican philosophers can benefit from a neo-Aristotelian outlook that appreciates the influence of cultural-historical tradition—he cites the work of Alasdair MacIntyre and Hans-Georg Gadamer (127). Beuchot reasonably asks: “If it is true that we live within a tradition, how can we advance in it or even oppose it if we do not have at least a minimum knowledge of it?” (114). The right hermeneutical balance can nevertheless be, in any specific context, difficult to strike; see O’Gorman, “Art or Monstrosity,” on understanding Aztec archeological artifacts, specifically the magnificent statue of Coatlicue in the National Museum of Anthropology in Mexico City.

31. NE I, 4, 1095b4–6.

32. NE I, 3, 1095a4–6.

33. The valuable collection of essays in Miller, The Reception of Aristotle’s Ethics, provides a good beginning here—it considers the historical reception of Aristotle’s ethics—except that there is no consideration of the European encounter with the Americas or the Latin American world. A valuable corrective can now be found in Aspe, Aristóteles y Nueva España.

34. Florentine Codex 12.20, 53–54.

35. The historiographical study in Park, Africa, Asia, and the History of Philosophy: Racism in the Formation of the Philosophical Canon, and the work of Robert Bernasconi, Walter Mignolo, and Charles Mills (among others) have helpfully gotten the discussion going; as have recent attempts to generate “new histories of philosophy”—but see also Allais, “Kant’s Racism,” and Ameriks, “Kant and Dignity: Missed Connections with the United States,” for helpfully more sympathetic views of the late eighteenth century and Kant in particular. Recent work on Aztec ethics and Aristotle (Purcell, “Eudaimonia and Neill-Ziiltzi: Aristotle and the Aztecs on the Good Life”) and Aztec metaphysics (Maffie, Aztec Philosophy: Understanding a World in Motion) illustrates another type of void waiting to be filled.

36. Latin American philosophy provides an epistemological opportunity for Anglo-American philosophy: to scrutinize its own historical development from the radically alternative perspective of world-historical marginality. This theme in Latin American thought—the theme of marginality—has been especially emphasized in the work of Leopoldo Zea, Enrique Dussel, and Walter Mignolo. See, for instance, Zea, The Role of the Americas in History; Dussel, The Invention of the Americas: Eclipse of ‘the Other’ and the Myth of Modernity; Mignolo, Local histories/Global Desires: Coloniality, Subaltern Knowledges, and Border Thinking; see also Alcoff, “Philosophy and Philosophical Practice”; Schutte, Cultural Identity and Social Liberation in Latin American Thought. On the dangers that can prevent dominantly situated groups from taking advantage of such epistemological opportunities, see Milis, “White Ignorance”; Pohlhaus, “Relational Knowing and Epistemic Injustice”; and Fricker, “Epistemic Injustice and the Preservation of ignorance.”

BIBLIOGRAPHY


**Chicano/a Philosophy: Rupturing Gringo Anti-Chicano/a Paradigms and Philosophies**

Andrew C. Soto

HARTNELL COLLEGE, CALIFORNIA

---

1. **THE EMERGENCE OF CHICANO/A PHILOSOPHY**

Chicano/a philosophy axiomatically rejects the use of gringo paradigms, logic, reason, and philosophies. It is grounded in the axioms, paradigms, logics, circumstances, and history of the Chicano/a people. It is a field that has emerged from the interdisciplinary work of scholars across numerous fields, including activists, students, artists, musicians, educators, laborers, farmworkers, and several others dedicated to liberating Chicano/as from racist gringo institutions and gringo anti-Chicano/a systems of reason, morality, philosophy, and logics. This gringo system has terrorized Chicano/as by unwaveringly attempting to imprison them to US institutions grounded in a racist/imperial logics that has legally, logically, axiomatically, economically, socially, philosophically, and politically made their exploited existence, their every breath, gesture, and thought, a permanent threat to Western society.

Octavio I. Romano-V in his trailblazing article “Social Science, Objectivity, and the Chicanos,” is clear that in contrast to Western paradigms rooted in a Western logic that shapes and maintains the institutions, norms, ideals, standards, ethics, categories, and philosophies of Western man and woman, a Chicano/a paradigm is “the symbiotic relationship within the universe, that is the historical patrimony of Chicanos, revolves around a philosophical system about the nature of man and man, of man in nature, and man in the universe. In essence, this philosophy is non-Weberian, non-Hegelian, and it is very dissimilar to Greek ontology.”

Put another way, in opposition to a racist Western system of logic and reason that has historically constructed the Chicano/a as a threat, foreigner, and...
nonhuman in the imagination and psychotic institutions of the gringo, Chicano/a philosophy provides Chicano/as with a methodology, framework, and paradigm to analyze and theorize their intersubjectivity and historical relation to colonialism, empire, white supremacy, and Western anti-Chicano/a logic/reason.

Provided this as a point of departure, Chicano/as become creators and guardians of their own philosophical systems, logics, epistemologies, and institutions. Simultaneously, as Chicano/as are institutionally constructed as nonhumans within anti-Chicano/a Western paradigms, they implode, destruct, and contour the gringo world by recognizing, analyzing, and understanding it as a permanent psychological and physical threat to their existence and by rejecting the axioms and imperial/genocidal logics that undergird anti-Chicano/a Western paradigms and philosophies. Extending Octavio Romano’s thesis in “Social Science, Objectivity and the Chicanos,” Nick C. Vaca in “The Mexican-American in the Social Sciences,” Miguel Montiel in “The Social Science Myth of the Mexican American Family,” Alfredo Mirandé in The Chicanos: An Alternative Perspective, and Deluvinia Hernandez in The American Challenge to a Sacred Crow, shed light on the violence, institutional tyranny, and threat Western paradigms pose to Chicano/as. These scholars persuasively articulate how concepts in Western philosophy, such as objectivity and dualism, are used by gringo and gringo-minded academics of color to institutionalize, conceptually, and logically validate the permanent creation of the nonhuman Chicano/a subject.

As Alfredo Mirandé advances in The Chicanos: An Alternative Perspective, many early Chicano/a scholars rejected the use of Western paradigms as tools to theorize and analyze Chicoano/a culture. Instead, they aimed to create “new paradigms or theoretical frameworks that were] consistent with a Chicoano world view and responsive to the nuances of Chicoano culture.” They understood Western logic and reason, not as systems that transcend our environment or material world, but systems dependent on the very Western institutions they inform, shape, and create. Mirandé’s point is critical because it means that Western logic and Western institutions can only survive and maintain their dominance in conjunction with each other. In Octavio Romano’s essay “Social Science, Objectivity, and the Chicanos,” he explains the impossible task one must undertake to meet the standards of objectivity within Western logic and how this standard is the centerpiece to pathological and racist academic theories:

As a part of this historical process, there followed a renewed exploration into the concept of objectivity. As generally defined in Western thought, the concept of objectivity is impossible without a corresponding belief in man’s ability to separate his mind not only from his body, but also from all of his ecological surroundings, whether or not these ecological surroundings are human or physical. It is in this manner that the mind, when believed to be in its objective state, has come to be viewed as separable in Western science just as the soul has been seen as separable in traditional Western theology.

Romano adds, “Western man, then, in his quest for a pure objective reality (that is, to be objective) began to consider events, phenomena, and ideas as apart from personal self-consciousness, to be dealt with ideally in a detached, impersonal, and unprejudiced manner.” He understood that logic and the concepts of objectivity and dualism were central components to the gringo’s arsenal of violence, lies, and mythmaking. These concepts provide researchers and scientists with the false belief that if they apply a critical eye to their analysis it can be free of biasness and prejudice. In other words, in a system where gringos have control and power over reason, logic, research, and academia, objectivity and dualism are nothing more than concepts gringos use to manipulate the world to their standards, ethics, laws, and norms. Hidden behind these concepts, the gringo is able to create and advance pathological Chicano/a theories that are said to be neutral, fair, rigorous, and scientific. Ultimately, critiquing both Pythagoras’s and Plato’s dualism, “the belief in the dualistic nature of man, i.e., the separation of mind and body,” Romano exquisitely argues against the use of key Western philosophical axioms and underscores their racist and white supremacist functions.

In his 1968 essay, “The Anthropology and Sociology of the Mexican-Americans,” Octavio Romano articulates how both gringo and gringo-minded academics of color have for decades deliberately distorted history and created pathological theories about Chicano/as framed within Western logic, objectivity, dualism, and scientific theory, i.e., within their own racist anti-Chicano/a paradigms. Masked in academic theory, scholars advanced calculated myths that Chicano men were machismo, alcoholics, abusers, deadbeat fathers, criminals, and overall threats to American society and their own culture. While Chicanos were theorized as deviants and predators, Chicanas were theoretically pathologized by academics as “supernaturally-fatalistic with a touch of sado-masochism.”

Echoing Octavio Romano’s critique of Western gringo paradigms, Miguel Montiel in his 1970 article “The Social Science Myth of the Mexican-American Family” argues that studies of the Mexican-American family “have relied almost totally upon a psychoanalytic model in which there is an uncritical use of concepts like machismo . . . relegat[ing] explanations of Mexican family life to a pathological perspective.”” Borrowing Western ideas about the nature of man and Alfred Adler’s individual psychology, Latin American scholars such as Samuel Ramos, Leopoldo Zea, and Octavio Paz adopt these European ideas to “explain” the causes of Mexican and Mexican-American inferiority.” As Montiel notes:

Alfred Adler’s theory postulates that children born with hereditary organic weaknesses are inclined to compensate both physically and emotionally in the direction of the defective function. Furthermore, Adler’s theory asserted that “the whole human race is blessed with deficient organs,” and thus there exists a continual resistance to the establishment of a harmonious life situation. Specifically, he claimed that most individuals suffer from a “sense of female inferiority” and as a result “both sexes have derived an overstrained desire for masculinity.”
Adler’s theory universally subsumes all human beings as defunct, broken, and defective. This defectiveness, regardless of race, ethnicity, and culture, leads people to yearn for their masculinity as a way to cope with their inferiority.

Derived from the understanding of the European man and woman and the death, rape, and evilness Europeans wrecked on the indigenous and Mexicans, Samuel Ramos, Leopoldo Zea, and Octavio Paz adapted Adler’s concepts to theorize and ultimately pathologize Mexican and Chicano culture. As Alfredo Mirandé highlights in Hombres Y Machos: Masculinity and Latino Culture, Adler’s individual psychology is center to the overall pathological anti-Brown construction of the machismo Mexican and Chicano:

[The so-called cult of machismo developed as Mexican men found themselves unable to protect their women from the Conquest’s ensuing plunder, pillage, and rape. Native men developed an overly masculine and aggressive response in order to compensate for deeply felt feelings of powerlessness and weakness. Machismo, then, is nothing more than a futile attempt to mask a profound sense of impotence, powerlessness, and ineptitude, an expression of weakness and a sense of inferiority.]18

Chicano/a philosophy must resist the urge to theorize about the Chicano/a condition by adapting European concepts. The gringo imperial system of rationality is grounded on an anti-Chicano/a logic that axiomatically and incessantly categorizes Chicano/a people as threats, defective, and less than human. Theorizing Chicano or Mexican males within gringo logics permanently filters out pathologies that “remains imbued with . . . negative attributes as male dominance, patriarchy, authoritarianism, and spousal abuse.”19 Since Western logics and Western institutions are mutually dependent on each other, Chicano/as are permanently and necessarily imprisoned between anti-Chicano/a logics and anti-Chicano/a institutions. As long as Chicano/as continue to subscribe to Western philosophy, their very existence will always be controlled, manipulated, and exploited by the gringo.

In “The Anthropology and Sociology of the Mexican-Americans,” Octavio Romano asked, “What are these Mexican-Americans who have been created by Social Scientists?”20 After rigorously analyzing the pathological theorization by several academics, he concluded that their research was grounded in the same axiomatic anti-Brown logic as centuries before. As he highlights, “[t]hese opinions were, and are, pernicious, vicious, misleading, degrading, and brainwashing in that they obliterate history and then re-write it in such a way as to eliminate the historical significance of Mexican-Americans, as well as to simultaneously question the legitimacy of their presence in contemporary society.”21 Similarly, Maxine Baca Zinn in “Sociological Theory in Emergent Chicano Perspective,” Alfredo Mirandé in “A Reinterpretation of Male Dominance in the Chicano Family,” “Chicano Sociology: A New Paradigm for Social Science,” and “Rascuache Lawyering: A Chicana/o Vision of Rebellious Law Practice, Pedagogy, and Clients,” set the foundation for a paradigm shift in Chicano/a scholarship.22 Resisting the racist epistemologies, ethics, logics, and metaphysics of gringo theorization, scholars such as Octavio Romano, Alfredo Mirandé, Miguel Montiel, Nick C. Vaca, Deluvina Hernandez, and Maxine Baca Zinn, created new categories of knowledge grounded in a Chicano/a logic where Chicano/as are the creators of reason, theory, history, and philosophy.

II. CHICANO/A POWER, CHICANISMO, AND THE DEVELOPMENT OF CHICANO/A LOGIC

Ignacio M. García in Chicanismo: Forging of a Militant Ethos Among Mexican Americans highlights how key Chicano leaders, such as Reies López Tijerina, Rodolfo “Corky” Gonzales, and José Angel Gutiérrez, unapologetically rejected a liberal and assimilationist ethics. As García notes:

By the early 1960s, many Mexican Americans were disenchanted with traditional politics. These liberal politics centered on an active government that would provide economic development, protect civil rights, and guarantee cultural pluralism. It was an approach that required faith in the established institutions and patience in the face of slow change. It was a steady approach of government action, judicial litigation, and Anglo American leadership. It also required that Mexican Americans wait for the “real” civil-rights problems—those of Black Americans—to be solved before the focus shifted to them.23

Opposed to working within the American system, these scholars sought to create their own institutions framed in Chicano/a power, Chicanismo, and Chicano/a reason. They understood US institutions to be a priori racist and anti-Chicano/a. In José Angel Gutiérrez’s A Chicano Manual on How to Handle Gringos, he presents Chicano/as with a Chicano/a culturalogics to rupture gringo anti-Chicano/a categories and institutions. His manual provides a roadmap to minority empowerment through an effective use of analysis, practical experience, and anecdote. . . . Gutiérrez analyses various types of power and evaluates Chicano and Latino access to it at various levels of US society.24 Gutiérrez was clear that the only way Chicano/as would be liberated from the hands of the gringo was not by espousing an integrationist or assimilationist ethics, but by eliminating the foundation of the system altogether.25

At all points in the gringo system, as José Angel Gutiérrez precisely articulates, Chicano/a existence, his/her very ontology, is always under physical and psychological threat.26 As long as Chicano/as aim to live within gringo Institutions, they will always be foreigners and threats. This is not an error, ignorance, or a mistake in reasoning that can be corrected by a white supremacist anti-Chicano/a US education system, but a sound and valid conclusion where logic is always under the control and manipulation of the gringo. For example, despite the fact that the gringo is, and will always be illegal in the US, through manipulation of laws, logic, education, history, power, and reason, the illegal gringo became legal27 and as Gutiérrez notes in The Making of a Chicano Militant, “[t]hese “white” Anglo
people also began to reverse the definition of "illegal aliens." Illegal aliens now became the Anglo term to refer to Mexicans who ventured into "their" lands without their permissions. They made us, as Mexicans, foreigners; they made my ancestors illegal in their own homeland.

Similarly, Rodolfo "Corky" Gonzales, leader of the Crusades for Justice and author of I Am Joaquin/Yo Soy Joaquin, argued that Anglo-Americans had attempted cultural genocide of Chicanos. Not only was the gringo producing anti-Chicano/a theory, but grounding American institutions in its genocidal anti-Chicano/a logics:

By destroying the self-image of Chicanos, Anglo-Americans were able to rewrite history, demeaning hundreds of years of Chicano experience. Chicanos were caught up in a social reality where they had no past; their leaders were co-opted, their lands confiscated, and their culture demeaned. They were confused in that their anger and frustration were often directed against each other in gang violence and domestic conflicts.

Gonzales’s Crusade for Justice took a militant stance against American white supremacist institutions and "rejected white America in its entirety." Moreover, "It searched within the Chicano community for the answer to problems. To get rid of drugs and gangs from the barrio, Gonzales proposed barrio defense committees. These also would keep watch on police activity . . . [t]o develop La Raza’s self-esteem, Gonzales called upon the school system to teach Mexican American history, language, and culture."

In Gonzales’s speech, "El Plan del Barrio," he was clear that the US education system, at all levels, was a direct threat to the existence of the Chicano/a people.

The curriculum, facilities, resources, educators, and philosophies of the gringo have always been and continue to be central components in the construction and preservation of America’s white supremacist anti-Chicano/a education system. Gonzales knew the US education system was nothing more than a racist institution dedicated to forcing Chicano/a students to assimilate to a gringo anti-Chicano/a ethos. As he noted at the Poor People’s March in 1968 speech "El Plan del Barrio", he was clear that the US education system, at all levels, was a direct threat to the existence of the Chicano/a people.

We demand that our schools be built in the same communal fashions as our neighborhoods . . . that they be warm and inviting facilities and not jails. We demand a completely free education from kindergarten to college, with no fees, no lunch charge, no supplies charges, no tuition, no dues. We demand that all teachers live within walking distance of the schools. We demand that from kindergarten through college, Spanish be the first language and English the second language and the textbooks to be rewritten to emphasize the heritage and the contributions of the Mexican American or Indio-Hispano in the building of the Southwest. We also demand the teaching of the contributions and history of other minorities which have also helped build this country. We also feel that each neighborhood school complex should have its own school board made up of members who live in the community the school serves.

Gonzales was adamant that Chicano/as needed to turn away from looking for gringo approval, ideas, education, and philosophies. Similarly, as founder of the Mexican American Youth Organization (MAYO) and La Raza Unida Party, José Angel Gutiérrez sought to organize Chicano/a activists and Chicano/a youth to unwaveringly fight against white supremacist liberal and conservative US institutions. According to Gutiérrez in The Making of the Chicano Militant: Lessons from Cristal, other pivotal organizations, such as the League of United Latin American Citizens (LULAC) and the American G.I. Forum, were "wanting to become assimilated into the Anglo world, thus leaving behind their Mexican roots and culture. The thought of these organizations dividing our community on the basis of assimilation, culture, citizenship, and class status in addition to age and gender for membership was abhorrent."

Rodolfo "Corky" Gonzales and José Angel Gutiérrez understood that a new type of logic/reason needed to be developed to undergird Chicano/a institutions, grounded in the relationship between Chicano/a history, culture, power, resiliency, and gringo racist anti-Chicano/a institutions. As Rodolfo "Corky" Gonzales noted in his "Arizona State University Speech,”

When we start to think on those levels, when we start to teach in the barrio how we are colonized people, then we’re able to understand how we live in this country and how this economy is based on the farm workers’ struggle and the farm workers’ production. Our people still use their hands in a society that is the most advanced technologically in the whole world.

Gonzales knew that revolutionary action, philosophies, and institution building did not begin with Chicano/as in America, but was tied to deeper historical roots and traditions in Mexico. As Ignacio García notes, revolutionaries such as Emiliano Zapata, Francisco "Pancho" Villa, La Adelita, and La Corregidora were important figures "who had taken on the gringo or other oppressors and had held their own. These new heroes, as much as anything else, signaled a rejection of American political culture. No longer were young Chicanos to be taught to admire the Founding Fathers, American military heroes, or civilian elder statesmen."

Similarly, as Alfredo Mirandé articulates in Gringo Justice, Chicanos such as Tiburcio Vázquez and Joaquin Murieta, "were subject to the double standard of justice and were . . . victims of injustice, rebelled against the dominant order, took the law into their own hands, and were admired and respected by the Mexican population." They knew they could not rely on a gringo legal system where justice is nothing more than the maintenance of a white supremacy ethos and the criminality of the Chicano/a. While Chicano/as continue to be creators of history, they must also continue to recover a history that has been rewritten and stolen from them by the gringo. This is a necessary step in the rupture of pathological epistemologies that categorize the
Chicano/a as a permanent nonhuman and in the creation of new forms of knowledge grounded in the cultural logics of the Chicano/a.

Anti-Chicano/a epistemologies are upheld by a violent gringo legal system that a priori and necessarily constructs the Chicano/a as a permanent threat to its civilized and rational institutions. It is one of the most indispensable tools of dominance over the Chicano/a. The Treaty of Guadalupe Hidalgo, for example, has been pivotal in the construction and maintenance of the Chicano/a as a nonhuman, threat, and alien. As eminent scholar Rodolfo F. Acuña highlights in Occupied America, “[t]he most publicized of these incidents occurred on June 5, 1967, with the Tierra Amarilla Court House raid involving both the citizens and the local district attorney. This incident culminated with a shootout between authorities and Alianza members, which became a part of the legacy of deeds stimulating the Chicano movement.” As Tijerina articulates in They Called Me “King Tiger”: My Struggle for the Land and Our Rights, “[t]his Alianza of communities and land grantees would become the terror of those who had stolen our land and destroyed our culture.”

Chicano/a philosophy, at its most axiomatic foundation, must continue Alianza’s commitment to self-defense against gringo terror. Utilizing nonviolence and civil disobedience as a tactic “to actively refuse to obey unjust laws and injunctions,” César Chávez, “who in 1965 headed the National Farm Workers Association (NFWA) in California . . . called his first huelga (strike) against California growers. Citing the Grito de Dolores, or ‘Cry of Dolores,’ which launched Mexico’s independence movement, Chávez proclaimed a struggle to liberate the Mexican American farmworker.” Along with Dolores Huerta’s leadership and commitment to fighting gringo injustice, Chávez and the farmworkers “had taken it upon themselves to change their condition, and they were doing it despite harassment from law enforcement agencies. This audacity to challenge agribusiness and its allies in state government galvanized a whole generation of Chicano youth to join the struggle for Mexican American civil rights.”

A complete rejection of an integrationist and assimilationist logics, anti-Chicano/a gringo ethics, and philosophies centered on the need for gringo recognition, Chicanismo as a twenty-first-century axiom must seek to build on a Chicano/a logics that centers Chicano/as as creators of reason, knowledge, history, and philosophy in relationship to, as Tommy Curry articulates in his revolutionary and paradigm-shifting book The Man-Not, “the grammar of racism, its discursive logics, that legitimates and subtly produces the logics of (genocide)/violence.” Curry argues for the creation of new liberatory institutions, such as Black Male Studies, where Black male existence is the “bridge between the reality of the world and failure of theory/thought/reason to capture the world due to its own obsession.” Furthermore, as Curry asserts:

To be seen as he truly is, the Black male must be configured within a new history that tells of his complexity, his embracing of the mother right, his anti-imperialism, his anticolonialism, his Black socialism, his Pan-Africanism—simply put, his struggles to realize himself within his own experiences, meanings, and formulations. He must be thought of as a traveler of as reflective and deliberate but flawed—as a kind of human, not perfect, but not condemnable. He is not the perfect subject but a worthwhile subject/suspect of study and theory.

Similarly, Chicano/a philosophy must be a discipline in which Chicano/as can study and theorize about themselves
III. CHICANO/A PHILOSOPHY FOR THE TWENTY-FIRST CENTURY: BUILDING NEW CHICANO/A PARADIGMS WITH CRITICAL RACE THEORY AND RICHARD DELGADO’S RUPTURE OF GRINGO KNOWLEDGE, REASON, AND LAW

As a pioneer of Chicano/a intellectual thought and a founder of critical race theory (CRT), Richard Delgado’s ideas have been paramount to creating new concepts and epistemologies grounded in Chicano/a paradigms and logic. Injecting storytelling as a tool to dismantle the gringo legal system, Delgado’s The Rodrigo Chronicles brings to light the stories of the outgroup. As Delgado eloquently notes, “stories create their own bonds, represent cohesion, shared understandings, and meanings. The cohesiveness that stories bring is part of the strength of the outgroup. An outgroup creates its own stories, which circulate within the group as a kind of counter-reality.” Delgado’s stories subvert gringo anti-Chicano/a logics and ruptures the bundle of presuppositions, received wisdoms, and shared understandings against a background of which legal and political discourse takes place. The Rodrigo Chronicles shed light on the anti-Chicano/a nature of legal reasoning and its impact on creating and sustaining anti-Chicano/a institutions.

In Richard Delgado and Jean Stefancic’s article “Norms and Narrative: Can Judges Avoid Serious Moral Error?” they detail how the gringo legal system is limited in avoiding errors of logic, reason, and morality. Since most law practitioners have been trained to think about law from the perspective, norms, and logic of the gringo, “one obvious explanation for these mistakes is judicial inability to identify, imaginatively, with the persons whose fate is being decided.” Since judges construct their logic and norms within an already prescribed anti-Chicano/a social reality, to exist as Chicano necessarily entails existing as a threat, criminal, and machismo:

“We decide what is, and almost simultaneously, what ought to be. Narrative habits, patterns of seeing, shape what we see and that to which we aspire. These patterns of perception become habitual, tempting us to believe that the way things are is inevitable, or the best that can be in an imperfect world. Alternative versions of reality are not explored, or, if they are, rejected as extreme or implausible.”

Within a white supremacist American social reality/paradigm where Chicano/as are conceptually, logically, and institutionally constructed and categorized as necessarily nonhuman, there is no alternative version of reality that can be envisioned or constructed for them.

This system is institutionally, legally, and morally fixed within the permanency of gringo anti-Chicano/a logic. As such, “an unfamiliar narrative invariably generates resistance; despite our best efforts, counterstories are likely to effect at most small, incremental changes in the listener or reader.” Since equality, morality, justice, and fairness is the antithesis of the Chicano/a, complete freedom and liberation from gringo wrath is permanently illusory. As Richard Delgado highlights in his article “The Social Construction of Brown v. Board of Education: Law Reform and the Reconstructive Paradox,” “[b]ecause every social practice is part of an interlocking system of other practices, meanings, and interpretations, changing just one element (for example, school assignment rules) leaves the rest unchanged . . . [i]t is as though legal decisions take place against a gravitational field, with the pull being toward the familiar, toward stasis.”

Said differently, since the criminality and illegality of the Chicano/a is a priori logically and institutionally axiomatic in all aspects of America’s social reality, simply changing one aspect of it, e.g., through legal, education, or economic reform, does nothing to change the overall reality of the Chicano/a. His/her illegality and criminality is permanently upheld within an anti-Chicano/a logics that undergirds America’s white supremacist institutions. Utilizing Brown v. Board of Education as an example, Richard Delgado highlights the control and manipulative power gringo logic/legal reasoning has over legal reform in the United States:

“Any text, including a legal one, is interpreted against a background of meanings, presumptions, and preexisting understandings. If a parent tells a child “Clean up your room,” the terms “clean” and “room” have relatively well agree-upon meanings: The child knows he or she is not expected to launder the drapes or vacuum that attic space above the room. If an adolescent tells the parent, “I’ll be back by midnight,” both understand that “midnight” means tonight, not next week, and that “back” means inside the house. The same is true of legal commands. Thus, when Brown ordered school districts to desegregate “with all deliberate speed,” southern officials interpreted the decree in terms of their common sense. In hundreds of close cases, they constructed Brown to mean the only thing it could mean, consistent with their experience: integration . . . that left the school system as intact as possible.”

Reform, of any kind in America, will always seek the recognition and approval of the gringo. Delgado’s work provides Chicano/as with a treasure trove of insight. His ideas not only detail the limitations of an anti-Chicano/a US legal system, but also, much like the Chicano/a intellectuals and activists before him, provides readers with a blueprint to build Chicano/a logics and Chicano/a philosophy.

Several contemporary philosophers are working on issues that are centered around Chicano/a thought. In 2015, the Society for Mexican American Philosophy (SMAP) was officially established to examine issues and themes that pertain to the Chicano/a condition. As a founding member of SMAP and central figure of Mexican American philosophy, Carlos Alberto Sánchez, has played a pivotal role in changing the landscape of professional philosophy by leading the surge of Mexican and Mexican American
philosophy. In his article, “On Documents and Subjectivity: The Formation and De-Formation of the Immigrant Identity,” he discusses the phenomenological relationship between having papeles and “the possibility of assuming a stand-taking subjectivity.” As he notes:

When these formative documents are challenged . . . one’s identity or stand-taking subjectivity is also challenged . . . this is the case with legislative acts like Arizona Senate Bill 1070 which challenges the force of those documents by targeting all individuals who are reasonably suspected of not having them, such legislation directly puts into question the identity and subjectivity of those who, in fact, have the documentation that authorizes their trespass.72

Sánchez shows that being Chicano/a necessarily entails being questioned about one’s own existence and legality. Despite having one’s papeles, a Chicano/a may still be legally condemned to exist in a space of nonexistence.73

Utilizing “Edmund Husserl’s prescriptions for the phenomenological method,”74 Carlos Sánchez shows there is a relationship between being conscious of an object, e.g., green cards, and the value one places on it. Said differently, “the value of the document depends on the consciousness that perceives it.”75 As he notes, “[t]heir value resided in what they represent. In its textual configuration, a green card represents the legal and, ultimately, the authentication of the person to whom it belongs. Having a green card in hand is socially, culturally, politically, and existentially significant.”76 Possessing a green card, Sánchez believes, is significant because it authorizes one to operate within the sphere of American law and offers security.77 Provided this, “having a green card means that one is not a criminal. The others, who cannot present proof of this absoluteness, continue to exist as criminals—they are ‘illegal’ ‘aliens,’ external to the space of law or the space of intelligibility. Outside of this space, the document-less exist as incomprehensible.”78

The concept of the racialized legal immigrant is critical to Carlos Sánchez’s analysis. In other words, even those who possess their papers live a life of contingency. As long as one looks undocumented or possesses gringo-defined undocumented characteristics, then his/her existence is contingent upon gringo institutions. Utilizing Martin Heidegger’s concept/logics of equipment, Sánchez in “Illegal’ Immigrants: Law, Fantasy, and Guts” underscores how Mexican immigrants ultimately become obtrusive to gringo society and targets of gringo violence.79 As he posits, “The ‘illegal’ immigrant thus finds herself paradoxically outside the space of law but inside the realm of capitalist production, as equipment. She is both in and outside the space of life and culture while always outside the space of rights—she is both valued and reified.”80 Laws, policies, treaties, US programs, and bills that both exploit the use of “illegal” immigrants and Chicano/as, while simultaneously constructing them as nonhuman, indicates their bodies are “objectified so as to complement other household appliances; it becomes an extension of the vacuum cleaner, the lawn mower, or the hoe.”81

“Illegal” immigrants and Chicano/as are thingified and denied as Dasein/human beings.82 Carlos Sánchez contends they become equipment, and as such, their dispensability is at the hands of the gringo. Sánchez observes, “when the equipment does not meet our needs, it suddenly shows up everywhere; we see it every time we are set to engage in our tasks. It will frustrate us, as it is there, unable to help us with our project . . . equipment may also become obtrusive and obstinate when it fails to meet our demands.”83

Like equipment that is no longer needed, the US rids “illegal” immigrants and Chicano/as at their discretion. Gringo institutions, “thingifies” the ‘illegal’ immigrant, simultaneously values and devalues the very being of persons. It devalues a person by turning her into a tool; it values her by emphasizing her irreplaceable importance in the American economy, one in which her absence would be immediately felt.84

Carlos Alberto Sánchez’s work is a must read, including his monumental texts, Contingency and Commitment: Mexican Existentialism and the Place of Philosophy, The Suspension of Seriousness: On the Phenomenology of Jorge Portilla, and his most recent book co-edited with fellow SMAP founding member Robert Sánchez, Jr., Mexican Philosophy in the 20th Century: Essential Readings.85 Chicano/a philosophy must not stand alone in its relationship to the gringo world, but must also be grounded in its historical relationship to Mexican and Indigenous philosophy. Sánchez’s work is vital to uncovering and revealing these relationships. Despite Sánchez’s impact and influence, much of his work, such as “On Documents and Subjectivity: The Formation and De-Formation of the Immigrant Identity” and “Illegal’ Immigrants: Law, Fantasy, and Guts,” relies on the use of gringo theory to analyze the condition of people of color. In Contingency and Commitment: Mexican Existentialism and the Place of Philosophy, Sánchez notes the use of gringo ideas/theory is an act of appropriation.86 As he explains, “[a]propriation is not assimilation, or mimicry, but a simultaneous taking and altering for the sake of some end. But the end of the taking possession-of of appropriation is not to preserve, it is not an embalming of what is possessed; rather, the end is transformation, of world or one’s place in it.”87 Appropriating a system of logic that is axiatically anti-Chicano/a and controlled by the gringo is not transformative, but the creation of a stronger system of anti-Chicano/a knowledge and reason that is legitimized by people of colors’ appropriation of it. Contrary to Sánchez’s approach, the gringo system in its entirety must be rejected for transformation to take place.

In “Illegal’ Immigrants: Law, Fantasy, and Guts,” Carlos Sánchez’s thesis relies heavily on the Heideggerian concept of Dasein/human being. Sánchez appropriates this concept to explain the recognition and value the “illegal” immigrant receives as equipment in the gringo world. As Sánchez notes, “[t]he ‘illegal’ immigrant thus finds herself paradoxically outside the space of law but inside the realm of capitalist production, as equipment. She is both in and outside the space of life and culture while always outside the space of rights—she is both valued and reified.”88 Chicano/a philosophy axiatically rejects the logic that the Chicano/a can ever be recognized/valued in gringo society, even as equipment. The Chicano/a in gringo
society is always/permanently denied his/her Dasein. Dasein and gringo ethics should not be understood as Chicano/a possibility, but rather, Chicano/a impossibility. Said differently, in the gringo world, the Chicano/a is always an impossibility, always outside the space of life. Applying gringo theory, as Sánchez has, provides the “illegal” immigrant with an illusion, the possibility to be valued and recognized as equipment. Always the nonhuman, the Chicano/a has no value and possibility within the gringo world. Chicano/a value emerges, not by working within or with anti-Chicano/a gringo institutions, but by the creation of Chicano/a institutions, reason, knowledge, and logic that axiomatically grounds their culture, history, and existence as necessarily possible.

Similarly, in “On Documents and Subjectivity: The Formation and De-Formation of the Immigrant Identity” and “Illegal Immigrants: Law, Fantasy, and Guts,” Sánchez relies on a Hegelian logic to develop a “phenomenology of documents.” Ultimately, this leads him to conclude there is a level of recognition that is achieved by people of color who have their documents. In other words, once the “illegal” immigrant or one who shares similar characteristic has documents, he/she becomes visible. As Sánchez posits, “[c]odified in this way, the legal immigrant is the most visible (opposed to the invisibility of the “illegal” immigrant), and as such, subject to the most scrutiny, surveillance, and harassment when the necessarily racial prerogatives of anti-immigration law are set loose.”

Turning away from Hegelian logic and toward Chicano/a logic, it is clear the “illegal” immigrant and Chicano/a is never visible or recognized as an actual human in the gringo world, but as permanently invisible, “constrained/ castrated within the existing taxonomies/categories built on the assumptions of a white Western man.” Theorizing from Chicano/a philosophy as a starting point allows Chicano/a to create in their reality and intersubjectivity as opposed to being constructed from anti-Chicano/a logic, epistemology, and white violence.

IV. CONCLUSION

Keeping in mind the logic and principles that have already been set in place by Chicano/a intellectuals before me, a twenty-first-century Chicano/a philosophy must be grounded in the following axioms: 1) gringo theory, logic, and paradigms are psychological and physical threats to the Chicano/a people; 2) Chicano/as create their own axioms, paradigms, logic, and epistemologies independent of contamination from the gringo world; 3) Chicano/a philosophy is a complete rejection of gringo recognition, norms, humanity, logic, and reason; 4) Chicano/a philosophy must be constructed from the histories, logics, and epistemologies of the Chicano/a people, not from the axioms of their gringo nonhuman construction; 5) Chicano/as are permanent threats and foreigners in the gringo world; and 6) the gringo world is not broken or ignorant, but operates as the racist and violent institution it was created to be. As Chicano/a philosophy moves forward, it must turn away from seeking gringo recognition and validation. It must be a complete rejection of gringo institutions, norms, reason, logic, and morality. Chicano/a philosophy is unapologetically and unwaveringly a discipline for the study and theorization of Chicano/as and for the construction of Chicano/a institutions grounded in a Chicano/a logics that necessarily entails Chicano/a possibility and existence.

NOTES


4. Ibid., 2.


6. Ibid., 31.

7. Ibid., 30.


15. Ibid.


17. Ibid.


19. Ibid., 66.


21. Ibid., 54.


25. García, Chicanismo, 57.

26. Ibid.


28. Ibid.

29. García, Chicanismo, 34-35.

30. Ibid., 35.

31. Ibid.

32. Ibid.

33. Ibid.

34. Gómez-Quiñones, Chicano Politics, 114.


40. Ibid.

41. Alfredo Mirándé, Gringo Justice (Notre Dame: University of Notre Dame Press, 1987), 77; see also García, Chicanismo, 50.

42. Mirándé, Gringo Justice, 77.

43. It is important to understand that I am claiming that US institutions are indeed constructed to be civilized and rational. It is not the case that these institutions are broken or in need of rehabilitation. Said differently, to be civilized and rational within Western logics is to necessarily be gringo. Whereas, to be Chicano/a within Western logics and the US necessarily entails existing as a permanent nonhuman. Another way to highlight this: concepts such as reason, rationality, and knowledge are constructed within Western logics and the US to necessarily entail the construction of the Chicano/a as criminal, irrational, deviant, and alien.

44. Acuña, Occupied America, 51.


46. Acuña, Occupied America, 53.

47. Rendon, Chicano Manifesto, 72.


49. Ibid.

50. Ibid.

51. Ibid., 116.

52. Ibid.


54. Acuña, Occupied America, 297.

55. Ibid.

56. García, Chicanismo, 30.

57. Ibid., 30-31.


59. Ibid., 225.

60. Ibid., 226.

61. Ibid., 222.


64. Ibid.

65. Ibid.


70. Ibid., 246.


72. Ibid.

73. Ibid., 204.

74. Ibid., 200.

75. Ibid., 201.

76. Ibid.

77. Ibid.

78. Ibid.


80. Ibid., 6.

81. Ibid., 5.

82. Ibid.

83. Ibid.

84. Ibid.

85. Carlos Sánchez, Contingency and Commitment: Mexican Existentialism and the Place of Philosophy (Albany: State University of New York Press, 2016); Carlos Sánchez, The
Sánchez, Contingency and Commitment, 4.
87. Ibid.
90. Ibid., 204.
91. Ibid.
92. Ibid., 204-205.

BOOK REVIEWS

Coloniality of the U-S///Mexico Border: Power, Violence, and the Decolonial Imperative


Reviewed by Manuel Chávez
MONMOUTH UNIVERSITY

The title of Hernández’s book itself is an intentional subversion of the common label of “US-Mexico border,” which reflects his critique of the border between Mexico and the United States as an inevitable feature of nationhood and territory. The history of the border has long been studied, but Hernández argues there is a logic of power that predates and shapes this history of the border and continues to frame how it is perceived today. Although he situates his book “at the interdisciplinary crossroads of urban studies, border studies, and ethnic studies,” Hernández’s work can be read as a post-Occidentalist philosophy of history (27). His main concern, however, is the causes and effects of violence among those who live at or near the border between Mexico and the United States. The origin of Hernández’s study can be found in his childhood experiences growing up in San Ysidro; particularly the 1984 massacre at a local McDonald’s restaurant sowed the seeds for his thinking. While his work is an analysis of the historical power dynamics that construct the current border, it also is a call for the decolonial imperative as a way to resist the logic of violence that underlies it.

Relying on the concept of coloniality articulated by Aníbal Quijano and María Lugones, Hernández identifies the historical logic of power that operates through the conflated social identities of class, race, ethnicity, and gender/sex identity, and is both obscured and reproduced by the nation-state centered ideology of modernity. Modernity/coloniality is a “matrix of power” that is rooted in colonialism, but endures in society—at the level of social ontology and epistemology—after the end of formal political independence. While recognizing the political and economic changes over time, Hernández applies the coloniality framework to foreground the evolution of a persistent domination manifested through the mutually constitutive ideological, territorial, corporeal/racial, and cultural/symbolic practices of violence (23). The concept of coloniality operates as a critique of modernity, while also inferring the decolonial imperative.

As social theory, Hernández claims coloniality is best understood as what Kryíakos M. Kontopoulos names a “heterarchial theory,” that is, one that overcomes the “macro-micro divide,” by recognizing “a multiplicity of overlapping mechanisms at various levels—what geographers would see as multiscalar analyses” (23). Consequently, modernity/coloniality does not refer to a single deterministic system, but rather the “totalizing” conditions of the historical relations of power. These conditions include boundaries of social identities as well as nation-states (10-11). The border, as a construct of modernity/coloniality, serves to separate the “the modern” and the nonmodern, “the presumably backward, traditional, or primitive” (26).

Violence at the border, Hernández contends, is an effect and cause of modernity/coloniality. While the border region is portrayed in the media as an area of natural lawlessness, he argues recognition of the coloniality of the border reveals “the utility of violence as a tool wielded by many in variegate pursuits of power and domination, both physical and symbolic” (12). Coloniality reveals “a continuum of violence” from the colonial period to the present. Consequently, he claims the border region “is politically and materially marked by different forms of racial/colonial violence, particularly for Mexican@s, Chicán@s, and Latin@s who live near or on the both sides of, or who cross, the U-S///Mexico border” (186). The significance of the decolonial imperative is that it considers, to borrow a term from Alicia Gaspar de Alba, an “alter-Native” world without borders.²

In the first chapter, Hernández explores how coloniality of power frames the immigration debate. He traces how the modern notion of nationhood, conceived through the metaphors of home and family, is based on the distinction between “insiders” and “outsiders.” It is embedded in the “gendered discourses of home and family [that] function to reproduce Eurocentric and heteronormative narratives of nation, property, citizenship, and belonging” (39). He contends anti-immigrant nativists invoke this colonial logic in their political ideology: “Mexican migrants appears as heir to the role of the Indigenous ‘savages’ of yesteryear” (28). However, he argues “insider” critics of border vigilantes, while rejecting their “extremist” political ideology, nevertheless accept the colonial logic of the nation-state (e.g., contemplating “immigration” instead of “migration”). “The result is an effective legitimating of a kinder, gentler racism on the ‘inside,’ including the mainstreaming of anti-migrant politics/legislation” (39). In contrast, the decolonial imperative implies the need to question “the premises, the logics, and the episteme that underpin boundaries” (65).

In chapter two, Hernandez examines how the coloniality of power operates at the local, regional, and global level
ultimately and interactively. In his examination of the local history of annexation of San Ysidro under San Diego, Hernandez not only highlights its relation to expanding capitalist globalization and its proximity to the border, but also the racialization of geography and the marginalization of communities of color, primarily Mexican but also Black, Filipino, and Muslim. The “metropolitan colonial” situation of San Ysidro serves as the backdrop of his examination, in chapter three, of the mass shooting that took place at a local McDonald’s in 1984. While mainstream media accounts individualized the attack, Hernandez brings to bear a structural analysis that takes into account the factors of globalization and race which led an unemployed white Anglo male from a deindustrialized town in Ohio to shoot up a restaurant near the border. He argues “the shooter’s actions were a manifestation of recurrent colonial logic in which Indian-hating on the frontier is sublated into its heir: Mexican-hating on the border” (30). The coloniality of power is not only revealed in this violent event, but just as significantly, also in its aftermath, in the struggle to memorialize the massacre.

The debates over the monument to the nineteen victims of the shooting point to the coloniality of memory and knowledge. In chapter four, Hernandez extends this argument by bringing attention to the phenomena of feminicide in Ciudad Juárez. He contrasts the official dominant discourses surrounding the deaths of women at the border with opposing subaltern viewpoints represented in music and literature. Cultural texts, such as the corrido “Los Crímenes de Juárez” by Los Marineros del Norte and the novel Desert Blood: The Juárez Murders by Alicia Gaspar de Alba, offer counter decolonial perspectives on making sense of the violence against women along the border. In chapter five, he further elaborates his argument for escaping the “the epistemic/cartographic prison of modernity/coloniality” (167). He criticizes the “borderlands academic complex” that reproduces epistemologies abstracted from the “embodied understandings and the intersections of violence and colonization” (157). Echoing Alejandro Vallega’s “decolonial aesthetic turn,” Hernandez’s close examination of the “sonic geography” expressed in songs by Tijuana No! featuring Kid Frost, Los Tigres del Norte, and Aztlán Underground suggests a resistant conceptualization grounded in the materiality of the borderlands which disrupt the logic of modernity/coloniality. The decolonial imperative entails an “epistemic and cartographic disobedience” that shifts to a different way of knowing.

What Hernandez offers is a philosophically provocative way to rethink the hegemonic discourses about the politics of the border. The problematic he puts forward allows us to take into account the links between the legacy of colonialism and the 1984 McDonald’s shooting as well as the 2019 Walmart shooting, family detention centers, and family separation policies. He makes clear the decolonial imperative does not imply a type of cosmopolitanism or “open border” politics. In order to challenge modernity/coloniality, it is the demand to shift “differentially” between ethics and meta-ethics by way of taking indigenous social ontologies as an epistemic starting point. Ultimately, Hernandez provides a political and moral “philosophy born of colonial struggle” that can help us imagine and create resistant possibilities together.

NOTES


4. See Chela Sandoval, Methodology of the Oppressed (Minneapolis: University of Minnesota Press, 2000).


Lessons in Exile


Reviewed by Roy Ben-Shai

Sarah Lawrence College

When I first saw Pereda’s book, I thought, “Oh, that’s nice. It’s a short book.” I was excited to read it and have a conversation about it. But I’ve been deceived. This is not a short book; it’s a journey. I understand better now why the Prelude to the book is called “A Map for the Road.” While it is common, of course, that good philosophy books should have a developmental aspect, it is rare to find a philosophy book with so many twists, turns, and surprises. I mean, radical surprises.

As I began reading the book, I was already aware that I was going to come here and present it for you. So, while reading, I was already thinking about a way of presenting the book. At some point, two or three chapters in, I felt like I found it. Yes, I believed to understand the essence of the project and the manner by which it was unfolding and came up with an angle for sharing it with you as well. But then, as my reading progressed beyond the first few chapters, I could feel my face taking on the expression of someone who is thinking to himself: “No, this is not going according to plan . . . [and to the author:] What are you doing?”

And a few pages later, the dawning of renewed understanding: “Oh, now I see, this is the plan. This has been the plan.” But this did not last very long either. To the very last pages I had to go on reassessing the nature of this work.

So, I now have a problem. It is very difficult to discuss this book while doing justice to what I think is the most essential and distinctive aspect of it: that you don’t know where it’s going until you get there. So, the first thing to say is that you should buy the book and read it from cover to cover. Because that’s how it works. In the meantime, what I’ll do
is share my process of revelation. To do this, I will start from my initial insight: what I was planning to say when I thought I knew what the book was about. And then I’ll describe how things changed and unfolded. So, here is my initial angle:

Let’s begin with the title: Lessons in Exile. Break it down: “lessons” in “exile.” The word “lessons” is already suggestive. It is an interesting term. Surprisingly, perhaps, it is not often used in titles of philosophy books. There is a difference between a theory of exile and lessons in exile. And the difference can be construed as that between a top-down approach—an umbrella definition of what exile is—and a bottom-up learning from exile, about exile. The latter approach is the kind that Kant called “reflective judgment”: We begin from experience and generalize from it.

Exile, then, is a form of experience, which brings us to the second term in the title. What kind of experience is “exile”? Already in the most general and ordinary understanding of the word, we can say that exile is an experience that is out of the norm, that breaks with what is customary and accustomed. Etymologically, all the word means is this: “wondering out.” And when you think about it, if any experience could teach us a lesson, it goes to reason that this kind of experience would have a special place. After all, regarding our customary and habitual mode of experience we perhaps don’t have that much to learn, because we are already well-versed in it.

But exile is always something unexpected, breaking with expectation and habit. In that respect, exile can even be thought of as a paradigm for lived experience as such. And indeed, the word “experience” is very similar to the word “exile.” It literally means “trying out.” The “ex-” factor is there in both of them. To experience, like exile, is to go out.

Well, I was thinking, initially, that I should probably be cautious about saying that exiles can be a paradigm for experience in general, because that’s a very broad statement for someone who has not yet learned the kind of lessons exile has to teach. I did not anticipate, in fact, just how far Pereda’s book would end up going in advancing this claim.

The title aside, another impression I got is that there is a certain note of lamentation, almost nostalgia, as Pereda recalls the existence of an old and illustrious tradition of lessons in exile. He finds it in the Hebrew Bible, and in Greek and Roman poetry and philosophy. And the concern that seems to me to animate the book from its inception is that we have lost this tradition.

In the first chapter, “Words, Words, Words,” Pereda distinguishes between “exile” and related words like “refugees” and “emigrants.” And he warns us, “not to overlook a . . . maxim, often repeated yet seldom heeded: Be careful with words!” (4). Pereda’s point is that words resonate in a certain way. They appear within a certain discursive context. Specifically, the word “refugee,” is one of the main items of political discourse in our time. Now, here, I take some liberty, and I may be mistaken in my interpretation, but this is what the distinction between the resonances of the words “exile” and “refugee” made me think about, and in relation to what I take to be Pereda’s lament. The manner by which “refugees” are invoked and discussed in our mass media is always by way of massification. There are thousands, millions of refugees; faceless, nameless, crowding refugees. And this massification of the phenomenon (which is, indeed, massive) goes together with a certain type of sensationalism about the horror of destitution.

The horrifying aspects of the refugee condition notwithstanding, one thing we simply cannot do with it is learn. There is little lesson to be had. And part of the problem is that, insofar as we think about it this way, talk about it this way, there is no subject to such massified condition either. And, for Pereda, the existence of a subject is very, very important. Where there is no subject, there is no experience, let alone reflection or judgment.

So, the word refugee, as it is most often invoked, resonates with massification, and I would add to this that, on the opposite extreme, we have the equally discursive tendency for radical individualization and emotionalization of experiences of victimization; an over-subjectification of experience: “This is my experience, and you cannot have it. You have no right to it, no access to it. You are too privileged and would not understand.” All we have left is to be shocked, but no lessons, nothing to learn.

And so, in a somewhat “Peredaesque” fashion, I would formulate this into a principle: two conditions must hold for there to be something like lessons in exile, or in any kind of experience: one is that there must be a subject. And the other is that it cannot be simply about this subject. The experience, and the subject’s testimony and reflection about it, is but the ground for learning and teaching. It must go out.

I should note that one of the central methodological principles that Pereda postulates, right at the opening of the book, is, and I quote: “Strive to rid social phenomena of their own inner barriers and strive to do this again and again” (1). What I understand this to mean is that, yes, experiences, especially those involving victimization, can be very, very difficult; very singular, and almost impossible to share. They have these inner boundaries, and yet we have to strive not to break these boundaries but to expand them from within, to press out against them. And we must to do this not once and for all, but, rather, again and again. And I think that is what the book keeps doing, again and again, expanding those boundaries of experience.

Much later in the book, in Chapter 6, “Words Say, Words Resonate,” Pereda returns to the theme of words to advance what appears to be a theory of resonance and resonation. Resonance is akin to such expanding. We begin from a particular experience. We expand it to certain types of experience. And then we expand it to a certain perspective on experience in general. But we cannot grasp experience, especially such exceptional experience, from the top down or in abstraction. We have to work from within and to push out like that. Again and again.

This sense of resonance can help to understand a fascinating distinction. In the second chapter, Pereda
draws a distinction between “testimonies” and “meta-testimonies.” The idea that, if a testimony is the account of a particular experience, meta-testimony—which is a form of writing Pereda essentially associates with poetry—is one that already takes a step beyond the particular. And why poetry? I take it that it is because poetry is art. There is, let us say, an element of stylization about it. We make an art of the experience, and that is no longer a straight-up testimony but something more, something beyond. I am almost tempted to say that this notion of meta-testimony harbors a potential critique of the Adornian dictum (or at least the superficial reception of it) that there can be no poetry after Auschwitz. Well, there must be poetry after Auschwitz, if we are to have something like a meta-testimony. And the kind of meta-testimonies that Pereda is particularly interested in are those poetic accounts of exile where the first person becomes third person, where the poet suddenly rises beyond herself and her experience.

Later in the book, it becomes clear that there’s a certain idea of how an experience that is had by someone becomes an experience that is had, or can be had, by anyone. This “anyone,” like the poet’s third person, does not symbolize an abstract universality, like a transcendental subjectivity available to all. It is, rather, an expansion, a resonance, a reaching out and rising past.

And now comes the point where surprises begin to amass, where, for a minute, I had lost the thread of the book, or what I thought the thread was. The first surprise occurred as I was reading through Pereda’s truly masterful literary analyses of meta-testimonies, which occupy Chapters 3 to 5, the epicenter of the book. It gradually dawned on me that through these poetic accounts, Pereda was conducting something like a critique of the experience of exile, and even of the testimonies themselves. “What are you doing?” I was asking him in my mind. Isn’t the whole point that we should learn from them? Isn’t the point that we should learn from the testimonies what it means to be in exile?

But, no, I was mistaken; perhaps I was still not sufficiently “careful with words.” Slowly, I remembered that third word, the middle word, in Pereda’s title: the word “in.” It became clear to me that what he was aiming at is, quite literally, lessons in exile, not about it. The question is how to be in exile (the book’s title in Spanish, Los aprendizajes del exilio, makes the point even stronger, as it suggests an apprenticeship, which is, indeed, a kind of learning from and through experience, through practice). But why, I am still asking, are we to learn about how to be in exile, if we are not in exile? What kind of lesson is this, and why should we heed to it?

Let me first give you a brief overview of Pereda’s analysis of the meta-testimonies. Focusing his analyses on a “few meta-testimonies that . . . belong to an immediately present past, a memory that is still living” (29), namely, the exile during and following the Spanish Civil War, and the even more recent exiles of the Latin American Southern Cone, he divides them into three groups, thus forming a typology of sorts of characteristic experiences, or perhaps dispositions, in exile: exile experienced or regarded as loss, exile as resistance, and exile as a new beginning. Each of them, as it turns out, is animated by a dominant affect: melancholy (loss), anger (resistance), and elation (new beginning).

Reading past these daunting sections of the book, entering the last third of it, a new realization lights up. This is not simply a critique of exile; it is a critique of reason. It is a critique of the relationship between affect and reason. And there is something much bigger that is happening, because the concern for Pereda is the sort of thing that Kant calls “pathology,” which is a certain form of reasoning or rationalization that stems from, is fed by, and enslaved to dominant passions—melancholy, or anger, or elation—which simply reiterate, and reiterate, and reiterate. By Chapter 6, where he concludes his reading of the poems, Pereda names this form of reasoning “arrogant reasoning” (78) as an antidote to which he promotes, throughout the book, an “art of self-interruption” (25).

In light of this, I find his critique of the three types of exile reminiscent of Nietzsche’s early essay on “The Advantages and Disadvantages of History for Life.” There, too, Nietzsche distinguishes between three attitudes towards history and historiography (monumental, antiquarian, and critical), and there, too, rather than making a choice or drawing up a hierarchy between them, he points out what is truthful about each of them, and at what point they become excessive, losing sight of why and how history matters at all. Similarily, for Pereda, there is a truth to this experience of loss and of resistance and of new beginning, but it cannot go too far. And here we find the lessons, the actual lessons, that Pereda articulates in the form of principles: “There is a time to be involved,” that is, for dwelling in loss, “and a time for stepping away,” that is, for moving on (40). “There is a time to resist and a time to break with the situation being resisted as well as with resistance itself” (54). “There is a time to welcome, even encourage, large and small discontinuities in our experience and a time to carry on with already proven plans and deep-seated routines” (71).

Lessons in exile. Whose exile is it? For whom these lessons, and by whom?

I will give you a taste of what happens in the book’s final chapter (“Nomadic Cultures and Personhood”). Pereda refers to it as a theoretical section of the book, but in my view, this is something of an understatement: this section is in effect a crescendo. At this point, it occurred to me that the book could have been titled differently. Not “Lessons in Exile,” but “A Critique of Nomadic Reason.”

For in this chapter, Pereda embarks upon an interpretation of Kant’s Groundwork for the Metaphysics of Morals. He re-envision this book as if it had been written by an exile. And this is pretty astonishing. Let me read it to you in Pereda’s own words:

I will avoid engaging with Kant’s technical, powerful and perspicacious arguments, and I will try to read him as if he were an exile who, terrified and in a rush to leave, had left the manuscript at home. I will also imagine the exile Kant as retaining some rather vague idea of the norms he proposed in his famous book, and reworking, even
transforming, them from this new heartrending situation. So, instead of a “close reading” like the one I proposed for exile poetry, I recommend now a “distant reading.”... (94)

So, what is the point here, of this innovative and indeed distant, if not irreverent, rereading of Kant? If Kant is known for one thing, it is for the rigid separation between transcendental and empirical subjectivity, and for his unflinching grounding of practical reason and the critique of practical reason in the transcendental. What is Pereda telling us, then? What if the ground for practical reason were nomadic? What if practical reason went on exile, forced into exile, and out of its element of native stability? Pereda’s exilic, nomadic Kant, however, is still very much Kant. He does not leave himself behind. He still needs to find certain normative principles by which to judge the passions that arise out of new situations. But he has to contend with this new moving ground.

And thus, the figure of the exile, merged with the person of Immanuel Kant, finally does become a paradigm for subjectivity (Pereda calls it “personhood”). But this new paradigm means that subjectivity is neither purely transcendental, for it is nomadic and by no means self-grounding or even grounded, nor merely empirical, for it is has ventured out of itself, exiled, beyond the boundaries of its inner experience. This personhood is somewhere in-between the transcendental and the empirical. The practical reasoning that ensues from this nomadic reflection—the lessons in exile—also offers a hybrid of Aristotelian (prudential), and a Kantian (principled), ethics.

By the end of the book it all becomes clearer; this was the plan all along; the “theoretical section” represents nothing short of a dialectical reversal; it is the telos of the book which informs its beginning, shedding light also on the significance of that old and illustrious tradition of lessons in exile, which dates as far back as history and the written word itself.

As is perhaps fitting, I will conclude my presentation by reading a couple of passages from the end of the book. First, “So, if I am not mistaken, the experiences of exile and of friendship” (friendship also plays an important role in Pereda's lessons in exile) “can also be used not to solve, but to dissolve the dilemma of personhood.”

Successfully realizing that solution would allow us to claim that a person is that animal who fashions itself out of its own actions and narrations using both biological and social concepts. Yet even while in the midst of acting or narrating, a person is capable of stepping back and reasoning in normative terms as if they were “any” human being. A person making the most out of the opportunity can negotiate and examine their desires, emotions, social systems of belonging and even their most intimate imaginings. (106, italics added)

And, yes, “The lesson is difficult, but [it is also] necessary.” For, Pereda writes (perhaps with an eye to Kant and the Kantian), if we do not give ourselves permission to destabilize our routines (including our theoretical routines), welcoming uncertainty and puzzlement so that later we can work our way through each situation towards a reflective solution, then we will almost certainly lose everything that matters most. (108)

And the very last paragraph of the book, in a brief concluding section bearing the understated title “Clarification,” reads, “One might object, considering my intentions to be too wildly ambitious;”

Not only did I attempt to outline certain approaches to these problems by not avoiding the relevance of all stories, metaphors, analogies, fragments of poems read as metatestimonies and even reflections on culture and people, I also did not shy away from a desire for generality. Of course, without subsequent reasoning—that is, without arguments that are expounded one premise after another—approaches of this type can often dissipate into thin air. But if we refuse to allow ourselves to participate in a little disorder from time to time, especially with certain investigations, we run the risk of being condemned to tasks that, out of routine, are guided by nothing more than centripetal movement, when not by repetitive and hollow paraphrase. (110)

And with these words, the journey ends.

AUTHOR BIOS

Roy Ben-Shai is assistant professor of philosophy at Sarah Lawrence College in New York and editor of The Politics of Nihilism: From the Nineteenth Century to Contemporary Israel. He has published essays on nineteenth- and twentieth-century European philosophy in Telos, The European Legacy, French and Francophone Philosophy, and others.

Noell Birondo is currently associate professor and chair of the philosophy department at Wichita State University. He completed his PhD at the University of Notre Dame and his BA at the University of California, Berkeley. He is coeditor (with S. Stewart Braun) of Virtue’s Reasons: New Essays on Virtue, Character, and Reasons.

Manuel Chávez Jr. is a philosophy lecturer at Monmouth University. He is a graduate of the Philosophy, Interpretation, and Culture Program at the State University of New York at Binghamton. His research focuses on the intersections of decolonial theory, social and political philosophy, Latinx and Latin American philosophy, and ethics.

Andrew C. Soto is a tenure-track philosophy instructor at Hartnell College. He received his PhD in philosophy from Texas A&M University. He specializes in Chicano/a philosophy, Mexican/Latin American philosophy, critical race theory, LatCrit, Africana philosophy, and philosophy of education.
In this newsletter we offer first an article by Andrea Sullivan-Clarke of Windsor University in Canada. In “Empowering Relations” she critically considers recent discussions of allyship and the limitations inherent in them, so as to build a less ethically naïve concept for allying work with indigenous peoples and communities.

We then offer three views of two recent books: Brian Burkhart’s *Indigenizing Philosophy through the Land: A Trickster Methodology for Decolonizing Environmental Ethics and Indigenous Futures* and Shay Welch’s *The Phenomenology of a Performance Knowledge System: Dancing with Native American Epistemology*. Both authors work to further the project of Native American philosophy, and their work intervenes into academic conversations in environmental philosophy, epistemology, philosophy of language, phenomenology, and performance studies. At the same time both books alert us to the always-already-also political framing of these sub-disciplinary and interdisciplinary boxes that can fail to catch the accomplishments of Native thinking.

We start with Joseph Len Miller’s review of Burkhart. Miller, of the University of Washington and Elon University, provides a chapter-by-chapter outline and scaffolding in terms of three key insights. Many readers should find this very useful for orienting readers to the specifics of Burkhart’s project. Likewise, the questions Miller poses open to further conversation about the impact of a localized concept of philosophy on questions not just for environmental thinking but also metaethics. After Miller’s orientation to Burkhart, Dennis McPherson and J. Douglas Rabb of Lakehead University orchestrate a dual look at some of the shared concerns animating Burkhart and Welch. Deploying some trickster methodology of their own, McPherson and Rabb point us to questions of both framing and reception. Finally, Lorraine Mayer of Brandon University provides some more exclusive focus on Welch’s key moves, following the implications of the insight that for some knowledge systems, truth is metaphorical and embodied.

We look forward to further considerations of each book and other new work in Native American, American Indian, and Indigenous philosophy, along with your scholarly articles, discussions of teaching, etc.

---

### Submission Guidelines and Information

We invite you to submit your work for consideration for publication in the *Newsletter on Native American and Indigenous Philosophy*. We welcome comments and responses to work published in this or past issues. We also welcome work that speaks to philosophical, professional and community concerns regarding Native American and indigenous philosophies and philosophers of all global indigenous nations. Editors do not limit the format of what can be submitted; we accept a range of submission formats including and not limited to papers, opinion editorials, transcribed dialogue interviews, book reviews, poetry, links to oral and video resources, cartoons, artwork, satire, parody, and other diverse formats.

In all cases, however, references should follow the *Chicago Manual of Style* and include endnotes rather than in-text citations. For further information, please see the Guidelines for Authors available on the APA website.

Please submit material electronically to Agnes Curry (acurry@usj.edu). For consideration for the Fall 2020 newsletter, please submit your work by June 15, 2020.

---

### Article

**Empowering Relations: An Indigenous Understanding of Allyship**

Andrea Sullivan-Clarke  
**University of Windsor**

Many non-Indigenous people naively assume that living in a post-conquest society entails living in a post-colonial one as well. Colonization, however, is an encompassing presence in the lives of the Indigenous people of North America. The effects of colonization linger, creating what Kyle Whyte describes as *Our Ancestors’ Dystopia Now*. Federal policies, motivated by a settler colonial agenda, have comprehensively damaged the myriad of relationships (cultural, epistemic, familial, etc.) originating within the ancestral lands of Indigenous people.

What most non-Indigenous people fail to realize is that “colonization is war.” As Whyte points out, the drive to
settle Indigenous land was “sustained, strategic, and militaristic.” The colonizing strategies invoked against the Indigenous people

Include[d] both war-like violence and the tactics for suppressing populations that are used alongside belligerence, from assimilative institutions (e.g., boarding schools) to containment practices (e.g., reservations) to the creation of dependency (e.g., commodity foods).

While some may claim that the policies of assimilation and the removal of Indigenous communities to reservations/reserves remain part of our unenlightened past, we need not look far to find contemporary acts of aggression and examples of Indigenous sovereignty being met with threats of violence from militarized police and private security forces. In the fall of 2016, the #NODAPL (No Dakota Access Pipeline) water protectors faced threats of violence from the local police and security forces, evoking images of previous conflicts, such as the Oka Crisis, Burnt Church Crisis, Idle No More, Ipperwash Crisis, Gustafsen Lake Standoff, and Wounded Knee (1973). In 2019, headlines in the media reported a raid on the unceded lands of the Wet’suwet’en people in British Columbia by the Royal Canadian Mounted Police. One disturbing feature—at least when one considers the Canadian government’s supposed commitment to Truth and Reconciliation—was the creation of “exclusion zones,” which prevented access to the media as RCMP officers dismantled two camps of Indigenous people protesting the placement of the LNG Coastal GasLink pipeline by Coastal GasLink, a subsidiary of TransCanada Corporation.

Police actions and militarized responses in Indian Country are so numerous that forward progress regarding relations with one’s colonizers seems virtually impossible. Other pernicious threats—such as revoking tribal recognition, contesting foster care placement laws, proposed reductions in treaty obligations, the failure to pursue cases of missing and murdered Indigenous women and girls, and the taking of Indigenous land—loom daily, profoundly affecting the lives of Indigenous people in North America. In order to eradicate the institutions of systemic oppression, contemporary Indigenous communities require long-term advocacy; I refer to this sustained support as allyship or being an ally. The concepts of ally and allyship, however, have come under scrutiny, and rightly so. Historically, the term has been misused. Given that the need for allies is great, how we define allyship for Indigenous Country is of critical import. In this paper, I discuss why being an active bystander is not sufficient for the needs in Indian Country and I present the failings of the current understandings of ally and allyship. As a solution, I offer a decolonized and indigenized concept suitable for the needs of the Indigenous people of North America.

An individual who wishes to support Indigenous people might act in a variety of ways. They might address an immediate need, and depending upon the circumstances, a response may be isolated or it may be of short duration. For example, an individual might provide a coordinated response to a threat of violence, such as inserting themselves between the private security forces and the water protectors at Standing Rock. I refer to this type of support as being an active bystander. Unlike a bystander, who witnesses a situation but does not act, an active bystander “witnesses a situation [and] takes steps to speak up or step in to keep a situation from escalating or to disrupt a problematic situation.”

Active bystanders are positively motivated to preventing or addressing bias, prejudice, and threats of violence. Of course, the actions of active bystanders need not necessarily be immediate; their response may take place after an incident, such as when someone listens to a victim or provides them with medical care. Although critical to the goals of social justice, the actions of an active bystander do not seem to have the sustained commitment I have in mind, nor do they have the investment (dedication for a particular individual or group). Generally speaking, active bystanders are not committed to act beyond the moment and once completed, it seems their work is done.

I take this to be the primary distinction between allies and active bystanders—we expect more from our allies. Active bystanders may not always be present and aware of oppressive situations—such as when the assaults to Indigenous sovereignty are kept private (through restricted media presence or behind doors of political committees/courtrooms). To overcome systemic oppression, Indigenous people would benefit from both active bystanders and allies.

The concepts of ally and allyship, however, are problematic. They suffer from overuse. A quick search at the library will reveal such titles as Earthworms and Their Allies, The Church’s Natural Allies, and The Working Class and Its Allies. More importantly, their meaning is often ambiguous. For example, the concepts conjure up an association with military alliances, such as the Allies in WWII or the various alliances between the Indigenous nations and early colonial powers, like England or France. This conceptualization often assumes a shared or common goal, such as the defeat of Nazism or mutual protection and trade. Once that goal is attained, however, the alliance has little value. In a letter to George Washington, Chief Cornplanter of the Seneca Nation expresses his confusion and feelings of betrayal resulting from the alliance his people had with England and subsequently the colonies.

Initially, the colonists demanded that the Seneca enter into an alliance with England. Chief Cornplanter notes, “When you kindled your thirteen fires separately [sic], the wise men that assembled at them told us you were all brothers, the children of one great Father who regarded also the red people as his children. They called us brothers and invited us to his protection.” Yet, after the colonists secured their independence from England, the support of the Seneca was no longer needed. Their lands, however, were. According to Chief Cornplanter, “[Street of Niagara] told us, that our Lands had been ceded by the King and that we must give them up.” Previously, the Seneca were entitled to rents for land used/settled. After the war, their “allies” forced them to surrender all their land without compensation.

An additional worry associated with allyship is its propensity for epistemic injustice. For example, allyship...
is ubiquitous within “queer politics and activism,” and as Rachel McKinnon points out, it poses harm to individuals from socially marginalized groups, such as individuals in the transgender community. McKinnon’s worry is that when individuals undergo training to be an ally, it becomes a part of their identity and when they “behave badly,” they do serious harm to already vulnerable individuals.

This type of allyship is often cultivated through some form of training, such as the Trevor Project or the Safe Zone Project. When allyship becomes part of an individual’s identity, it can whitewash their failings and even embolden their estimation of their performance. For example, McKinnon identifies a form of epistemic injustice known as gaslighting, which is associated with the use of an ally identity. Gaslighting is when an individual “doesn’t believe, or expresses doubt about, a speaker’s testimony,” which prompts the tester to question their experience of reality. Epistemic injustice affects members of marginalized communities whenever a hearer doubts the reliability of a speaker’s firsthand account of a harmful experience or unjust treatment based on some aspect of the speaker’s social identity. The identity of the “ally” can come into play as well.

If an “ally” fails someone, as in McKinnon’s example of a transgender person being mispronounced, the identity of being an ally can be used to discredit the victim’s account of the harmful experience, and, worse, it can even be used to deflect the hearer’s offense. For example, the hearer might respond with “Are you sure they did that? I know he/she is an ally” or “I don’t believe that since they are an ally.” According to McKinnon, the hearer has failed “to afford the first person (epistemic) authority of disadvantaged speakers its appropriate epistemic weight.” The hearer privileges their own knowledge and fails the speaker by not accepting their account as credible even though it is a report of the speaker’s firsthand experience. The set of worries associated with allyship as part of one’s social identity is only the tip of the iceberg.

Recently, Indigenous activists condemn the development of an “ally industrial complex” (AIC), calling into question the motives and sincerity of its members. Individuals who are a part of the AIC exploit and commodify allyship in order to preserve and enhance their professional goals, reputation, and personal agendas. Under the AIC, allyship is reduced to a purchasable identity, rendering it “ineffective and meaningless.” Allyship borne out of the AIC permits the preservation of romantic ideals, such as being a savior to the oppressed victim, and can result in the tokenizing of the oppressed (by challenging their autonomy). It encourages paternalism amongst allies: those with “official” training are granted a higher epistemic status than those who are actually undergoing the oppression. The existence of the AIC relies on oppression—something without which the “allies,” and their capitalist venture, would no longer be needed. Thus, it is reasonable to ask whether individuals are genuinely motivated to end oppression.

Another worry associated with the ability to purchase the label of ally is that there is little to ensure that those individuals seeking ally status are actually sincere or even capable to serve as an ally. Training programs vary and there is reason to worry about the ability of these programs to sufficiently address the diverse issues and people in Indian Country. There are over 1,200 federally recognized Indigenous communities in the United States and Canada, and although some problems may be shared amongst colonized people, they often face diverse challenges. Sometimes, interests clash amongst the Indigenous communities. Training may be too general and not sufficiently detailed to enable someone to be a competent ally to a specific Indigenous community.

When considering the training of allies, the variety of motivations individuals have for undertaking training must be subject to critical examination. One’s commitment to being an ally may be correlated with whether the individual independently sought out the training, the training was mandated by an employer, or there was an additional benefit/s associated with the training (such as whether it confers marketability in a job search, promotion, or provides additional compensation). In her critique of allies, McKinnon is quick to point out the swag—pins, buttons, and signage—often associated with being an ally. Thus, although there may not be a direct link between the motivation for undergoing training and an individual’s sincerity as an ally, it is reasonable to question whether an individual’s motivation is misplaced, as when an individual attends allyship training to be acknowledged amongst or even be thought morally superior to their peers. In short, completion of training does not guarantee a certain level of commitment.

Notably, training does not ensure cultural competency. Despite sincere intentions, an individual may fail, or they may acquire limited skills. In such cases, there will be times when harm is produced. The problem with allyship being construed as a social identity is that the harms may go unchecked or uncorrected. As was the case in McKinnon’s example, the bad behavior may be excused, or even deflected, given that the guilty individual is known to be an ally. Excusing the behavior given one’s social identity reduces the likelihood that the individual will learn to do otherwise. Instead, the bad behavior continues unchecked.

Not only does the concept of allyship suffer from commodification and abuse, upon closer analysis it also preserves colonial hierarchical structures. Some versions privilege the contributions made by individuals of the dominant social group. Take, for example, a version found in sociological research. It defines allies as “dominant group members who work to end prejudice in their personal and professional lives, and relinquish social privileges conferred by their group status through their support of nondominant groups.” This version renders the contributions from individuals from outside the dominant social group invisible.

Members from non-dominant social groups, such as Black Lives Matter (#BLM) and Veterans Stand for Standing Rock (VSSR), served as allies to the water protectors during the events of #NODAPL. Although many of their actions could be described as those of active bystanders—ministering
to logistical needs and acting as human shields between the Sioux people and the militarized security forces—many served as allies as well. For example, members of VSSR initiated relationships with the water protectors at Standing Rock by attending prayer ceremonies and participating in rituals. Aware of the U.S. government’s continued oppression of Native Americans, some members of VSSR sought to acknowledge past injustices and seek a way to move forward. The organizer of VSSR, Wesley Clark, Jr. noted “some veterans will take part in a prayer ceremony . . . during which they’ll apologize for historical detrimental conduct by the military toward Native Americans and ask for forgiveness.”

Members of #BLM in New York City served as active bystanders by “collect[ing] piles of donated items—including school supplies, tents, blankets, sleeping bags, medical supplies, and coats” to be delivered to Standing Rock. Yet, other members of #BLM recognized that Standing Rock is only one in a multitude of issues facing colonized people in the United States. Tara Houska, national campaign director of Honor the Earth, identified the shared motivations behind joining the protest at Standing Rock, stating, “We know that our communities are not only targeted by police and killed at a disparate rate by police officers, we also know that our communities are targeted at a disparate rate by these projects.” Historically colonized people have a shared starting point from which to cultivate their activism.

The relationship between Black and native activists has relatively recent historical roots. Wilma Mankiller, former chief of the Cherokee Nation of Oklahoma, described the shared problems of American Indian and Black families in the early 1970s. Indigenous families were relocated to the cities by the federal government and lived amongst Black families in the poorer neighborhoods, like Hunters Point, San Francisco. According to Mankiller, these groups shared similar experiences: poverty, high rents, unemployment, and broken families. The afflicted communities organized responses to these problems through militancy and grassroots projects, such as creating “breakfast programs and alternative schools” for all members of the community.

At Standing Rock, members of #BLM worked with the water protectors, sharing their prior experience of organization at the grassroot level as well as strategies for handling large protests, like that in Ferguson, Missouri. The #BLM NYC chapters continue to support Indigenous children through the collection of school supplies. As a result of this relationship, members of #BLM and the Sioux nations are able to communicate their needs and support each other.

The history of Indigenous communities in North America is replete with instances of (to co-opt the title of McKinnon’s article) allies behaving badly. The actions of missionaries, national governments, and others who believed that “civilizing” the Indian was in their best interest resulted in a myriad of policies such as removal, assimilation, and termination. Then and now, Indigenous people endure the negative effects of those policies, which were often framed in terms of improving their lives. So how do these cases of “allyship” differ from the examples of #BLM and VSSR? I propose that these putative allies acted without attending to the experiences, culture, and even the true needs of Indigenous people. Their efforts and the outcomes—whether intentional or not—ultimately preserved the social hierarchy and insulated them from critique. Their fundamental error as allies was failing to stand in the proper relation to the Indigenous people they claimed to serve.

The actions of the members of #BLM and VSSR are distinct from the above examples because their actions were those of decolonial allies—and it is their relationship with the Water Protectors that distinguishes them. In addition, the members of #BLM and VSSR were recognized by the Water Protectors as allies. Generally speaking, Indigenous people stand in relation to all things, as is evident in their creation stories and lived narratives. The knowledge of Indigenous people often stresses, or provides guidance for, the way for the people to be in the world. In “What Coyote and Thales Can Teach Us,” Cherokee philosopher Brian Burkhart presents such a story.

Using a story of Coyote, Burkhart relates to the reader the lesson that Coyote did not learn. Yet, it is a lesson that Indigenous allies must learn. Coyote, in his anger to seek revenge on some prairie dogs that poked fun at him, failed to realize that his wishing for enough rain to flood the prairie dog town would adversely affect him too. Simply put, Coyote did not recognize his connection to the world and, more importantly, his connection with the prairie dogs.

Relationships are a fundamental part of Indigenous philosophy; people stand in relation to everything in the universe—objects, other people, places, and spirits. As Burkhart describes, “meaning and value arise in the intersection between us and all that is around us.” In this way, our actions put meaning and value into the world. Thus, an Indigenous epistemology often includes a normative component—the universe is moral, and all of our relations provide knowledge as to how to live. As the actions of allies are critical to the success of the relation, I propose that an Indigenized conception of allyship should be understood as a relationship that promotes the well-being of those being served.

To be an ally to Indigenous people is to embark on a relationship with the people; it is a relationship with the “We.” It is not a commodity for purchase, although it does require investment. To stand in relation, individuals must know themselves while at the same time be willing to learn about the people with whom they enter into relations. Simply put, a decolonized understanding of ally requires an epistemic commitment. To counter the worry of paternalism, allies must respect the sovereignty of Indigenous people, which enables members of each group in the relationship to be epistemic equals. By doing so, allies can be guided and informed by Indigenous communities regarding future collaborations.

As we see from the examples at Standing Rock, the members of VSSR and #BLM were not necessarily from privileged groups and yet they served as allies. The social makeup of the membership of VSSR included a good number of military veterans from historically disadvantaged groups. Several of the members of VSSR were actually Indigenous people.
An allyship borne out of 45 As marginalized 46 Although traveling a great distance and joining 47 Personal posts to Facebook cited in 50 describe individuals who entered the 55 Examples 45 to cultivate their activism. For example, Kim Ortiz, organizer of #BLM reveal similar commitments of a decolonial ally. 395 Being a decolonized ally does not mean that there may not be shared common interests, but it allows for learning from diverse experience. For some, members of VSSR and the water protectors found common ground in the experience of post-traumatic stress disorder (PTSD) and sought healing together, as well as ways to help each other cope with the symptoms. 44 From this example, we see the respectful exchange of information and support. Although these interactions may appear to be similar to the WWII type of allyship, there are notable differences—such as the willingness to reveal one’s vulnerabilities.

The actions, rhetoric, and motivations of certain members of #BLM reveal similar commitments of a decolonial ally. Working together to overcome similar obstacles does not necessarily entail wishing well for the other, but it can bring marginalized/oppressed communities together to jointly improve their individual circumstances. Reporting on #BLM relations with Standing Rock, Ashoka Jegroo notes

Much like the indigenous tribes they’re supporting, Black Lives Matter activists from New York City see themselves engaged in a fight against state violence, oppression, and exploitation. For these activists, black and indigenous struggles are intimately tied together. 45

In this case, the activists from both sides—Indigenous and non-Indigenous—have a shared starting point from which to cultivate their activism. For example, Kim Ortiz, organizer of NYC Shut It Down, states, “we really need to stand in solidarity with the tribes out in Standing Rock because we know very well that all of our struggles are connected, and until we unite, we’re never going to win.” 46 As marginalized members of society, #BLM did not surrender their privilege or advantage to support those at Standing Rock. Rather, they identified as colonized people seeking to show solidarity with indigenous protestors, who also experience the negative effects of colonization. Both groups face police brutality, disproportional incarceration rates, and racism in society—the interactions amounted to a respectful exchange of knowledge to deal with a particular challenges, such as documenting inhumane treatment on social media to motivate a public response.

As mentioned previously, being an ally to Indigenous people requires learning their history, struggles, and needs. In addition, it requires knowing yourself. Whyte identifies a key issue that must be addressed in order to ensure the efficacy of non-native individuals to serve as allies: undermining the resilience of settler privilege. 47 Settler privilege, per Whyte,

means that some combination of one’s economic security, U.S. citizenship, sense of relationship to the land, mental and physical health, cultural integrity, family values, career aspirations, and spiritual lives are not possible—literally!—without the territorial dispossession of Indigenous peoples. 46

Settler privilege prevents non-native individuals from comprehending the dystopian and post-apocalyptic realities their colonial lifestyle has thrust upon contemporary Indigenous people. 48 Whyte suggests that in order to decolonize their thinking, potential allies must surrender their romanticism of native people along with the whitewashing of history that has erased the unique experience of native people from social memory. 49 Examples of so-called “allies”—those who failed to check their privilege—were present at Standing Rock. Accounts in the news media describe the “trashing [of] camps, mooching privilege—were present at Standing Rock. Accounts in the news media describe the “trashing [of] camps, mooching donations and treating the anti-pipeline demonstration like a Burning Man-style festival for hippies” by non-natives at #NODAPL. 51 Personal posts to Facebook cited in The Washington Times describe individuals who entered the camps, using food and resources “without any desire to participate in camp maintenance and without respect of tribal protocols.” 52 Primarily non-native “students, environmental activists and agitators,” the actions of these individuals stand in stark contrast to the actions of the VSSR and #BLM. 53 Although traveling a great distance and joining the protest for the environment, these individuals fail to understand their privilege and to recognize the actual living conditions of the protectors at Standing Rock. They did not stand in relation with the water protectors. (Given their trashing of the camps, it seems they failed to act as an ally to the environment as well.)

Potential allies must critically reflect on not only their motivations for being an ally, but they must be cognizant of their position in the social hierarchy. In this case, treating #NODAPL like a concert event, with its attendant cultural appropriations, compounds the damage done to Indigenous people. “Meaningful alliances aren’t imposed, they are consented upon.” 54 An allyship borne out of investing in a relationship will value Indigenous people as equals, and not perpetuate a romantic ideal that deflects critical discourse. 55

ACKNOWLEDGEMENT

I am indebted to the following groups and individuals who provided feedback and critique with earlier versions of this paper: The University of Windsor Philosophy Department, the University of Alberta Philosophy Department, the University of Washington’s Feminist Philosophy Reading Group, and the attendants at American Philosophical Association (APA) panel on Decolonizing and Indigenizing Philosophy, arranged by the APA Committee on Native American and Indigenous Philosophers.
REFERENCES


Biso, Derrick. Student Paper. MS, University of Windsor.


NOTES


4. Ibid., 3.

5. Derrick Biso, MS, unpublished manuscript, University of Windsor, 2019 unpublished manuscript.


7. Ibid.

8. I use the term "Indian Country" to denote not only the geographical spaces related to the American Indians and First Nations people, but also the political spaces and issues that affect them.

9. "Active Bystander Strategies," Safety Net Coalition, accessed January 2019, https://www.luc.edu/safetynet/resources/bystander/. At the University of Windsor, we refer to them as pro-social bystanders; they provide a positive social response in a given situation.


12. Ibid.

13. Ibid.

14. Coined by Miranda Fricker, epistemic injustice refers to the unfair treatment regarding knowledge attribution, particularly when communicating of knowledge. It often tracks gender, race, and other historically marginalized social categories.


16. Ibid.


19. Ibid.

20. Ibid.

21. Ibid., 170.


23. Ibid., 1–10.

24. Ibid., 2.

25. Ibid.

26. There is something about allyship as a social identity that misdirects the focus. A person’s behavior seems permitted given their status as an ally, but it really seems that how they
perform for an oppressed group should be the focus. McKinnon seems to be driving at the same point by supporting the use of active bystander and by her comments regarding the use of pins, badges, posters, etc. See McKinnon, “Allies Behaving Badly,” 167 & 174.


31. Ibid., 7.
32. Wilma Pearl Mankiller and Michael Wallis, Mankiller: A Chief and Her People, 154.
33. Ibid.
36. See McKinnon, “Allies Behaving Badly.”
38. Ibid.
39. Ibid., 17.
40. Ibid.
46. Ibid., 2.
48. Ibid.
49. It is important to note that settler privilege applies to all non-native people. I wish to thank an anonymous reviewer for this point.


52. Ibid.
53. Ibid.

BOOK REVIEWS

Indigenizing Philosophy Through the Land: A Trickster Methodology for Decolonizing Environmental Ethics and Indigenous Futures


Reviewed by Joseph Len Miller
UNIVERSITY OF WASHINGTON, SEATTLE, AND ELON UNIVERSITY

INTRODUCTION

As someone trained in contemporary, analytic, Western philosophy, Brian Burkhart’s Indigenizing Philosophy Through the Land was delightfully challenging. However, it was a challenge for which I’m thankful and one I think most analytic philosophers should welcome. It’s challenging not just because it’s a style of philosophy that is unfamiliar in the Western canon (i.e., Indigenous philosophy), but most of the challenge lies in the book’s vast scope. While Burkhart specifically names a number of historically important philosophers in his illustration of the traps and limits of Western philosophy, most of the challenge lies in the fact of how easily these criticisms apply to Western philosophy as a whole. While the latter half of the book focuses on decolonizing environmental ethics, the buildup to that discussion requires covering a good amount of Indigenous metaethics. Since metaethics itself spans Western topics including metaphysics, epistemology, philosophy of mind, philosophy of language, etc. as they relate to ethics, this book covers an enormous amount of philosophical territory.

In the book Burkhart argues that relationships to the land not only ground Indigenous ontology, epistemology, and morality, but that this relationship is also the source for decolonizing academic philosophy and Indigenous futures. By focusing on the land and our relationship to it, Burkhart provides an excellent illustration of Indigenous philosophy that serves as a contrast to Western philosophy. Given that Indigenous philosophy is grounded in the land, emphasis and focus on the land can provide us with a framework for decolonizing environmental ethics.
Given the depth of this book’s scope, and the incredibly rich illustrations offered therein, I would like to focus on three main contributions that I found to be particularly insightful. These are (1) the distinction between, and utilization of, delocalization and localization; (2) the introduction of Trickster Methodology; and (3) conceiving of ethics without appealing to value. Before discussing these topics, however, I would like to start with an unfortunately brief outline of the book.

**BRIEF OUTLINE**

Part 1 of the book is comprised of chapters 1–3, and focuses on the colonization of philosophy, whereas part 2, comprised of an interlude and chapters 4–6, is about decolonizing environmental thought and Indigenous futures.

With such little attention in academic philosophy being paid to both Indigenous philosophy and the relationship between people and the land, Burkhart begins in Chapter 1 by offering an explanation as to why Western philosophy has continually ignored these topics as a focus of philosophical investigation.

Chapter 2 focuses on decolonizing academia and indigenizing native studies. By clarifying the misunderstandings that early Indigenous scholars have concerning the works of Vine Deloria Jr., Burkhart discusses how some common practices and questions in Native studies are the result of coloniality. These include questions concerning authenticity, identity, identification, and sovereignty. Focusing on how epistemology, semantics, and ontology are rooted in the land (as Deloria Jr. does) would help to limit colonial influences on academia.

In Chapter 3, Burkhart discusses how ontology and epistemology can be understood as being rooted in the land by referencing Black Elk and the stories of Iktomi, the Spider Trickster. He then discusses how these approaches to ontology and epistemology work in practice. By contrasting Western ontology and epistemology with ontology and epistemology as rooted in the land, Burkhart continues to provide examples of the limits of Western philosophy (i.e., philosophy that doesn’t recognize the land) in the forms of opposing, mutually exclusive binaries (e.g., propositions are either true or false in Western philosophy).

Part 2 begins with an Interlude that focuses on Iktomi’s commentary on the history of Western environmental ethics. This commentary helps to explain why there’s a distinction in Western philosophy between intrinsic and instrumental value and helps to set up the discussion in Chapter 4 regarding how we can have a system of ethics that doesn’t rely on values. This is done by showing how we can avoid anthropocentrism, misanthropy, and the distinction between intrinsic and instrumental value.

The focus for Chapter 5 is about the metaphysics of morality rooted in the land. There’s discussion about avoiding primitivism and hypocrisy in Indigenous morality, and about the roles of unity and completeness when assessing moral theories. Working through these requires changing our understanding of how a moral theory is structured and functions instead of just offering an alternative moral theory.

Lastly, Chapter 6 is focused on how Indigenous morality can be seen as a kind of metaethical naturalism. There’s an extensive discussion about what “natural” means regarding morality, with an emphasis on the role of relationships and reciprocity in Indigenous morality. Grounding morality in an Indigenous understanding and approach to science, Burkhart contends that morality is a natural enterprise.

Overall, this is a well-structured, engaging, and illustrative book that engages and encourages a lot of further discussions. It’s a much-needed addition to Indigenous philosophy, metaethics, and environmental ethics. There’s no way my brief summary can do justice to the rich contents therein, but I hope that I can do enough to excite curiosity given the book’s scope and ambition.

**THREE INSIGHTS: LOCALITY, TRICKSTER METHODOLOGY, AND ETHICS WITHOUT VALUE**

I’d like to focus on three insights that appear throughout the book. Each of these insights is used to highlight some of the traps and limitations of Western philosophy. As such, I’d like to explain my understanding of these concepts, as well as provide some illustrations of how they work in the book.

### 1. **LOCALITY**

Throughout the book one of the key distinctions that Burkhart utilizes is that between locality and delocality. According to Burkhart:

> Locality is being-from-the-land and knowing-from-the-land. I use “locality” as a term of art in this book as a way to reference the manner in which being, meaning, and knowing are rooted in the land. Locality as a root of being is a part of each of us and speaks through us and from our historical and geographical place in the world regardless of how our identity is constructed in relation to culture or nation.¹

Delocality is the attempt to understand meaning, being, and knowledge as “floating free from the land.” Essentially, trying to understand these concepts as Western philosophy does involves delocality. Working with universal, abstract concepts require conceiving of subjects as having no relationship to land.

As Burkhart mentions in Chapter 1, since Western philosophy has focused on seeking abstract, universal concepts or principles, anything that is concrete and particular, like land, has been ignored. As part of the process of colonization, common concepts in Western philosophy are delocalized as they have come to be defined as being abstract and universal despite their being rooted in particular Western ideals. Given their roots in Western Europe, these concepts assume ideals and relations that are common to Western Europe. When applying these concepts to those not from Western Europe—i.e., Indigenous peoples—one easily falls into philosophical traps (e.g., thinking in terms of competing dualisms) and contradictions.
To illustrate this part of colonization, Burkhart mentions some key figures in the history of Western philosophy including Descartes, Kant, and Locke. For Descartes and Kant, key concepts in their arguments assume abstracted universality but identify universality in terms of Western concepts. Descartes’ in the cogito is identified as being a landless, cultureless subject. Anyone whose identity is bound with their land or culture—i.e., Indigenous peoples—is then not subsumed under Descartes’ (i.e., they’re not a subject). Similarly, Kant’s conception of rationality is defined in terms of “maturing” from an Indigenous (or “savage”) worldview/thought process to a Western worldview/thought process. In both cases, these concepts are defined in such a way so as to not apply to Indigenous peoples, thereby marking Indigenous peoples as non-human. As for Locke, although he proports to write a political treatise justifying property and land ownership, it’s really just an attempt to justify settler colonialism and taking Indigenous land without consent. Since, according to Locke, each person owns their body and the labor associated with it, anything mixed with their labor becomes their property. To justify the taking of Indigenous lands, Indigenous peoples were thought of as hunter gatherers (ignoring thousands of years of farming practices) as existing in the state of nature. This would make Indigenous land free for colonial taking and would morally justify the taking of Indigenous lands. These are just three examples that serve to highlight how universal concepts in Western philosophy are delocalized. They are defined in Western terms that rely on, and reinforce, harmful characterizations of Native Americans, as well as perpetuate ignoring Indigenous philosophy and the land.

One of the most hopeful or inspiring things the Burkhart mentions is that locality can never be entirely removed. In the case of colonization, delocalized concepts and practices are used to try and replace Indigenous concepts and practices, but, since Indigenous concepts and practices (i.e., localized concepts and practices) result from a relationship to the land, those concepts and practices will never completely be removed from the land. Indigenous philosophy can be ignored but not removed.

2. TRICKSTER METHODOLOGY

Another interesting introduction in this book is that of trickster methodology. Trickster methodology is an approach to showing how certain concepts or methodological approaches lead to philosophical traps. With reference to Indigenous tricksters, Burkhart uses stories about Iktomi, the Spider Trickster, to illustrate the limits and delocalization of Western philosophy. As Burkhart describes it:

Indigenous tricksters teach their relatives about the contours of locality and so help them put their feet back on the ground so that they do not continue to fall into the holes. Indigenous tricksters walk both sides of locality and delocality. Through humorous and creative failings Indigenous tricksters, like Coyote, Jisdu, Raven, and Iktomi, are able to deconstruct the epistemology and ontology of delocality from the inside. More simply, a trickster like Iktomi can lead you to spin and wrap yourself in the same webs that he spins around himself. As the Spider Trickster, however, he can do this in such a way that he will show you how you wrapped this web around yourself in the first place. This creates the space for you to be able to see how to get out of the web of your own making.\(^2\)

This methodology allows one to step back and see the paths of their reasoning. Seeing how they arrived at their current epistemic location, as well as seeing where their thought-process is likely to lead, can help us to understand the contradictions and philosophical problems that we face. Acknowledging and understanding the historical structuring of these problems can help us to avoid them by tracing our way back to the point where the problem was created.

Stories involving Iktomi appear throughout Burkhart’s book. One example that I’d like to discuss is Iktomi’s “story of the Western Thinker/trickster and the Western understanding of land and value.”\(^3\) In this story we hear the history of Western environmental ethics. Beginning with the Last Man thought experiment, we see why Western philosophy struggles to morally consider the land. Since there is no question in Western philosophy that humans have value, the last man clearly has moral standing. However, our intuition is that it would be wrong to destroy the earth. To explain this intuition, we make a distinction between intrinsic and instrumental value. Humans clearly have intrinsic value: they’re valuable in and of themselves (without reference to anything else). This being the case, since other things have value to humans, their value is merely instrumental. Wrongness is always made with reference to humans.

Iktomi wonders why anyone would continue to listen to the Western Thinker after displaying such trickster logic (thinking that presented oneself as both the justification and conclusion of an argument). People have been ignoring Iktomi since time immemorial for just this reason.\(^4\)

Eventually, the Western Thinker started to think that maybe the land itself has intrinsic value. Maybe non-human animals do too, maybe ecosystems as well. All of this is done, however, within the framework of the intrinsic and instrumental distinction. As such, explanations are offered as to what properties are required to have intrinsic value. According to Burkhart:

The Western Thinker wants to deny the intuition of the last man. He does not want to allow anything but instrumental value for the nonhuman world. The Western Thinker wants to limit the value of the nonhuman world to human-centered, instrumental value. The Western Thinker wants to think about the value of the nonhuman world within a pragmatic anthropocentrism, where the value of nature lies in its relation to the good life or human well-being.\(^5\)

This story is told with interspersing reactions from Iktomi suggesting and identifying various tricks in logic along the way. By the end,

Iktomi wonders whether the Western Thinker has just followed his own trickster logic back to
the same place where he started. Isn’t he now back standing next to Aristotle facing the original question: Why is value human-centered?

This story shows us why the distinction between intrinsic and instrumental value arises, how it’s utilized, and brings us back to the beginning so that we can wonder whether it’s a distinction we need.

3. ETHICS WITHOUT VALUE

One way of avoiding the distinction between intrinsic and instrumental value is to remove value as the basis of ethics. “A nonvalue theory of ethics I define as any theory where notions of value do not play a significant role in determining what makes one act right and another wrong or play a role in grounding this determination.”

By referencing Chief Seattle’s claim that everything is sacred, Burkhart goes on to explain how ethics without value is conceptually understood. Understanding the concept of “life” as “the capacity for kinship,” what counts as living expands drastically compared to the Western conception of life.

To say that life might be understood not as a property of things but as something that exists in the relationship between things belies two standard western philosophical assumptions. First, the space between things seems to be mostly thought of as a secondary aspect of our understanding because reality is primarily built up out of things in the Western philosophical narrative. The second philosophical assumption is seen in the standard move, which seems to arise, in a way, out of the first, of defining a notion such as “life,” “value,” or “right” by looking for a feature of a thing that does not merely indicate the presence of that thing but is what “life,” “value,” “right,” and so on truly are. To claim that life is not a property that anything has, because it is not a feature of things, runs directly against some of the most standard Western philosophical assumptions.

With this conception of life things like rocks, streams, ecosystems, nonhuman animals, etc. are considered alive. Being alive is a kind of relationship that things have to their surroundings. This helps to clarify and dispel misunderstandings/misreadings about the claim that everything is sacred. Everything, on this conception, has value since everything is sacred.

Everything that is alive is sacred and everything that is sacred is alive since both of these propositions reference being in kinship relations. Further, since every single thing, every grain of sand, is sacred, there are no levels of value. Everything has all the value there is. Everything is sacred.

This eliminates the distinction between intrinsic and instrumental value. Instead of identifying humans as the focal point of value, then expanding value to other things that are relevant to us, if everything is alive, everything has value. If everything has value, then there’s no need to distinguish between degrees of value (e.g., intrinsic v. instrumental value). In other words, if everything has value, then nothing does—or the need for the concept disappears entirely.

Instead of a foundational model of value where there are things that have primary value (intrinsic value) and other things that have secondary value (instrumental value), there is a web of value. Things have value only in terms of this web. This is a way of thinking about value in locality. Connectedness or continuity is what gives a thing value. The amount of value that a thing has is not determined by its place on the web. Yet if everything has exactly the same value—everything is sacred—value cannot form the basis of a moral theory.

Notice, this doesn’t suggest the elimination of value, but it points out that it’s not the basis of morality. With regards to environmental ethics, morality being relational eliminates the concerns about anthropocentrism as well as misanthropy, while also granting moral standing to nonhumans.

QUESTIONS

The main questions I have concern the audiences for this book. While I think this is something that most Western philosophers would do well to read, I wonder how Burkhart anticipates their response. Regarding trickster methodology, I’m concerned that showing someone the paths of their thinking (i.e., its history and structure) won’t be enough to convince them to adopt a view that doesn’t lead to the particular problems they face. Although, I’m not sure this is Burkhart’s aim. If it is, though, facing certain problems, however seemingly insurmountable, may be more epistemically comforting than trying to alter or surrender the more foundational beliefs on which those problems depend. From the viewpoint of Western philosophy, the answer to the question, “Why should we decolonize environmental ethics?” may have to do more than appeal to insurmountable problems. Another way of articulating this concern is asking Burkhart whether this is intended to convince Western philosophers to adopt the Indigenous views being presented? If not, I could see the book as an example for early career Indigenous scholars for how they can decolonize philosophy and environmental ethics.

Another question I have concerning the audience is what they should take away, personally, regarding epistemic locality. As I mentioned earlier, it’s easy to see how the concepts employed in the history of Western philosophy are delocalized, but where do we go from here? Localizing the concepts employed by western philosophers seems like it would restrict the scope of the arguments by those delocalizing those concepts. This may lead to epistemic humility, but it seems like delocalized concepts lead to a doomed philosophical enterprise. Is there any place in philosophy for universal concepts? Or is the search for these kinds of concepts inevitably doomed?

Lastly, as someone interested in metaethics, I’m interested in seeing how localized epistemology and ontology would
change discussions concerning contemporary positions in Western metaethics. It would be interesting to see how localizing ethical, epistemological, and ontological concepts would influence other metaethical views like expressivism (and other forms of non-cognitivism), non-naturalism, divine command theory, etc. Burkhart discusses naturalism, realism, and G. E. Moore’s open question argument (see Chapter 5), but this discussion excited me to the point of wondering about the implications for other views. I imagine some of these views would be untenable with localized concepts, which, if Burkhart is right about localizing philosophical concepts, so much the worse for those views. However, I’d be interested to hear more about how an Indigenous metaethics would be situated among more Western metaethical views.

CONCLUDING REMARK
As stated in the beginning, given the ambition and scope of this project, my review can’t do the book or Burkhart justice. The reason I focused on these three topics is because they are the basis for, or foundation of, many of the topics in the book. Understanding these ideas helps to not only understand a lot of the problems that Burkhart highlights, but they can also be used to identify problems beyond the explicit problems that Burkhart articulates. They also help to build a foundation for future Indigenous philosophy within academia. While this challenges the structure and many of the problems of Western philosophy, they are challenges that need to be addressed. Given the structure and scope of this book, I hope it’s something that becomes required reading in courses on Indigenous philosophy, metaethics, and environmental ethics.

NOTES
1. Burkhart, xiv.
2. Ibid. xxiii.
3. Ibid., 167.
4. Ibid., 169.
5. Ibid., 176.
6. Ibid.
7. Ibid., 204.
8. Ibid., 194.
9. Ibid., 195.
10. Ibid., 200.
11. Ibid., 203–204.


Shay Welch (Cham, Switzerland: Palgrave Macmillan, 2019).
Brian Burkhart (East Lansing, MI: Michigan University Press, 2019).

Reviewed by Dennis H. McPherson and J. Douglas Rabb
LAKEHEAD UNIVERSITY

In 2015 a comprehensive History of American Philosophy somewhat tentatively concluded: “It would not be misleading to say that American Indian philosophy is a ‘new’ field. Although Native American thinkers have been discussing and publishing on questions easily recognized as philosophical for centuries, few of these thinkers have carried out their discussion and writing in the context of the North American academy.” In 2019 the publication of two books on Indigenous philosophy, the one by Shay Welch the other by Brian Burkhan, may well change forever the way philosophy is done in the academy.

Burkhard focuses on Indigenizing philosophizing through epistemic locality.” He engages and critiques the delocalized epistemological structure of Western philosophy in both history and current practice and in the context of broader Indigenous philosophical practices.” He notes that “Indigenizing philosophy is then, in part, making it more active and dynamic. . . . In the context of Indigenous decolonial philosophizing, the context of this book, this philosophizing movement is a movement back to the land, regrounding our language, being, knowing, meaning, and so on back in the land.” Burkhart draws on the work of Vine Deloria Jr., whom he considered his Elder and mentor and who was in fact the external consultant/examiner on his Ph.D. dissertation. As Burk Hart explains, “One of the reasons that Deloria was so insistent that I complete a Ph.D. in philosophy, and in his constant encouragement of Native students to do the same was his sense of the need to underscore Native American philosophy with a deeper sense of its metaphysical and epistemological base. His belief was that Native American philosophy will always be seen as . . . representing the stage in human development in which superstition and ignorance reigned supreme . . . unless philosophers are given a bigger picture of the deeper unity and completeness of Indigenous philosophy through locality.” This is exactly what these studies by Burkhart and Welch are intended to provide.
Welch draws extensively on Burkhart’s book quoting passages at length. It is obvious that she is using a prerelease version of his book, which is listed in her “Bibliography” as “forthcoming.” Sure enough, when his book was published, among the “puffs” on the back cover was one by Shay Welch, associate professor of philosophy at Spelman College, USA. Her book is, of course, very much more than a commentary on Burkhart. For one thing, her principal thesis is that “dance is the most unmediated and clearest mode through which to generate and communicate knowledge and Truth from the perspective of Native American epistemology.” As Welch explains: “That my aim is to account for (an) epistemology within a Native American worldview substantiates why it is that I cite as extensively as I do.” She draws on our own study, Indian from the Inside, on a number of occasions. Here we must make a couple of minor corrections. She writes: “In the chapter ‘Dancing with Chaos: Phenomenology of a Vision Quest’, McPherson and Rabb (2011) interview a Blackfoot Mētis man named Douglas Cardinal to demonstrate how it is that supposed ‘mythical’ and ‘magical’ Native experience, typically discounted by Western culture and theory, actually shares common features with many other similar embodied phenomena.”

Epistemic locality is as much about ethics as it is about epistemology or ontology. As Welch explains, “Primarily, the purpose of pursuing knowledge is to help guide individuals along the right path. Relatedly, knowledge has as its end the nurturing of relationships between individuals and community members, including non-human persons and the environment.” Though neither Welch nor Burkhart use the exact expression “relatives not resources,” Burkhart can certainly be read as attempting to give philosophical depth to this cliché. He critiques moral philosophers from Aristotle to Peter Singer arguing, for example, that Singer’s defense of vegetarianism constitutes nothing less than a reductio ad absurdum of Utilitarianism, at least from an Indigenous perspective. Burkhart presents and defends a kind of relational ethics in the context of a process philosophy. He draws positively on an expanded version of Martin Buber’s I-Thou relationship to explain the Indigenous notion of kinship. Following Deloria, Burkhart references Alfred North Whitehead in explicating an Indigenous process philosophy. His arguments deserve close study. We can see his book being used in philosophy graduate seminars.

Welch, like Burkhart, is critical of the abstract nature and de-locality of much of Western philosophy; but also, like him, she draws positively on a number of Western philosophers to make her case. For example, she uses Gilbert Ryle’s famous distinction between “knowing how” and “knowing that” in her discussion of procedural versus propositional knowledge, the former being most compatible with dance. She calls upon philosopher of cognitive science Mark Johnson, making considerable use of his discussions of embodied cognitive science. The fact that we are embodied beings has important implications for philosophy, suggesting that much of our abstract thought is unconscious and metaphorical in nature. Welch explains:

For both Native Philosophy and embodied cognitive theory, meaning is grounded in corporeality. According to embodied cognitive theory, meaning is phenomenological and stems from embodiment in that it comes together for us through unconscious and mostly unaware bodily perceptions of space, movement, and environmental qualities. Movement, specifically, grounds our ongoing connection to and interaction with others and the world; it is what keeps us in touch with the world.
These notions of embodiment and movement prove fertile ground for Welch in her discussion of the role of dance in Indigenous epistemology and the Native narrative tradition. As she puts it: "Embodied cognitive theory provides crucial material for drawing the necessary connections to explicate in detail how dancing is a quintessential way of knowing in Native epistemology."17

Trickster stories are an integral part of most Native American narrative traditions. As Welch puts it,

Traditional Native American and Indigenous storytelling serves many functions within a community. They are tellings of a people’s origins, their relation to the land or water, the origin of the universe and stars and thunder beings, they relay cautionary tales about proper ethical behavior and the consequences of misdeeds and missteps. . . . Most importantly, they are metaphorical because most forms of education, formal and informal, are not comprised of directives that must be remembered and obeyed. The goal is for individuals, especially children, to eventually unravel the meaning of the stories in their own time and in their own way to guide them on their right path. . . . Burkhart analyses this marriage between metaphor and knowing through his development of the Trickster methodology. Indigenous tricksters, e.g. Coyote, Jidsu [sic. Jisdu? Jistu?], Iktomi, are our misbehaving or derpy relatives who, by leading—or rather, failing—by example, teach our kin about ‘the contours of locality’ so as to ‘deconstruct the epistemology and ontology of de-locality from the inside.’"18

Burkhart’s Trickster methodology is an integral part of his argument. He calls upon Iktomi, the Lakota Spider Trickster, though Coyote, Raven, and Jisdu, the Cherokee Rabbit Trickster, also make an appearance. We somewhat reluctantly invoke the Trickster from our local Ojibwa narratives, Nanabozho, Nanabijou, or just plain Nanabush. Our reluctance comes from the well-known fact that in our stories Nanabozho never knows when to shut up. Nanabush has some questions for Iktomi. But first, Nanabush wants to commiserate with Iktomi about the time Iktomi was “chased with a hot fire poker when he tried to point out that seemed to be obvious bias against Native epistemologies. The Western philosopher slammed his fists on the table and proclaimed, ‘we have a right to be biased. It’s just the truth.’ When Iktomi refused to submit, that’s when the hot fire poker came out.”19 Nanabush thinks that waving a hot poker is hardly the way to make a philosophical point. But s/he likes the indirect way Iktomi draws attention to the fact that some Western philosophers are actively hostile to Native philosophy. That said, Nanabush wants to ask, Why Indigenize philosophy? Burkhart rails against the attempt to Europeanize Indigenous land, where “the land itself is understood to be conquered; it is thought to become European land, as described by the names ‘New England,’ New York,’ ‘New Jersey,’ and so on.”20 Nanabush would like to add “Nova Scotia,” in recognition of what Mi’kmaq kin have had to put up with. “New Scotland,” but they wrote it in Latin just so we would not miss the European superiority. Nanabush thinks that is as silly as calling an Indigenous Law Degree a JID, a Juris Indigenarum Doctor. The Europeanization of Indigenous lands and concepts is ubiquitous. But Nanabush wonders if we should stoop to their level and go about Indigenizing things like philosophy. Does this not go counter to Indigenous pluralism and noninterference? The very last sentence in Welch’s book concludes with the insight that “there are unproblematic incommensurabilities between Native American philosophy and Western philosophy.”21 Everyone else talks about the “incommensurability problem.” That’s why Western philosophers cannot understand their Indigenous colleagues, or even believe that Native people would have a philosophy. Welch simply dissolves the incommensurability problem through Indigenous pluralism, polycentrism, kinesthetic knowledge, and so forth. There are unproblematic incommensurabilities! It makes Nanabush feel like dancing! Nanabush is tempted to claim that Burkhart is not really Indigenizing philosophy, he is rather engaged in Indigenous philosophizing, though Nanabush is willing to admit that that may well be the same thing.

Nanabush is pleased to see that Burkhart treats Elders like Black Elk, Lame Dear, and even Chief Seattle with philosophic respect and not ethnographic dismissal or ethnographic containment in which “Indigenous voices only have meaning as a form of ethnography, which would mean that the truth or value of their words is determined by the ethnographic authenticity of their words rather than truth or value in a broader sense.”22 But Nanabush wishes Shay would stop using big unpronounceable words like “perspicaciously.” Nanabush just likes a more colloquial writing style. That goes for the title of her book as well. Nanabush cannot believe that her publisher let her get away with the title that she used. Nanabush thinks her book deserves to be widely read but doubts that the title will contribute to that end. Nanabush also wonders why Brian insists on repeating an expression like “a world always already in motion” over and over and over again. Why can’t he just talk about “the world unfolding as it will” while recognizing the agency of locality? Nanabush also thinks that the word “locality,” itself, may be too abstract and general for the particularity of power and agency it is intended to capture. But Nanabush thinks s/he’s thinking too much, and so is off to the local bar for beer and bannock.

Readers need to take seriously, and experience for themselves, the hilarity of Iktomi and the Trickster methodology. We have had the privilege of team teaching Indigenous philosophizing by Welch and Burkhart would be well received there. Students would certainly get the Trickster methodology. We have experienced more laughter in these classes than in any others that we have taught, anywhere. More tears too, which is why these studies are both needed and welcome.

NOTES

1. Erin McKenna and Scott Pratt, American Philosophy from Wounded Knee to the Present, 295.
3. Ibid.
The Phenomenology of a Performative Knowledge System Dancing with Native American Epistemology

Shay Welch (Cham, Switzerland: Palgrave Macmillan, 2019).

Reviewed by Lorraine Mayer
BRANDON UNIVERSITY

Seldom, if ever, have I come across a book that stole my attention from beginning to end. Shay Welch's The Phenomenology of a Performative Knowledge System: Dancing with Native American Epistemology not only captured my attention but the way she weaves philosophical discourse with Native American discourse was fluid, comprehensive, and extremely educational. Shay takes our Native world and brings it straight to the heart of philosophy.

Few people, scholars included, recognized the significance of dance for Native Americans, reducing it to savage gyrations that required prohibiting banishment. The dance and the social, political, and spiritual reasons for conducting dances were meaningless to cultures without the sophistication for understanding dance at deeper philosophical levels. Welch tells us how dance, “whether as social or ritual performance, has always been a cornerstone of cultural practice and education and communal relationship strengthening.” Shay brings Native philosophy to light with her mixed-blood heritage, her storytelling ability, and her unique ability to integrate Western philosophical methods with Native American philosophical methods. She teaches the reader dance as a mode of Native American epistemology—an epistemology of dance she sees being in solidarity with others as an act of resistance, both in the academy and on the stage as an urban Native American sky dancer.

Welch explicates an epistemology, specifically in terms of its being a performative knowledge system, by moving through the works of scholars, Dennis McPherson, Doug Rabb, Brian Burkhart, Thomas Norton Smith, and many others. In true pluralist form, she also introduces Monica Mojica, George Lakoff, Mark Johnson, Rosy Simas, and Steven Winter, again, among many others.

She cleverly weaves through cognition and metaphors to emotion, intuition, and storytelling and, of course, dance to demonstrate knowledge as performative. One of her goals is to demonstrate how Truth can be constituted by the performance of an action, rather than by mere statements. Borrowing the earlier work on metaphors by George Lakoff and Mark Johnson, she is able to demonstrate action as knowledge, and the Native American epistemology of procedural knowledge as opposed to propositional knowledge. She skillfully weaves a Western analytic process with a Native American epistemology. For example, she is able to take Lakoff and Johnson’s idea of movement as grounding connections and interactions with others to her pivotal theory of performative knowledge of dance as procedural knowledge. Shay is unique in that she does

---

BIBLIOGRAPHY


American philosopher she also recognizes storytelling as valid, and a primary medium through which we come to know. Story, she argues, "operates through the oral tradition and relies on the sharing of pluralist individual experiences for knowledge construction." As part of her careful study and explication of story, she takes us to the world of Tricksters and stories as performativity. Shay herself has told a story of a Native American phenomenology as beautifully and enjoyably as any Native American storyteller. Yet her method of explication, while appealing, was still academically successful.

Shay Welch expertly constructs a Native American epistemology as she takes us through her journey to the phenomenology of a Performative Knowledge System while dancing with Native American Epistemology. Moreover, I thoroughly enjoyed her way of sashaying into chapters.

NOTES
1. Welch, 2.
2. Ibid, 64.
4. Ibid., 76.
5. Ibid.
As co-editors of the newsletter, it is with a deep sense of sadness that we announce the death of the Black philosopher Kenneth Allen Taylor (1954–2019), the Henry Waldrgrave Stuart Professor of Philosophy at Stanford University. From 2001 to 2009, Taylor was the chair of Stanford’s philosophy department. He was the first African American to chair a philosophy department at an Ivy League institution. While individuals such as Alain Locke, Eugene Holmes, Richard McKinney, Thomas Freeman, Jesse Taylor, Samuel W. Williams, and Francis A. Thomas served as chairs of philosophy departments, few had reached the heights that Taylor had. Since 2005, Taylor co-hosted a national syndicated radio show/podcast Philosophy Talk with John Perry. After serving as chair, Taylor became the director of the Symbolic Systems Program at Stanford, which blends computer science, psychology, linguistics, and philosophy, in a “study of the mind.” Taylor’s body of work, mainly in the fields of philosophy of language and philosophy of mind, include three books: Truth and Meaning: An Introduction to the Philosophy of Language; Reference and the Rational Mind; and Meaning Diminished: Toward Metaphysically Modest Semantics. While Taylor wrote relatively little that was immediately connected to the philosophy of the Black experience, we should not take it to mean that he didn’t have determinate views on the philosophy of the Black experience. One of the co-editors (Ferguson) remembers a spirited discussion I had with Taylor in April of 2019 while visiting at Stanford University. Taylor and I went back and forth for about forty minutes about whether race was metaphysically real. Taylor was a defender of Anthony Appiah’s Argument from Illusion or racial eliminativist argument; Appiah was coincidently at Stanford—only a few tables away—when we were engaged in our spirited conversation. After our discussion, I remember his wife, Claire Yoshida, saying to me, “I’m glad you are challenging him on his views about race and racism.” She further explained that their son, Kiyoshi Taylor, had also been raising questions about the validity and veracity of Ken’s views on race and racism. We hope in the future to dedicate an issue to Taylor’s philosophical legacy. If anyone is interested in contributing an article on Taylor’s philosophical contribution, please contact either myself or Dwayne Tunstall.

For this issue, we begin with our annual “Footnotes to History” spotlighting Joyce Mitchell Cook—the first African American woman professional philosopher—who passed away in 2014. Next, we have a contribution from Anwar Uhuru (Monmouth University). Uhuru’s essay-review of Anthony Neal’s 2019 book Howard Thurman’s Philosophical Mysticism: Love against Fragmentation explores Neal’s reading of Thurman’s philosophical mysticism and its place in African American philosophical history. Next, we have a philosophical dialogue between Michael L. Thomas and Alfred Frankowski on the relationship between lynching, extra-State violence and genocide. And our last contributor, Dr. Leonard Harris, is, of course, no stranger to us; he is a former editor and book review editor of our newsletter. Harris’s article, “Purdue University and President Mitch Daniels: Confession of a Rare Creature,” offers “philosophical musings” about what it means to be a Black philosopher in the academic world in the twenty-first century.

SUBMISSION GUIDELINES AND INFORMATION

The APA Newsletter on Philosophy and the Black Experience is published by the committee on the status of Black philosophers. Authors are encouraged to submit original articles and book reviews on any topic in philosophy that makes a contribution to philosophy and the black experience broadly construed. The editors welcome submissions written from any philosophical tradition, as long as they make a contribution to philosophy and the black experience broadly construed. The editors especially welcome submissions dealing with philosophical issues and problems in African American and Africana philosophy.

All article submissions should be between 10 and 20 pages (double spaced) in length, and book reviews should be between 5 and 7 pages (double spaced) in length. All submissions must follow the APA guidelines for gender-neutral language and The Chicago Manual of Style formatting. All submissions should be accompanied by a short biography of the author. Please send submissions electronically to apa.pbe.newsletter@gmail.com.

DEADLINES

Fall issues: May 1
Spring issues: December 1
Joyce Mitchell Cook (1933–2014)

Until the recent explosion of Black women professional philosophers, there were relatively few women in the field. There are the well-known names such as Anita Allen, Angela Davis, and Joy James. But before these women, there was Joyce Mitchell Cook. In 1965, after receiving her doctorate from Yale University, Cook became the first African American female professional philosopher. This was the same year that Malcolm X was assassinated at Manhattan’s Audubon Ballroom. Cook’s road was not an easy one—given the vampire of sexism at Historically Black Colleges and Universities (HBCUs) in addition to predominantly white universities and colleges.

Cook was born on October 28, 1933, in Sharon, Pennsylvania, and died on June 6, 2014. She was the ninth of twelve children of Reverend Isaac William Mitchell, Sr., and Mary Belle Christmas. According to a recently published essay by George Yancy, her father was affiliated with the Church of God (Anderson, Indiana)—a non-denominational sect of the “holiness Christian movement” with roots in Wesleyan pietism and also in the restorationist traditions.

After graduating from high school, Cook attended Bryn Mawr College. She graduated from Bryn Mawr in 1955 with a degree in philosophy. [It is worth noting that in 1940 the Asian-American activist and philosopher Grace Lee (Bogg) received her doctorate from Bryn Mawr.] Cook later received a bachelor’s and master’s degree from Oxford University in 1957 and 1961, respectively, with a double major in psychology and philosophy. While at Oxford, she became acquainted with some of the prominent analytical philosophers of her day. She studied with renowned philosopher Peter Strawson, attended lectures by John Austin, and was tutored by Mary Warnock and B. A. Farrell.

And while at Yale working on her doctorate, Cook studied with a range of well-known philosophers such as Wilfrid Sellars, F. S. C. Northrop, Rulon Wells, John Smith, and Paul Weiss.

Her dissertation focused on the American philosopher Stephen C. Pepper’s Theory of Value. Yancy explains: “It was Wilfrid Sellars who assisted Cook in formulating her dissertation prospectus before he left Yale to teach at the University of Pittsburgh, and then Paul Weiss stepped in as Cook’s dissertation director. In the end, however, it was under the directorship of philosopher Rulon Wells that Cook wrote and completed her dissertation.”

In 1966 she began teaching in the philosophy department at the “Capstone of Negro Education,” Howard University. The distinguished historian David Levering Lewis notes that Howard and Harvard were the most misogynistic universities in the country. Cook also taught for one year at Wellesley College (1961-1962) and two years at Connecticut College (1968-1970). In terms of other work experience, for about a year and a half, Cook worked for the State Department, where she worked as an analyst covering the affairs of various African countries. After leaving the State Department, she worked for the now defunct Office of Economic Opportunity. She left governmental work in 1966 to pursue a career in academia.

In the early 1970s, during the formative historical period when only a small number of Black philosophers were working on the conceptual parameters of what constitutes the field of Black philosophy, Cook was actively involved in a number of significant panels and conference discussions dedicated toward that end. This is paradoxical given Cook’s tendency towards what Yancy describes as a “race transcending cosmopolitan spirit.” In November of 1970, at a conference on philosophy and the Black experience, at the University of Illinois at Chicago Circle, Cook delivered a paper entitled “A Critique of Black Experience.” Later, in February of 1976, at the John A. Johnson Foundation in Madison Wisconsin, Cook, along with Black philosophers William R. Jones and Robert C. Williams, engaged in an extensive and insightful conversation about the nature of Black philosophy.

Unfortunately, Cook was denied tenure at Howard University. Under the dark cloud of racism and sexism, many of the early Black women philosophers face unbelievable hardships. As Anita Allen observes: “Angela Davis was unlawfully fired by the University of California, Los Angeles, in 1970 for her radical politics; Joyce Mitchell Cook was denied tenure at Howard University, a historically black institution. LaVerne Shelton, now an inspiring poet, was denied tenure at Rutgers University and left academia in 1996. Adrian Piper, who got her PhD the same year I got mine, was denied tenure at the University of Michigan, and after a series of other good jobs in philosophy, continued to an enormously successful international career as an artist.”

After being denied tenure, Cook worked for four years (1977–1981) at the White House under President Jimmy Carter.
Carter as a speech writer and correspondence editor.

George Yancy has noted: "Cook, while not participating in the sphere of professional philosophy, had never abandoned philosophy. Philosophy was her vocation, her calling. She was brilliant, a prolific reader and possessed a remarkable memory for details. Her knowledge of the history of western philosophy was impressive in its breadth and depth."

Ironically, she received the 2004 Alain Locke Excellence Award in Africana Philosophy presented at Howard University. I say ironically because Locke (who was also gay) was one of the most misogynistic men of his era.

Yancy reports that Cook was an avid pianist and loved the work of Frederic Chopin.

Her general areas of expertise were value theory and social and political philosophy. At the time of her death, she had been working on a manuscript on the concept of the Black experience.

When the Collegium of Black Women Philosophers was formed in 2007, Cook was honored for her contribution to the advancement of Black women in the discipline of philosophy. However, the recognition of Cook’s contribution has not resulted in any articles or books discussing her philosophical legacy or contribution. George Yancy, who is the executor over her books collection in addition to her published and unpublished papers, is the only person to have written about her.

SELECTED WORKS BY OR ABOUT JOYCE MITCHELL COOK


NOTES


ARTICLES

Textual Mysticism: Reading the Sublime in Philosophical Mysticism


Anwar Uhuru

MONMOUTH UNIVERSITY

I would like to begin this essay with a quote by philosopher Blanche Radford-Curry. She states, "philosophers should not acquire analytic or linguistic skills for their own sake, but as a means for solving existential problems to make life more worth living." Making life more worth living is what I believe is Anthony Neal’s intervention and reason for reading Howard Thurman. Neal’s book is not a mere attempt to intervene; it actually does the work. His book Howard Thurman’s Philosophical Mysticism: Love against Fragmentation goes beyond the intersections of time (the historical timeline of Thurman’s writings) and disciplinary boundaries. The goal of this essay is to critically read Anthony Neal’s work as a case for critically reading mysticism, and why it matters in Africana philosophy. My response to Neal’s reading of Thurman answers the question: Why should we read it? We should read it because, as I argue, not only is Neal arguing for the imperative to read philosophical mysticism and the contributions of Africana philosophers in the field of mysticism, but also how that work is a contemplation of what is the sublime. For Neal reads Thurman’s take on the sublime as not being just a mere moment of ecstasy, abundance, and bliss, but it is achieved when reading and engaging with texts. I would argue that in Africana philosophy and the way Neal reads Thurman is that the sublime is achieved when reading and engaging with texts. Hence those moments of engagement while reading, thinking, and responding are what I call textual mysticism.

Currently, scholars working in Afro-Diasporic thought and intellectual history are working through the ideas of recovery and re-memory. The work of recovery is what scholars in Africana Intellectual History are working through and at the same time abiding by and honoring the limitations of fragments and missing histories of people of African descent. Yet, working to rewrite histories and narratives that go against the dominant approach, which is
to paint people of African descent as victims, dispossessed, and melancholic, to portray them instead as survivors and thriving against those odds. I borrow Toni Morrison's term of re-memory, but I expound on that concept because to remember is a protest of whiteness that forces those who are racially and historically marginalized to forget. What Morrison and I would agree on is that memory is uncomfortable, disruptive, and forces those who benefit from "forgetting" to reckon with that memory and do the work to repair. In thinking about recovery and re-memory it is revisiting moments in history that have become canonical and hyper-formulaic. Consequently, that methodology of thinking has created a narrative of erasure and excuse. Erasure of the complexity of Black life before, during, and after the Third Wave of the Civil Rights Movement—the first being 1866, the second being 1875, the third 1964, and fourth 1968. Prior to this moment in political legislative victory for Black/African Americans, we only had the philosophical contribution of Alain Locke. He states,

The Negro has been more a formula than a human being—a something to be argued about, condemned or defended, to be “kept down,” or “in his place,” or “helped up,” to be worried with or worried over, harassed or patronized, a social bogey or a social burden. The thinking Negro even has been induced to share this same general attitude, to focus his attention on controversial issues, to see himself in the distorted perspective of a social problem.

Moving to a telos of whole being and moving beyond a problem for white folks is how and why the reading of Thurman's work is essential to contest white social constructions of a Black/Negro as “the problem.” Therefore, there is no Negro problem; there is only a problem with Negroes.

The telos of whole being and absolving one's self of being a problem is why it cannot go without being mentioned that Neal's work is an in-depth undertaking of doing the work of recovery and re-memory. It also creates a different framework for those who look at after-life studies to actively engage in countering the narrative of canonized figures within the movement. Because for Thurman, it is the philosophical teachings of his grandmother Nancy. Examples of the experiential as philosophy are texts that are often read as narratives such as the writings of Harriet Jacobs, Frederick Douglass, Maria Stewart, David Walker, Ottobah Cugoano, and Phillis Wheatley. Their experiences are philosophical questions that are first approached through contemplation and then further explored through the written form. The textual and inter/textual mystical experience that Thurman writes is why Neal reads the ontological epistemology of Blackness in Thurman's writings. Neal states, "The study of black(ness) as a type of consciousness or being in the world, not the color of a person's skin, but that which is rooted in the experience of blackness, particularly in the American context, is the fundamental aim of my investigations." As a philosopher, Neal reads Thurman from the perspective of the experiential instead of the social construction of Blackness, largely because Thurman bears witness to his grandmother’s accounts of having been enslaved. By doing so, Neal's reading complicates the argument that Blackness in America begins and ends with slavery. Yet, Neal sees Howard Thurman's grandmother Nancy as the absolute point of origin on how Thurman philosophizes but also proves that Blackness is before and after being enslaved. The discourse on theology and mysticism is removed from the period as being both a philosophical guide and strategy that informs thought and ideologies. Instead, discourse on theological and mystic approaches are relegated outside of the movement and are siloed to thinkers such as Kierkegaard, Well, and Stein.

How does mysticism from the perspective of a non-Catholic and non-white thinker inform philosophy but, more importantly, inform others? As Neal claims, "I have no intention of placing all of Thurman's events on a canvas, presenting but never firmly establishing justifications for texture, tone, and scheme. I am a philosopher." Instead of a biographical or a mere insertion of a non-white theologian in the canon of mystics, Neal's work is a hermeneutic undertaking in which he reads Thurman as a philosopher. Yet, he cannot ignore Thurman's Blackness. I argue that Neal purposefully uses the term "textual mysticism" as his inclusion of the hermeneutic and the mystical. Textual mysticism is the transformatory process in which a person engages with texts as a process to obtain enlightened bliss. Yet, for Thurman, his bliss is not just from nor is textual mysticism obtained from printed texts. It is the sonic and oral texts that Thurman and Black philosophers encounter. What better way to grapple with the experiential than by reading the mystical? How can our experiences challenge how we live in the world? By thinking through those experiences of living in the world, who better to philosophize than a Black American who, as Neal notes, is "a thinker who has some religious, some poetic, and some philosophical works".

One cannot begin to philosophize the Black experience without those three elements: the poetic, which I argue is the personal; the religious, which I argue is humanity's attempt at explaining the unexplainable; and philosophy as a set of ideas that are used to conceptualize the self and society. Neal further notes, when forming frameworks and contextualizations, I have ignored the words of those who could only speak about blackness as other, and not as self. What could I say about the experience of being white without first having someone else who experiences it.

I would add that this is from the vantage point or perspective of seeing but not knowing through experiential ties. Thinking about Blackness as the site and genesis of thought and not as the problem is already an anti-racist and Black-affirming methodology, partially because Neal doesn't philosophize from the perspective of contextualizing and correcting the Western Canon. However, in placing Blackness of embodiment and philosophizing, it is necessary to acknowledge the influence in its duality, which is the positive and negative impact that white philosophers have on African American philosophy and the way that African Americans philosophize. During his time at Columbia University, Thurman was able to experience the impact that John Dewey had on the department and the ways philosophers approached how they philosophized. Neal notes,
philosophers, particularly at Columbia University, were employing the techniques of philosophy to achieve solutions to the social and moral problems of their day. This philosophical aim was rooted in the propensity of the department at Columbia towards metaphysics. In this citadel, Thurman would be invested with the tools to struggle with and solve many of the issues which thwarted the very existence of the multitudes of blacks in the United States causing Thurman to refer to the bleakness of black existence as \textit{the Luminous Darkness}.\textsuperscript{17}

If Thurman considered himself a philosopher, what did he consider a philosopher to be? \textit{More so, what was Howard Thurman thinking about especially when he defines Blackness as}

\textit{“the Luminous Darkness”?}\textsuperscript{18} Instead of seeing Blackness as the absence of light, he sees Blackness as the absolute source of light. For Thurman and the way in which Neal reads his body of work, that Blackness is the sublime. His mysticism or textual intervention is an articulation of Blackness as the sublime.

Despite his personal experiences and bearing witness to or tertiarily absorbing those experiences of the social state of Blackness, Thurman does not see Blackness as bleak or a void but the point of illumination. That brings us to the question of what was Thurman thinking? Neal posits that Thurman was thinking of three things:

1. What did it really mean to be human?
2. Is there a best way for humans to live together?
3. Is there any significance in personal experience?

I will begin to answer the questions that Neal asks in his reading of Thurman. First, what did it really mean to be human? Neal himself begins with this statement:

On one level, it can be said that African American philosophy arises from the American moment and is rooted in the denial of the foundational ideal of freedom to enslaved African bodies, but on another level African American philosophy can be said to have arisen from the rejection of the lived experience created by slavery/oppression and an affirmation of the desired experience which extended from the minds of these black bodies as they ushered in their own Modern Era.

Despite being born in 1899, which is three years after the \textit{Plessy v. Ferguson} ruling (1896), Thurman’s personal and inherited legacy situates him in what Neal argues as going beyond philosophy as “white-only” and being the only “human perceptual framework(s).” Neal notes, African American philosophy rejected the white-only as human perceptual frameworks. The very relevance of African American philosophy is derived from the context of slavery. Therefore for Thurman as philosopher establishes not only his humanity but anyone who has autonomy and the ability to imagine.\textsuperscript{19}

Neal’s second question is, Is there a best way for humans to live together? To answer this, Neal reminds us that Thurman read Plato’s \textit{Republic} and referred to it throughout his life. Plato was required reading, and still is, I may add, if you are a student at Columbia University. The utopic structure of Plato’s \textit{Republic} is what plants the seed of thinking as to what are the best ways for humans to exist amongst each other. Second, in the structure of a utopia, Thurman would be able to contemplate the state of Blackness. Third, Neal also notes that the mysticism in Plotinus’s revival of Plato would influence how Thurman would read \textit{The Republic}. For example, Neal points out that Thurman feels that “Plotinus felt, taught, thought that all of life, all creation was in God.”\textsuperscript{20} Hence, Plotinus’s intervention of Plato’s work allows Thurman to see utopic formations of a society as being an ideal place for humanity. Just like Plato and Plotinus, Thurman used philosophy to work through his concerns about finding the best means for humans to live together. Is there any significance in personal experience? I would say in thinking through Thurman, absolutely. To not think through primary, secondary, or tertiary experience is to think from a limited and Western, \textit{Vitruvian} perspective. By thinking through the complex non-monolithic elements of experience is to unify the self, community, and the cosmological from a non-Western perspective. As for philosophy and what it means to philosophisize is to think through modes of being that do not center the white gaze (Vitruvianism) or seeing Blackness as monolithic. It would also mean seeing texts and those who respond to texts as having continued moments of the sublime.

\textbf{NOTES}
\begin{enumerate}
\item Blanche Radford-Curry, \textit{"The Reality of Black Philosophy," The Black Scholar 43, no. 4 (Winter 2013)}: 146–51.
\item Anthony Neal, Howard Thurman’s Philosophical Mysticism: Love against Fragmentation (Lanham, MD: Lexington Press, 2019).
\item I argue that mystics such as Soren Kierkegaard, Simone Weil, Edith Stein, etc. speak of moments of the sublime as a moment bigger than humanity’s ability to measure. However, in thinking through Africana discourse and mysticism, the ability to read, interpret, and write that moment is the sublime, especially when Blackness and being are under constant contestation in opposition and overcoming whiteness.
\item The African American Intellectual History Society (AAIHS) is a scholarly organization that aims to foster dialogue about research, writing, and teaching Black thought and culture through scholarship, interdisciplinarity, inclusiveness, public engagement, and media.
\item Nathalie Etoke, \textit{Melancholia Africana: The Indispensable Overcoming of the Black Condition} (Lanham, MD: Rowman and Littlefield, 2019).
\item https://history.house.gov/Exhibitions-and-Publications/BAIC/ Historical-Data/Constitutional-Amendments-and-Legislation/
\item \textit{Ibid.}, 6.
\item Soren Kierkegaard, \textit{Fear and Trembling} (Denmark, 1846).
\item Edith Stein, \textit{On the Problem of Empathy} (Germany, 1916); and \textit{Finite and Eternal Being: An Attempt at an Ascent to the Meaning
\end{enumerate}
To get us started, could you give us some background on how this new work bridges from your previous book project, *The Post-Racial Limits of Memorialization*?

---

**Al Frankowski:** I was working on a piece thinking about genocide and anti-Black violence and how framing things in terms of genocide gives us a different way to think about political violence. At the same time, I was also shifting to thinking about architecture and spatiality and histories of racism spaces of violence, and places of racism. So, moving away from memorials, which we tend to think of as being located in space, but remarkably as public sites of memory and history, that gave me a way of thinking about things temporally, but I was thinking about special location and exploring what could be thought of in terms of not just pasts or history, but spatial workings of histories of racial violence. So I’m looking at the history of anti-Black racial violence as a type of architecture, as a racialized terror issue, that occupies that sort of space. Another way of thinking about the background is that I was thinking about spectacle terror lynching as not only continuous with the history of slavery, but something discontinuous with it. One of the things that troubles me is that when people talk about anti-Black violence or white supremacy in the United States, they talk about a sort of continual trajectory from slavery to Jim Crow to Mass Incarceration. There’s something right about that. I think that’s the right context. But what seemed peculiar to me was that spectacle terror lynching marks a different type of racialization, a different type of public violence than slavery. So, I’m trying to be attentive to those differences.

**MT:** To build off of that, I’ve been thinking about the notion of violence at work in your previous book. There’s a conversation happening about sensibility in which it can reify itself. Reading this piece, it seems that what’s at stake is not a particular form of historical sensibility, the issue is types of violence and atmospheres of violence. Where are you on thinking of what those forms may look like? After the first book, we can think about historical sensibility, of how post-racial sensibilities disconnect a sense of the violence of the past. One way to combat that would be to cultivate a sense of historical violence. But you seem to be saying something different.

**AF:** The memorial work tries to locate a type of present violence that is historicized or a historical sense of violence that animates why it’s important to think about a historical sense of violence. What post-racial memory seems to do is to cut in the wrong directions. It seems to historicize a type of violence that is still present. It specifically doesn’t remember exactly what it claims to be remembering. So, its present is one that continually reenacts certain types of violence. What may have gotten lost there or what I hadn’t anticipated and am trying to be more attentive to is that the spaces which we occupy are not simply cleansed of violence just because they’re historical. Our present occupation of spaces, the present experience of spaces is highly politicized in direct relationship to the types of racial violence that don’t necessarily disappear just because racial violence has been in its past. Lynching spaces are particularly prone to this. If we think architecturally, there are spaces in the south and north that had large African-American communities. Once a lynching happens, everyone moves out, and you have the creation of a white space. That’s an architectural feature of that space. So, what you would have currently is people who have lived...
in the fact of white spaces, that have been architecturally constructed as white spaces. Those white spaces were only made possible through this severe anti-Black expression of political violence. So it’s no longer really a question of what do we remember and what do we forget, it’s now a question of do how we live in the spaces that we occupy.

MT: You begin your article with a discussion of the narrator of Ellison’s “Party down in the Garden.” Particularly, in the narrator’s experience, there’s a tension between the public and the private dimensions of those spaces, which opens us to thinking of the public and private dimensions of people. In this case, it seems the spectator is an outsider to lynching but an insider to white supremacy, meaning they are foreign to our direct experience of violence, but at home with anti-Black violence in the U.S. Is this how you’re interpreting that experience?

AF: What’s complex about the Ellison character is that he’s clearly white and being invited into this event, where he becomes an insider. What strikes him as odd isn’t the absence of Black people or what’s happening to this particular person. There is a weird transformation of his own sensibility that is laying claim to him. So, he is not an outsider to white supremacy, he’s an outsider to a very particular aestheticization that is being made public. I quote one of the characters who says, “there are no Christians here, Only Americans,” and at this point there is a dropping of the façade. “Let’s not fool ourselves anymore, we’re all here. This is what we’re here for. We’re only American.” There’s a sense where that’s what’s being made public. That’s what makes [the narrator] sick. He’s literally being turned inside out. But, it’s not a sickness to the violence being cast on the Black body.

So there’s something interesting about the revealing of the political through public space. It’s a revealing that can only take place in that particular square. So we lose something when we talk about the generality of each lynching. Yes, they’re really highly scripted. Yes, they’re very routine. But each one is attached to a very particular place.

MT: So, you’re emphasizing that this is a place of public assembly. We tend to think of spaces as abstract terms, as vacuous. What role does the public play in constituting the space? How do we move from space as the abstract place where things happen and that space where this public gathers to mark it for itself?

AF: We think of spaces as being abstract so that we don’t think of them as places where the political is operative. In her work on assembly, Judith Butler is emphatic about the fact that the ability to publicly assemble is the work of the political. There’s no reason to believe that the places that we assemble and the reasons why we assemble have to be good or bad. But, acknowledging that spaces actually don’t exist unless they are potentially sites of the political is important for understanding why something like a protest or why police presence is important. Think of what’s going on with Walmart or college campuses where you are confronted with the fact that these places of assembly are places of the work of the political. What goes on there is also a type of working out of the political.

MT: What does sovereignty, a concept central to your article, mean within that political space of assembly. What does it mean for that to be a white supremacist space?

AF: One thing I find problematic when looking up terror spectacle lynching is that public spaces seem to be redefining themselves politically not simply as white supremacist spaces. They are redefining themselves as they are redefining what is possible through anti-Black forms of assembly. If we go back to the spectator in Ellison’s work, the spectator is becoming something. It’s not that he’s radically changing his whiteness, he’s becoming something different. It doesn’t mean that after seeing the lynching he only understands himself as white supremacist. The lynching becomes a sort of boundary. It’s always an extreme. The fact of its possibility and that it operates in a space as something the state can’t legislate against. It means that the public performance is also a reconstruction of a sovereignty that isn’t finished or terminated in the lynching. It becomes real in the displacement of the people. It becomes real in the people who occupy those spaces normatively the next day, weeks later, or months later. It becomes real in their children. It becomes real in the way in which that public space is configured around that event. Through erasure, through downplaying it, through putting up markers. It becomes real after that. It becomes real for us in the way that it’s only a dream or a promise for the folks who attend the lynching. This is how Arendt talks about sovereignty and the general will.

MT: In the article, you trace an evolution of the idea of sovereignty from Rousseau, through Schmidt, to Arendt. What happens in that movement between the three?

AF: For Rousseau, what is sovereign has to do with the question of the General Will as a type of agreement. He has a bare bones notion of the limit of the political. Schmitt rightly capitalizes on that particular point, that the General Will is needed by the state to understand itself politically. Without an expression of absolute sovereignty, don’t have state power, you have potential state power. Just like, without the absolute expression of God’s will breaking all of God’s own rules, you don’t have divine power of of any sort, you just have potential. The state has to be able to dispose of its people at will, for any reason or no reason at all in order for it to embody sovereignty.

Schmidt means this literally, that it has to be able to send people off to die. Some people have to die in a war, or it has to dispose of them as an expression of its sovereignty. That has to be a real possibility. In religious terms, it has to be possible for your God to be able to tear you down. Otherwise, it’s not really a God. If God can only work to do the good, then I have no fear of my God, and I can’t really stand in awe of its power. Schmidt is working on something similar there.

Arendt takes up the same notion of sovereignty between the two because she’s pointing out that it’s in the public gathering of the General Will that any sovereign action is never completed in the gathering. It’s always a promise at the same time. This doesn’t have to be for bad or for good. Whatever relates us, what organizes that public affect,
is not just an act or an event, it’s also a promise that can be completed later on. That’s true of those revolutionary politics which we are much more comfortable. It’s also true of the white supremacy of spectacle terror lynching. It’s in the destruction of the Black body that we see the architecture of a future form of white supremacy that’s being made public.

MT: This helps make sense of the example of Brandon McClendon, who killed in Paris, Texas after being dragged to death behind a car. You seem to claim that this murder is a lynching despite lack of a gathering or a public being present. There is still an extent to which lynching is more than the noose. What is specific about it as a form of anti-Black violence?

AF: It’s a form of anti-Black violence that signals a type of gathering that it doesn’t have to have. It doesn’t have to follow the same patterns. So, when people say we no longer lynch, well, yes and no. It depends. Maybe that’s not the most important question. The more important question we need to ask is, “What form of sovereignty are we continually reinscribing as a public?”

MT: It seems that what makes lynching look exceptional to many people is that it is a ritualized form of violence. How is that ritual aspect still present?

AF: We have to be careful about what’s embedded in our everyday rituals. If the public gathering is not necessarily good nor bad beforehand, if we pay more attention to how our habits are relating to types of public reenactments of anti-Black violence. If those haven’t changed, then there’s no reason to think of them differently. Take the example of twitter activism where “dragging” is metaphorically used for publicly shaming users who post racist, homophobic, or other forms of violent tweets. The stakes aren’t if it’s effective, the importance is what’s being reinscribed there. A lot of it has to do with signaling a public and getting them enthusiastic around a particular type of violence or destruction.

MT: When we discuss the ritual aspect, there’s a habitual aspect, where our everyday practices reinforce what’s been promised to us in those spaces. There’s also a ritual aspect in the sense used by Durkheim and Bataille of the affect that holds the space together which holds that space together. Is that right?

AF: Yes. Plus, if it’s a promise, then we don’t really know how we’re completing or moving against it. We shouldn’t be so confident that we know our way of participating of political action is moving against that promise or working towards fulfilling it. When we think of our agency, various technologies give us a type of agency that we didn’t have before. How much do we know about that agency? How inquisitive are we about those particular methods? Those are all things that we should be cautiously questioning.

What we get out of looking at terror spectacle lynching as a type of promise of a particular formation of public sovereignty as opposed to it being white supremacist or specifically anti-Black is that we have to think more carefully about practices of genocide. This is why framing it as genocide is necessary. It may be true that all forms of anti-Black violence ought to be framed as genocidal practices. But I’m not sure that they all establish the same type of genocide.

MT: Are we thinking about genocide in its codified definition, or is there something more there?

AF: There’s something more there. I don’t want to discount what’s been said of genocide’s legal definitions and the cautions about expanding the term beyond its general meaning. If we’re being careful about how we’re defining something, we have to notice that there are clear ways that genocide is always referring to cultural practices. In Limpkin, there is a way that he could not think about genocide without thinking about colonialism and without thinking about how its affected interactions between the colonizer and colonized and what cultural effects it’s had over time. He’s right up to the point of talking about it as being embodied in one’s affective relationship with forms of political violence, whether or not that entity has any real investment in destroying people or not. This is particularly in his references to Native Americans, but you also see it when he talks of Germany’s relationship to Africa.

I’m perfectly fine with saying there is genocide and there are genocidal patterns of violence. I’m also fine with saying you can’t have an anti-Black society and think that it’s just about discrimination of individual groups without also thinking about how that type of discrimination has coded in it the language of genocide.

MT: Could you say more about the connection between anti-Black violence, discrimination, and the language of genocide? How is the language of genocide encoded in that language of discrimination?

AF: If we think about colonialism as not just an abstract set of political relations, but one that was enabled by slavery, and not just slavery abstractly, but African slavery, we have to think about it as a practice that brings in the world anti-Black practices. Those practices aren’t just making possible a world that disadvantages Black people, it’s a world that is anti-Black in its most essential sense, which is the elimination of Black people. It’s a world in which Black people aren’t just a second tier, or third tier, or an absolute bottom, as Derrick Bell will say. It is a world that is antagonistic to Blackness. It has a tense relationship to slavery because you need the slave, but the slave is completely disposable. This is why I would say that there’s a discontinuity. Because that tension is one that is particular to slavery that’s lost in Jim Crow. You don’t distinctly need the labor of the slave, what’s needed in Jim Crow is the disposability of the Black body. Sovereignty is only made real through that particular form of anti-Blackness. So we can say slavery is genocide, but then we have to then acknowledge that Jim Crow is genocidal in a different way. That may tell us a lot more currently about the status of these various types of violence and their relationship to the political that we couldn’t see as genocidal without that difference.
MT: Recent footage emerged (on 9/20/19) of a Black woman being gang beaten by four men at a Pittsburgh gas station. Apparently, it started when she knocked over a display of chips after the owner refused to transfer her payment from a broken pump to a working one. The owner was not white, but it raised the question for me of what could animate that sense of violence as a legitimate form of recourse, particularly against the body of a woman. This seems to help explain that event. You've already presumed a certain sovereignty over the situation in a social order that makes this sovereignty possible.

AF: These are all spatial relations, so it's about property. She knocked over the chips. She's occupying space in a way that makes an immanent claim. The history is almost irrelevant, because they're occupying the space of rights, of that political sovereignty. You have to believe that there's a type of relation that goes into that. It dispenses with the idea that they're white supremacist, or they hate Black people, or they have an encoded doctrine working underneath everything. It's a form of relation that is present when they're able to exercise that sense of the public. Maybe on it's the small scale, it's when you're the owner of a convenience store, but it's also present at Trump rallies. This is why he's accurately describing them as a movement of love. It's a movement of White love, predicated on anti-Blackness.

MT: I explained to my wife that when I say, "it hasn't felt this way for a while" during the Trump presidency, I'm referring to the fact that there's an affect that's present which had dampened since I was young and in Louisiana. As soon as I stopped moving and changing jobs to settle in a place, this rhetoric up, and I feel it again. There's a present immanent threat that I'd thought I'd avoided.

To conclude, why is it important to use the language of genocide to make these connections?

AF: The term genocide neatly reframes the way we think about these things so that we're not thinking about a particular type of violence that is associated with the United States or with a particular racial relationship that lends itself to the language of sovereignty. You can't really talk about lynching without also thinking about a practice deeply wedded to colonialism, not an abstract colonialism, but anti-Black colonialism, and not just American form, which would particularize it. There are the acts that form our colonialism globally. They are also commensurate with the sensibility of that colonialism.

A world in which lynching practices are possible as they are in our world is a world in which they make sense. It's the sense of framing them as genocide that I'm trying to antagonize. We're shocked by lynchings because they're extreme. That's such a mediated relationship to something that allows us to be shocked by lynching, and that's how we make sense of it. When we start to frame it as genocide, that's what antagonizes the sense that we're making of them. That's the difference of taking an aesthetic approach to the problem rather than the rational. We're looking at what's going on at the level of the sensate. It cuts across that. It's not that people who are talking about lynching as genocide are stretching the definition, we're hitting at the sensual level, at what doesn't make sense to them.

Purdue University and President Mitch Daniels: Confession of a Rare Creature

Leonard Harris
Purdue University

On July 5, 1852, the abolitionist Frederick Douglass wondered what a July 4 celebration of American independence and democracy meant to America’s enslaved or freed slave population. Facing a new year, I wonder what the past 150 years of my university’s Great Leaps and contributions to intellectual life means to me?

I can never hope to be a named endowed chair at my 150-year-old Research One University. I am a scholar. I am not an administrator nor a grantsman with a track record of receiving major awards from the government or private foundations. There is a trait that I have which guarantees I will not be seen as a rare creature, that is, an African American scholar. I am an African American scholar. This trait is undeniable. I have been graced with a lifetime achievement award for scholarship and academic leadership and awards for promoting American philosophy, Africana philosophy, and numerous recognitions bespeaking my contribution to litera humanitas. Seven books to my name. I may not be qualified to be a named endowed chair, but I would be like to believe it is possible based on my merits. Others, whether of African, Japanese, Indian, Chinese, or of white heritage, have been named and endowed chairs for scholarship. What once were termed “honorary white” people in apartheid South Africa, (Japanese, Koreans, Taiwanese, and Chinese) is a designation that eludes me: I am invisible in a world where the exotic ‘other’ can be granted status but the native, in this case, African American, must be a singular racial kind and fit the appropriate stereotypes. It is only African Americans at my university that must be an administrator, in addition to merit-based scholarship, that have been awarded a named or endowed position. Wealthy African American slave owners were occasionally treated as honorary whites in the ante-bellum south (Negro, but not really, a sort of honorary white for convenience); white women slave owners were occasionally as vicious, ruthless and powerful as men and accorded special status at slave auctions or in court (women but not really, a sort of honorary male for convenience). If I were the absolute best in my field I might have a chance. I cannot just be great, like my white and honorary white colleague, I must be really rare. Sometimes I think I am invisible because I am not sufficiently productive. If I suffer from a multitude of micro-stresses that draw my mind away from scholarship and add to my likelihood of early onset high blood pressure or ulcers, none of this can be neatly establsh and can never be compensated. My shortfall in productivity gains me nothing but additional invisibility. Even if I am not a rare creature, I am not sufficiently unique. Despite the fact that there are, by any count, very few tenured African American Professors in Philosophy (4.3
percent in 2016), no matter how rare or accomplished, my kind in any field have only been awarded at my university a named or endowed chair if they have two jobs: in addition to any scholastic or academic merit, they must also be a senior administrator. For all the money, promotion of the university as a leading institution of higher learning and dedication to enhancing its image by the alumni of my university, I, nor anyone of my kind, in 150 years, have ever received a dime or the benefits of honors for their scholastic endeavors par excellence, simpliciter.

I teach the same number and types of courses as first year assistant professors in my department. I have access to the same support for research, travel or contribution to the university as other teachers. I do not get to be considered innocent and I do not get to be considered an exemplary teacher if I use degrading, demeaning or insulting stereotypes to characterize my students in my class when I use examples of ethical problems. I get reprimanded and I am not likely to get a substantive salary increase despite whatever other accomplishments I may achieve.

Another dollar to the university’s division responsible for promoting diversity will never fix the problem. Only senior administrators, professors, named and endowed chairs decide who will be named and endowed chairs. The white and honorary white people decide. Their decisions, for 150 years, have been pretty predictable.

I teach a course in philosophic anthropology that features concepts of racism by African American and Third World philosophers. The authors proffer different and competing concepts of racism. Racism is defined by J. Garcia and Anthony Appiah, for example, as malicious intentions, ill-wills toward the well-being of others or irrational reasoning. Angela Davis and Tommy Curry define racism by social structures and ill-wills or the interplay of class and prejudice. I define racism as a form of necro-being. Racism, on my account, is a way of killing and forcing persons to live egregiously facing death, preventing persons from being born and stealing assets needed to sustain healthy living. Intentions are nearly completely irrelevant on this definition. All the definitions of racism, however, consider the authority falters, it is not like the failure of authorities that are responsible for leading all of us all. When the authority falters, it is not like the failure of persons without public personas and responsibility. Their failure is especially egregious. To whom much is given, much is expected.

The slave and freed slaves in 1852 could only hope that the future would be radically different than the past.

REFERENCES

I wanted my picture to be on the front cover of my book, A Philosophy Born of Struggle: Leonard Harris Reader (Bloomsbury Publishing Company, 2020). The publishing company refused. After three months of arguing, and their willingness to compromise and put my picture on the back of the book, I relented. One reason they had for not wanting my picture on the cover was compelling: potential buyers would assume that a Black face meant the contents were for, about, and ultimately only created by an essential racial type, a Black being. If the cover had a white person’s face, buyers would assume that the contents had a general application and an impartial source. The marketing department had, in fact, done their homework. Stereotypes cause financial, personal and social harm. Stereotypes are not innocent images. Their authors are not innocent.

During a conversation with African American student the President of my university, Purdue, Mitch Daniels, on November 20, talked about recruiting minority faculty and told students that “At the end of this week, I’ll be recruiting one of the rarest creatures in America—a leading, I mean a really leading, African-American scholar.” My president has since apologized (December 4): “I retract and apologize for a figure of speech I used in a recent impromptu dialogue with students,” Daniels wrote in a letter sent to the NAACP, Purdue’s Black Caucus of Faculty and Staff, Latino Faculty/Staff Association, the Black Student Union and the Latino Student Union. “My reference was in praise of a specific individual and the unique and exciting possibility of bringing that particular individual to Purdue,” Daniels wrote. “The word in question was chosen and imprecise and, in retrospect, too capable of being misunderstood. I accept accountability for the poor judgment involved.”

I have every reason to believe his comment was made without malicious intentions and that his apology was sincere. So what if the president’s intentions were honorable and apology sincere? Unintentional consequences are consequences, nonetheless. The spread of stereotypes often influences loss of hope, depression and thereby increased blood pressure and onslaught of ulcers.

I teach a course in philosophic anthropology that features concepts of racism by African American and Third World philosophers. The authors proffer different and competing concepts of racism. Racism is defined by J. Garcia and Anthony Appiah, for example, as malicious intentions, ill-wills toward the well-being of others or irrational reasoning. Angela Davis and Tommy Curry define racism by social structures and ill-wills or the interplay of class and prejudice. I define racism as a form of necro-being. Racism, on my account, is a way of killing and forcing persons to live egregiously facing death, preventing persons from being born and stealing assets needed to sustain healthy living. Intentions are nearly completely irrelevant on this definition. All the definitions of racism, however, consider the authority falters, it is not like the failure of persons without public personas and responsibility. Their failure is especially egregious. To whom much is given, much is expected.

The slave and freed slaves in 1852 could only hope that the future would be radically different than the past.
PREFACE

Peter Boltuc
UNIVERSITY OF ILLINOIS–SPRINGFIELD
THE WARSAW SCHOOL OF ECONOMICS

The aim of this last issue of the Newsletter of the APA Committee on Philosophy and Computers is to feature some of the articles we were able to publish over the years. We published original papers by the late Lynne Rudder (2008), Jaakko Hintikka (2011, 2013) and John Pollock (2010) as well as by the eminent philosophers Gilbert Harman (2007), James Moor (2007), Luciano Floridi (multiple), Gregory Chaitin (2009), Dominic McIver Lopes (2009), Terry Horgan (2013) and many others. We are using the archives to highlight some of the high points. Detailed discussion can be found in the note from the editor just below the main block of papers.

We end with a few pointers towards the next steps that the community of philosophy and computing may be taking some time soon. This includes creation of an Association on Philosophy and Computers, affiliated with the APA and potentially other initiatives.

FROM THE ARCHIVES:
AI ONTOLOGY AND CONSCIOUSNESS

The Shrinking Difference Between Artifacts and Natural Objects*

Lynne Rudder Baker
UNIVERSITY OF MASSACHUSETTS AMHERST


Artifacts are objects intentionally made to serve a given purpose; natural objects come into being without human intervention. I shall argue that this difference does not signal any ontological deficiency in artifacts qua artifacts. After sketching my view of artifacts as ordinary objects, I’ll argue that ways of demarcating genuine substances do not draw a line with artifacts on one side and natural objects on the other. Finally, I’ll suggest that philosophers have downgraded artifacts because they think of metaphysics as resting on a distinction between what is “mind-independent” and what is “mind-dependent.” I’ll challenge the use of any such distinction as a foundation for metaphysics.

ARTIFACTS AS ORDINARY OBJECTS

Artifacts should fit into any account of ordinary objects for the simple reason that so many ordinary objects are artifacts. We sleep in beds; we eat with knives and forks; we drive cars; we write with computers (or with pencils); we manufacture nails. Without artifacts, there would be no recognizable human life.

On my view—I call it “the Constitution View”—all concrete objects, except for “simples” if there are any, are ultimately constituted by sums (or aggregates) of objects. Technical artifacts—artifacts made to serve some practical purpose—are, like nonartifacts, constituted by lower-level entities. Constitution is a relation of unity-without-identity. Unlike identity, constitution is a contingent and time-bound relation. To take a simple-minded example, consider a wooden rod and a piece of metal with a hole just slightly bigger than the diameter of the rod. When the aggregate of the rod and the piece of metal are in certain circumstances (e.g., when someone wants to make a hammer and inserts the rod into the hole in the metal), a new object—a hammer—comes into being. Since the rod and the piece of metal existed before the hammer did, the relation between the aggregate of the rod and the piece of metal and the hammer is not identity. It is constitution.

Typically, artifacts are constituted by aggregates of things. But not always: a paperclip is constituted by a small piece of thin wire; and a $50 bill is constituted by a piece of paper. Nevertheless, the piece of thin wire and the piece of paper themselves are constituted by aggregates of molecules, which in turn are constituted by aggregates of atoms. So, even those artifacts (like paperclips) that are constituted by a single object are, at a lower level, constituted by aggregates of atoms. For simplicity, I’ll consider artifacts to be constituted by aggregates of things, not by a single object. Any items whatever are an aggregate. The identity conditions of aggregates are simple: aggregate x is identical to aggregate y just in case exactly the same items are in aggregate x and aggregate y.

DIFFERENCES BETWEEN ARTIFACTS AND NATURAL OBJECTS

Technical artifacts have proper functions that they are designed and produced to perform (whether they
by contrast, however the gold was originally picked it goes out of existence altogether. the metal and plastic cube that comes out of the crusher is not the same object (your old clunker of a car) that went in. since artifacts have intended functions essentially, they are what i call “intention-dependent” or “ID” objects; they could not exist in a world without beings with propositional attitudes.

natural objects differ from artifacts in at least three ways: (1) artifacts (and not natural objects) depend ontologically—not just causally—for their existence on human purposes. (2) Relatedly, artifacts are “intention-dependent” (ID) objects that could not exist in a world without minds. Natural objects, which can be deployed to serve human purposes, would exist regardless of human intentions or practices. (3) Artifacts (and not natural objects) essentially have intended proper functions, bestowed on them by beings with beliefs, desires, and intentions.

THE ONTOLOGICAL STATUS OF ARTIFACTS

Many important philosophers—from Aristotle on—hold artifacts ontologically in low regard. Some philosophers have gone so far as to argue that “artifacts such as ships, houses, hammers, and so forth, do not really exist.”6 Artifacts are thought to be lacking in some ontological way: they are considered not to be genuine substances. The notion of substance is a vexed one in philosophy, what I mean by saying that things of some kind (e.g., hammers, dogs, persons)—Fs in general—are genuine substances is that any complete account of what there is will have to include reference to Fs. I shall argue that there is no reasonable basis for distinguishing between artifacts and natural objects in a way that renders natural objects as genuine substances and artifacts as ontologically deficient.

I shall consider five possible ways of distinguishing between natural objects and artifacts, all of which are mentioned or alluded to by David Wiggins.7 On none of these, I shall argue, do natural objects, but not artifacts, turn out to be genuine substances. Let the alphabetic letter “F” be a placeholder for a name of a type of entity.

(1) Fs are genuine substances only if Fs have an internal principle of activity.

(2) Fs are genuine substances only if there are laws that apply to Fs as such, or there could be a science of Fs.

(3) Fs are genuine substances only if whether something is an F is not determined merely by an entity’s satisfying a description.

(4) Fs are genuine substances only if Fs have an underlying intrinsic essence.

(5) Fs are genuine substances only if the identity and persistence of Fs is independent of any intentional activity.

Let us consider (1) through (5) one at a time.

(1) The first condition—Fs are genuine substances only if Fs have an internal principle of activity—has its source in Aristotle.8 Aristotle took this condition to distinguish objects that come from nature (e.g., animals and plants) from objects that come from other efficient causes (e.g., beds). But this condition does not rule in natural objects and rule out artifacts as genuine substances. A piece of gold is a natural object, but today, we would not consider a piece of gold (or any other chemical element) to have an internal principle of change; conversely, a heat-seeking missile is an artifact, but it does have an internal principle of activity. So, the first condition does not distinguish artifacts from natural objects.

(2) The second condition—Fs are genuine substances only if there are laws that apply to Fs as such, or there could be a science of Fs—also allows artifacts to be genuine substances. Engineering schools have courses in materials science (including advanced topics in concrete), traffic engineering, transportation science, computer science—all of which quantify over artifacts. Since something’s being of an artifactual kind (e.g., computer) does not preclude a science of it, the second condition does not make artifacts less than genuine substances.

(3) The third condition is semantic: Fs are genuine substances only if whether something is an F is not determined merely by an entity’s satisfying a description. Demonstrative reference is supposed to be essential to natural-kind terms.9 The reference of natural-kind terms is said to be determined indexically; the reference of artifactual-kind terms is said to be determined by satisfying a description.10 Membership in a natural kind, it is thought, is not determined by satisfying a description, but rather by relevant similarity to stereotypes.9 The idea is this: First, Fs are picked out by their superficial properties (e.g., quantities of water are clear liquids, good to drink, etc.). Then, anything that has the same essential properties that the stereotypes have is an F. So, natural kinds have “extension-involving sortal identifications.”10 By contrast, artifactual terms (like those I used earlier—“beds,” “knives and forks,” “cars,” “computers,” “pencils,” “nails”) are said to refer by satisfying descriptions: “A clock is any time-keeping device, a pen is any rigid ink-applying writing implement and so on.”11 I do not think that this distinction between how words refer captures the difference between natural objects and artifacts.12 The distinction between referring indexically and referring by description, with respect to natural kind terms, is only a matter of the state of our knowledge and of our perceptual systems.12 However gold was originally picked out (e.g., as “stuff like this”), now we can pick it out by [what are taken to be] its essential properties: for example,
Gold is the element with atomic number 79. Not only might natural kinds satisfy descriptions, but also we may refer to artifacts in the absence of any identifying description. For example, archeologists may believe that two entities are both artifacts of the same kind, without having any identifying description of the kind in question. (Were they used in battle or in religious rituals?)

Thus, the third condition—Fs are genuine substances only if whether something is an F is not determined merely by an entity’s satisfying a description—does not distinguish natural kinds from artifactual kinds, nor does it rule out artifacts as genuine objects.14

(4) The fourth condition—Fs are genuine substances only if Fs have an underlying intrinsic essence—also fails to distinguish natural from artifactual kinds. Although some familiar natural kinds—like water or gold—have underlying intrinsic essences, not all do. For example, wings (of birds and insects), mountains, and planets are all natural kinds, but none of them has an underlying intrinsic essence. Their membership in their kinds is not a matter of underlying intrinsic properties. Something is a wing, mountain, or planet not in virtue of what it is made of, but in virtue of its relational properties. For that matter, something is a bird or an insect in virtue of its relational properties—its genealogical lineage. So, having an underlying intrinsic essence does not distinguish natural objects from artifacts.

(5) The fifth condition—Fs are genuine substances only if the character of F is independent of any intentional activity—is the most interesting. According to some philosophers, the “character of [a] substance-kind cannot logically depend upon the beliefs or decisions of any psychological subject.”15 Unlike the first four conditions, the fifth does distinguish between artifactual and natural kinds. An artifact’s being the kind of thing that it is depends on human intentions. Conceding that the necessity of intention is a difference between an artifact and a natural object, I ask: Why should this difference render artifacts deficient?

If you endorse what Jaegwon Kim has called “Alexander’s Dictum”—To be real is to have effects—there is no doubt that artifacts are real. When automobiles were invented, a new kind of thing came into existence: and it changed the world. Considering the world-changing effects of the automobile (and countless other kinds of artifacts), artifacts have as strong a claim to ontological status as natural objects.

What generally underlies the fifth condition, I believe, is an assumption that Fs are genuine substances only if conditions of membership in the substance-kind are set “by nature, and not by us.”16 But it is tendentious to claim that the existence of artifacts depends not on nature, but on us.17 Of course the existence of artifacts depends on us: but we are part of nature. It would be true to say that the existence of artifacts depends not on nature-as-if-we-did-not-exist, but depends on nature-with-us-in-it. Since nature has us in it, this distinction (between nature-as-if-we-did-not-exist and nature-with-us-in-it) is no satisfactory basis for a verdict of ontological inferiority of artifacts.

The Insignificance of the Mind-Independence/Mind-Dependence Distinction

There is a venerable—but, I think, theoretically misguided—distinction in philosophy between what is mind-independent and what is mind-dependent. Anything that depends on our conventions, practices, or language is mind-dependent (and consequently downgraded by those who rest metaphysics on a mind-independence/mind-dependence distinction). All ID objects, including all artifacts, are by definition mind-dependent, inasmuch as they could not exist in a world without beings with beliefs, desires, and intentions. Nothing would be a carburetor in a world without intentional activity.18 The mind-independent/mind-dependent distinction is theoretically misguided because it is used to draw an ontological line in an unilluminating place. It puts mind-independent insects and galaxies on one side, and mind-dependent afterimages and artifacts on the other.

A second reason why the mind-independent/mind-dependent distinction is unhelpful is that advances in technology have blurred the difference between natural objects and artifacts. For example, so-called “digital organisms” are computer programs that (like biological organisms) can mutate, reproduce, and compete with one another.19 Or consider “robo-rats”—rats with implanted electrodes that direct the rats’ movements.20 Or, for another example, consider what one researcher calls “a bacterial battery”21: these are biofuel cells that use microbes to convert organic matter into electricity. Bacterial batteries are the result of a recent discovery of a micro-organism that feeds on sugar and converts it to a stream of electricity. This leads to a stable source of low power that can be used to run sensors of household devices. Finally, scientists are genetically engineering viruses that selectively infect and kill cancer cells and leave healthy cells alone. Scientific American referred to these viruses as “search-and-destroy missiles.”22 Are these objects—the digital organisms, robo-rats, bacterial batteries, genetically engineered viral search-and-destroy missiles—artifacts or natural objects? Does it matter? I suspect that the distinction between artifacts and natural objects will become increasingly fuzzy; and, as it does, the worries about the mind-independent/mind-dependent distinction will fade away. More particularly, as the distinction between natural objects and artifacts pales, the question of the ontological status of web-based objects, for example, becomes more acute.

CONCLUSION

No one who takes artifacts of any sort seriously, ontologically speaking, should suppose that metaphysics can be based on a distinction between mind-independence and mind-dependence. In any case, technology will continue to shrink the distinction, and, with it, the distinction between artifacts and natural objects.23

*An earlier version of this paper was presented to the Society of Philosophy and Technology, Chicago APA, April 20, 2007.

NOTES

1. There is a lot of literature on functions. For example, see Crawford L. Elder, “A Different Kind of Natural Kind,” Australasian Journal of Philosophy 73 (1995): 516–31. See also Pieter E. Vermaas and

SPL-1252
Wybo Houkes, "Ascribing Functions to Technical Artifacts: A Challenge to Etiological Accounts of Functions," *British Journal for the Philosophy of Science* 54 (2003): 261–89. As Vermaas and Houkes point out, some philosophers take the notion of biological function to be basic and then try to apply or transform theories of biological function (which since Darwin are non-intentionalist, reproduction theories) to artifacts. I believe that Vermaas and Houkes are entirely correct to liberate the theory of artifacts from the notion of function in biology.


3. More precisely, a nonderivative artifact has its proper function essentially. The constitutive of an artifact inherits the nonderivative artifact’s proper function and thus has it contingently (as long as it constitutes the nonderivative artifact).


5. All the conditions either follow from, or are part of, the basic distinction that Wiggins draws between natural objects and artifacts. There is a complex condition that natural objects allegedly satisfy and artifacts do not: "...a particular constituent x belongs to a natural kind, or is a natural thing, if and only if x has a principle of activity founded in lawlike dispositions and properties that form the basis for extension-involving sortal identification(s) which will answer truly the question ‘what is x?’" According to Wiggins, natural objects satisfy this condition and artifacts do not. David Wiggins, *Sameness and Substance Renewed* (Cambridge: Cambridge University Press, 2001), 89. I am not claiming that Wiggins denies that there exist artifacts, only that he distinguishes between natural and artificial kinds in ways that may be taken to imply the ontological inferiority of artifacts.

6. A substance has “within itself a principle of motion and sustenance (in respect of place, or of growth and decrease, or by way of alteration).” Aristotle, *Physics* 192b8–23.

7. This claim is similar to the notion that natural-kind terms, but not artificial-kind terms, are rigid designators. (A rigid designator has the same referent in every possible world.) However, what makes the difference between “whale” and “bachelor” is not that only the former is rigid. Rather, only the former term “has its reference determined by causal contact with paradigm samples of the relevant kind.” There is no reason that the terms cannot both be rigid. See Joseph LaPorte, "Rigidity and Kind," *Philosophical Studies* 97 (2000): 304.

8. Although Wiggins is an Aristotelian, this is not Aristotle’s view. For Aristotle, nominal definitions are reference fixers, used to identify objects for scientific study; they contain information that a scientist has before having an account of the essence of the objects. Real definitions are discovered by scientific inquiry and give us knowledge of the essences of objects identified by nominal definitions. Nominal and real definitions are not accounts of different types of entities. Rather, they are different types of accounts of the same entities. Members of a particular natural kind have the same essence (underlying structure). See Robert Bolton, "Essentialism and Semantic Theory in Aristotle: Posterior Analytics, II, 7–10," *The Philosophical Review* 85 (1976): 514–44.


10. Ibid., 89.

11. Ibid., 87.

12. Aristotle would agree with me on this point, I believe. His reason for downgrading artifacts ontologically is that artifacts have no natures in themselves.

13. Moreover, indexicality should not be confused with rigidity, which does not concern how a term gets connected to a referent. For criticism of Putnam’s confusion of the causal theory of reference and indexicality, see Tyler Burge, “Other Bodies,” in *Thought and Object*, edited by Andrew Woodfield (Oxford: Oxford University Press, 1982), 97–120.

14. Joseph LaPorte also holds that some kind expressions (both natural and artificial) designate rigidly, and some designate nonrigidly. See his “Rigidity and Kind,” *Philosophical Studies* 97 (2000): 293–316.


17. In Chapter One of *The Metaphysics of Everyday Life* (Cambridge: Cambridge University Press, 2007), I argued that a distinction between what depends on nature and what depends on us is neither exclusive nor exhaustive.


21. The *New York Times*, September 18, 2003. The lead researcher, Derek Lovley, who coined the term “bacterial battery,” is a microbiologist at the University of Massachusetts-Amherst.


**Artifacts and Mind-Independence: Comments on Lynne Rudder Baker’s “The Shrinking Difference between Artifacts and Natural Objects”**

Amie L. Thomasson  
UNIVERSITY OF MIAMI


Against contemporary reductivist and eliminativist trends, Lynne Baker argues in “The Shrinking Difference between Artifacts and Natural Objects” (2008, following up on her 2007 book) that artifacts should be considered just as genuine parts of our world as natural objects are. I couldn’t agree more. Of five different ways in which one might attempt to distinguish artifacts and natural objects, she argues, four fail to distinguish them, and the fifth, while distinguishing them, does not warrant denying that artifacts are “genuine substances” (2008, 3-4).

The fifth criterion, which Baker admits does distinguish artifacts from natural objects, is that the “identity and persistence” of artifacts depends on human intentions (2008, 3). This fifth criterion admits of (at least) two interpretations, given two different senses in which artifacts are apparently mind-dependent. First, individual artifacts are existentially mind-dependent in the sense that no table, painting, or computer could exist in a world absent of human intentions—in Baker’s terms, they are “Intention-Dependent” objects, such that, as she puts it, “the existence of artifacts depends on us” (2008, 4). This intention-dependence, moreover, is not just a causal matter but a conceptual matter or metaphysical matter: the very idea of an artifact is the idea of an intended product.
of human intentionality (cf. Thomasson forthcoming). Second, artifactual kinds (such as chair, fork, and house) are often thought to be mind-dependent in the sense that what it takes for there to be members of the kind, and under what conditions members of the kind come into existence and cease to exist, are determined by conditions we accept as relevant, rather than forming discoverable features of the world (as the parallel conditions for natural kinds are supposed to). As Baker puts it, the “conditions of membership” in the substance-kind are set by us (2008, 4). So let me add to her case by arguing that neither sense of dependence should lead us to deny that artifacts are real parts of our world.

The first sense of dependence is that individual artifacts are existentially dependent on human intentions. But there are important differences among existentially mind-dependent entities. Imaginary objects might be said to be existentially mind-dependent (if they are allowed to exist at all), but they are the products merely of human thoughts and intentions. By contrast, artifacts such as tables and chairs cannot be brought into existence by thought alone, but also require real physical acts of hammering, assembling, etc., and depend on their material bases as well as on the human intentions that (e.g.) endow them with a function. This alone should help undermine the idea that allowing the existence of any kind of mind-dependent objects involves countenancing “magical modes of creation” (cf. my forthcoming).

Moreover, the thought that any mind-dependence undermines an (alleged) entity’s claim to existence is based on illegitimately generalizing from the case of scientific entities: If we found out that some posited scientific entity (say, a planet or a species of bird) was really just “made up,” a human creation, we might indeed have reason to say that Vulcan (or the Key Sparrow) doesn’t exist. But that reflects the fact that planets and animals are supposed to be mind-independent. The same does not go for artifacts: the very idea of an artifact (or work of art, fictional character, or belief or desire) is of a human creation, and so the fact that (e.g.) a table could not have existed were it not for the relevant human intentions does nothing to undermine its claim to existence.

Thus, a mind-independence criterion may be suitable for would-be natural objects, but not for artifacts (or many other sorts of thing). Since the very idea of an artifact is of something mind-dependent in certain ways, accepting mind-independence as an across-the-board criterion for existence gives us no reason to deny the existence of artifacts; it merely begs the question against them (see my forthcoming). In fact, considering artifacts gives us reason to be suspicious of proposals for across-the-board criteria for existence and suggests that we should instead address existence questions separately, asking in each case what it would take for there to be objects of the kind and then determining whether or not those conditions are fulfilled—while acknowledging that criteria for existence may vary for different sorts of thing (cf. my forthcoming).

The second, perhaps more controversial, sense of dependence is the sense in which the conditions for membership in an artifactual kind, and for the existence, identity, and persistence of its members, are themselves mind-dependent. For, as I have argued elsewhere (2003; 2007), what distinguishes the natures of artifactual kinds from those of chemical or biological kinds is (roughly) that we—the makers and users of artifacts of various kinds—determine what features are and are not essential to being a member of an artifactual kind (like chair, split-level, or convertible), in a way that we do not determine the particular features relevant to being a member of a natural kind (like tiger or gold).

It is often held, however, that possessing a nature that is entirely independent of human concepts, language, etc., which is open to genuine discovery and about which everyone may turn out to be ignorant or in error, is a central criterion for treating kinds as real or genuine parts of our world (Elder 1989, Lakoff 1987). If that’s right, we’re left with the options of giving up an ontology of artifactual kinds or giving up the idea that possessing discoverable mind-independent natures is the central criterion for “really” existing.

I have argued elsewhere (forthcoming) in favor of the latter route. The thought that, to be real, artifactual kinds must have mind-independent natures again comes from borrowing an idea suitable for realism about natural kinds and assuming it must apply wholesale. For while natural kinds may have to have mind-independently discoverable particular natures, to require this of artifactual kinds misconstrues what it is to be a realist about artifactual kinds. For, again, if the analyses I have offered elsewhere (2003; 2007) are correct, it is just part of the very idea of artifactual kinds (as opposed to biological or chemical kinds) that their natures are fixed at least in part by makers’ intentions regarding what features are essential to kind membership—and so ruling out the existence of any kinds with natures of that sort merely begs the question against artifactual kinds.

Let me close by raising one further issue. In her new book (2007), Baker has given us a detailed account of how we can understand artifacts and other everyday objects as constituted by sums of particles, though not identical with them. This is most welcome work, which takes us a good way towards understanding the objects we concern ourselves with in everyday life. But it doesn’t cover all artifacts—if we think of artifacts in the broad sense, as the intended products of human labor. For among the artifacts with which we are most concerned are those I’ve elsewhere (2003b) called “abstract artifacts”—such everyday objects as novels and laws of state, songs and corporations. While these, like other artifacts, are dependent on human intentionality, such entities as Microsoft, the Patriot Act, or Twinkle Little Star are not themselves constituted by sums of particles at all. In fact, it might be said that our interest is increasingly occupied by abstract artifacts rather than concrete ones, as paper money is replaced with abstract sums in our bank accounts, letters replaced with email messages, and billboards and copies of catalogues with websites. And beyond these replacements, of course, a whole range of new abstract artifacts have come to play central roles in our lives, including computer programs, databases, search...
Explaining an Explanatory Gap

Gilbert Harman
PRINCETON UNIVERSITY


Discussions of the mind-body problem often refer to an “explanatory gap” (Levine 1983) between some aspect of our conscious mental life and any imaginable objective physical explanation of that aspect. It seems that whatever physical account of a subjective conscious experience we might imagine will leave it completely puzzling why there should be such a connection between the objective physical story and the subjective conscious experience (Nagel 1974).

What is the significance of this gap in understanding? Chalmers (1996) takes the existence of the gap to be evidence against physicalism in favor of some sort of dualism. Nagel (1974) and Jackson (1982, 1986) see it as evidence that objective physical explanations cannot account for the intrinsic quality of experience, although Jackson (1995, 1998, 2004) later changes his mind and comes to deny that there is such a gap. Searle (1984) argues that an analogous gap between the intrinsic intentional content of a thought or experience and any imaginable functionalist account of that content is evidence against a functionalist account of the intrinsic intentionality of thoughts. On the other hand, McGinn (1991) suggests that these explanatory gaps are due to limitations on the powers of human understanding—we are just not smart enough!

A somewhat different explanation of the explanatory gap appeals to a difference, stressed by Dilthey (1883/1989) and also by Nagel (1974), between two kinds of understanding, objective and subjective. Objective understanding is characteristic of the physical sciences—physics, chemistry, biology, geology, and so on. Subjective understanding does not play a role in the physical sciences but does figure in ordinary psychological interpretation and in what Dilthey calls the “Geisteswissenschaften”—sciences of the mind broadly understood to include parts of sociology, economics, political theory, anthropology, literary criticism, history, and psychology, as well as ordinary psychological reflection.

The physical sciences approach things objectively, describing what objects are made of, how they work, and what their functions are. These sciences aim to discover laws and other regularities involving things and their parts, in this way achieving an understanding of phenomena "from the outside." The social and psychological sciences are concerned in part with such objective understanding, but admit also of a different sort of subjective understanding "from the inside," which Dilthey calls "Das Verstehen." Such phenomena can have content or meaning of a sort that cannot be appreciated within an entirely objective approach. There are aspects of reasons, purposes, feelings, thoughts, and experiences that can only be understood from within, via sympathy or empathy or other translation into one’s own experience.

Suppose, for example, we discover the following regularity in the behavior of members of a particular social group. Every morning at the same time each member of the group performs a fixed sequence of actions: first standing on tip toe, then turning east while rapidly raising his or her arms, then turning north while looking down, and so on, all this for several minutes. We can certainly discover that there is this objective regularity and be able accurately to predict that these people will repeat it every morning, without having any subjective understanding of what they are doing—without knowing whether it is a moderate form of calisthenics, a religious ritual, a dance, or something else. Subjectively to understand what they are doing, we have to know what meaning their actions have for them. That is, not just to see the actions as instances of an objective regularity.

Similarly, consider an objective account of what is going on when another creature has an experience. Such an account may provide a functional account of the person’s brain along with connections between brain events and other happenings in the person’s body as well as happenings outside the person’s body. Dilthey and later Nagel argue that a completely objective account of a creature’s experience may not itself be enough to allow one to understand it in the sense of being able to interpret it or translate it in terms one understands in order to know what it is like for that creature to have that experience. Such an account does not yet provide a translation from that creature’s subjective experience into something one can understand from the inside, based on one’s own way of thinking and feeling.

Nagel observes that there may be no such translation from certain aspects of the other creature’s experiences into

REFERENCES


possible aspects of one's own experiences. As a result, it may be impossible for a human being to understand what it is like to be a bat.

We are not to think of Das Verstehen as a method of discovery or a method of confirming or testing hypotheses that have already been formulated. It is rather needed in order to understand certain hypotheses in the first place. So, for example, to understand a hypothesis or theory about pain involves understanding what it is like to feel pain. An objective account of pain may be found in biology, neuroscience, and psychology, indicating, for example, how pain is caused and what things pain causes (e.g., Melzack and Wall 1983). But it is possible to completely understand this objective story without knowing what it is like to feel pain. There are unfortunate souls who do not feel pain and are therefore not alerted to tissue damage by pain feelings of burning or other injury (Cohen, Kipnis, Kunkle, and Kubansky 1955). If such a person is strictly protected by anxious parents, receives a college education, and becomes a neuroscientist, could that person come to learn all there is to learn about pain? It seems that such a person might fail to learn the most important thing—what it is like to experience pain—because objective science cannot by itself provide that subjective understanding.

Recent defenders of the need for Das Verstehen often mention examples using color or other sensory modalities, for example, a person blind from birth who knows everything there is to know from an objective standpoint about color and people's perception of it without knowing what red things look like to a normal observer (Nagel 1974).

With respect to pain and other sensory experiences there is a contrast between an objective understanding and a subjective understanding of what it is like to have that experience, where such a subjective understanding involves seeing how the objective experience as described from the outside translates into an experience one understands from the inside.

In thinking about this, I find it useful to consider an analogous distinction in philosophical semantics between accounts of meaning in terms of objective features of use, for example, and translational accounts of meaning.

For Quine (1960), an adequate account of the meaning of sentences or other expressions used by a certain person or group of people should provide translations of those expressions into one's "home language." In this sort of view, to give the meaning of an expression in another language is to provide a synonymous expression in one's own language. Similarly, if one wants to give the content of somebody else's thought, one has to find a concept or idea of one's own that is equivalent to it.

Imagine that we have a purely objective theory about what makes an expression in one's home language a good translation of an expression used by someone else. For example, perhaps such a theory explains an objective notion of use or function such that what makes one notion the correct translation of another is that the two notions are used or function in the same way. Such a theory would provide an objective account of correct translation between two languages, objectively described. (This is just an example. The argument is meant to apply to any objective account of meaning.)

To use an objective account of translation to understand an expression as used in another language, at least two further things are required. First, one must be able to identify a certain objectively described language as one's own language, an identification that is itself not fully objective. Second, one must have in one's own language some expression that is used in something like the same way as the expression in the other language. In that case, there will be an objective relation of correct translation from the other language to one's own language, which translates the other expression as a certain expression in one's own language. Given that the correct translation of the other expression is an expression in one's own language "E," one can understand that the other expression means E. "Yes, I see, 'Nicht' in German means not."

This is on the assumption that one has an expression "E" in one's own language that correctly translates the expression in the other language. If not, Das Verstehen will fail. There will be no way in one's own language correctly to say or think that the other expression means E. There is no way to do it except by expanding the expressive power of one's language so that there is a relevant expression "E" in one's modified language.

Let me apply these thoughts about language to the more general problem of understanding what it is like for another creature to have a certain experience. Suppose we have a completely objective account of translation from the possible experiences of one creature to those of another, an account in terms of objective functional relations, for example. That can be used in order to discover what it is like for another creature to have a certain objectively described experience given the satisfaction of two analogous requirements. First, one must be able to identify one objectively described conceptual system as one's own. Second, one must have in that system something with the same or similar functional properties as the given experience. To understand what it is like for the other creature to have that experience is to understand which possible experience of one's own is its translation.

If the latter condition is not satisfied, there will be no way for one to understand what it is like to have the experience in question. There will be no way to do it unless one is somehow able to expand one's own conceptual and experiential resources so that one will be able to have something corresponding to the other's experience.

Knowledge that P requires being able to represent its being the case that P. Limits on what can be represented are limits on what can be known. If understanding what it is like to have a given experience is an instance of knowing that something is the case, then lacking an ability to represent that P keeps one from knowing that something is the case.

About the case in which nothing in one's own system could serve to translate from another creature's experience to
one’s own, Nemirov (1980), Lewis (1998), and Jackson (2004) say in effect that one might merely lack an ability, or know-how, without lacking any knowledge that something is the case. For them, understanding what it is like to have a given experience is not an instance of knowing that something is the case, a conclusion that I find bizarre.

I prefer to repeat that a purely objective account of conscious experience cannot always by itself give an understanding of what it is like to have that experience. There will at least sometimes be an explanatory gap. This explanatory gap has no obvious metaphysical implications. It reflects the distinction between two kinds of understanding: objective understanding and Das Verstehen.

ENDNOTES

1. For additional discussion, see (Harman 1990,1993). I am indebted to comments from Peter Boltuć and Mary Kate McGowan.

BIBLIOGRAPHY


Formulating the Explanatory Gap

Yujin Nagasawa
UNIVERSITY OF BIRMINGHAM


Gilbert Harman (2007) purports to illuminate the intractability of the so-called “explanatory gap” between the phenomenal aspect of consciousness and an objective physical explanation of that aspect by constructing a parallel situation involving translation from one language to another. While I agree with several points that Harman makes regarding the nature of phenomenal consciousness, I have a reservation about his formulation of the explanatory gap. In what follows, I explain my reservation.

Harman’s formulation is based on Thomas Nagel’s well-known example of a bat, which Harman describes as follows:

Nagel observes that there may be no such translation from certain aspects of the other creature’s experiences into possible aspects of one’s own experiences. As a result, it may be impossible for a human being to understand what it is like to be a bat.

Harman then explains the structure of a possible translation that would fill the explanatory gap:

Suppose we have a completely objective account of translation from the possible experiences of one creature to those of another, an account in terms of objective functional relations, for example. That can be used in order to discover what it is like for another creature to have a certain objectively described experience given the satisfaction of two analogous requirements. First, one must be able to identify one objectively described conceptual system as one’s own. Second, one must have in that system something with the same or similar functional properties as the given experience. To understand what it is like for the other creature to have that experience is to understand which possible experience of one’s own is its translation.

Harman’s description of the explanatory gap in terms of translation from bat experience to human experience seems to face the same problem that Nagel’s description faces.

Nagel contends that it is difficult to know how physicalism could be true given that we cannot know what it is like to be a bat, or, that is, that we cannot know the phenomenal aspects of a bat’s sensory experiences. Nagel’s bat example is often said to be so effective because, to any intelligent person, it seems so obvious that a bat’s sonar is nothing like any sensory apparatus that we have.
But exactly why does a bat's having a unique sensory apparatus make it impossible to know what it is like to be one? There are two possible explanations here:

(1) We have to be bats, or at least bat-type creatures that use sonar, in order to know what it is like to be a bat. However, we are neither bats nor bat-type creatures.

(2) An objective, physical characterization of a bat does not tell us what it is like to have sonar, and hence what it is like to be a bat.

Consider (1). If (1) is true, it is difficult to see why physicalism is threatened by the fact that we non-bats cannot know what it is like to be a bat. While physicalism is the ontological thesis that, roughly speaking, everything in this world is physical in the relevant sense, (1) does not entail any significant ontological claim that could undermine physicalism or indeed any other alternatives. It implies only that no human theory, whether it is based on physicalism, dualism, or neutral monism, can tell us what it is like to be a bat, merely because human beings are neither bats nor bat-type creatures. Hence, if (1) is the basis of Nagel’s bat example, it is irrelevant to the cogency of physicalism.

Consider (2). If Nagel and Harman rely on this explanation, then, while (2) is relevant to the cogency of physicalism, ironically, the apparent vividness of the bat example and Harman’s illustration about a translation turn out to be irrelevant. For the plausibility of (2) remains the same even if we replace the term “bat” with, for example, “human being.” We know perfectly well what it is like to be a human being subjectively, but we have no idea how to characterize it objectively and physically. This in itself creates the explanatory gap between the phenomenal aspect of consciousness and an objective physical explanation of that aspect.

The explanatory gap is a very general problem about characterizing fully objectively and physically the phenomenal aspect of consciousness. Thus, it does not really matter whether the phenomenal aspect in question is related to our own type of experience or to those of other animals. It is therefore misleading to say that the explanatory gap is a result of our lacking “a completely objective account of translation from the possible experiences of one creature to those of another.” It is a problem of there being no completely objective account of any experience, whether it is bat experience or human experience.

Suppose we discover somehow that, surprisingly, there is a one-to-one correspondence between a bat’s phenomenal experiences and a human being’s experiences, and that what it is like to be a bat is identical to what it is like to be a human being. Alternatively, suppose that we are the only conscious creatures in the whole universe. The explanatory gap nevertheless remains unfilled because, again, we still do not know how to characterize fully objectively and physically what it is like to be a human being.

Harman’s formulation of the explanatory gap seems therefore to face the following difficulty: Either (i) it is irrelevant to the cogency of physicalism or (ii) if it is relevant, any talk of translation is otiose.

REFERENCES

Logic as a Theory of Computability
Jaakko Hintikka
Boston University


1. Different Frameworks for a Theory of Computability

A general theory of computability must rely on some conceptual framework or other in which the different steps of computation can be codified and by reference to which a theory of computability can be formulated. (For this theory, see, e.g., Davis 1958 or Clote and Krajčiček 1993.) Different frameworks bring different resources to bear on the structure of computations. For instance, the Turing machine framework relates the general theory of computation to questions of computer architecture. The lambda calculus framework has facilitated the development of the denotational semantics for computer languages. The present-day theory of computability can be said to have come about when the main types of framework were shown to yield the same concept of computability.

One frequently used framework is recursion theory. (See here, e.g., Rogers 1967 or Phillips 1992.) In it the notion of a computable function is captured by means of generalization of the familiar use of recursive definitions in elementary arithmetic. In this framework, computable functions are defined as (general) recursive functions. (In the following “recursive function” means what is sometimes called general recursive function.) The class of recursive functions is defined as the smallest class of functions containing zero, successor, and projection functions and closed with respect to the operations of composition, primitive recursion, and minimization. Here a projection function $p^n_m$ takes us from the n-tuple $x_1, x_2, \ldots, x_n$ to $x_m$. The minimization function $\mu_f$ associated with a two-argument function $g(x,y)$ has as its value $\mu_f(y)$ the smallest $x$ such that $g(x,y) = 0$, with the understanding that for all values of $z$, $\mu_f(y)$, $g(z,y)$ is computable but $\mu_f(y)$ is not. The possibility of forming functions by composition or by projection is built into the usual notation of logic. Recursive functions can be either total or partial, that is defined only on a subset of natural numbers.
As a preparation for the treatment of minimization functions it is useful to introduce explicitly the usual notion of minimum through the following recursive equations:

\[(1.1) \quad \min (x,0) = 0\]
\[\min (s(x), s(y)) = s(\min(x,y))\]
\[\min (x,y) = \min(y,x)\]

Here \(s(x)\) is the successor of \(x\). As with other recursive definitions, the value \(\min (x,y)\) is determined by the values \(\min\) for smaller arguments, and hence can be computed in the finite number of steps. Then \(x \geq y\) can be defined as the equation \(\min (x,y) = y\).

2. COMPUTATIONS AS LOGICAL PROOFS

There is one more class of formal operations of which you can ask whether they can serve as a framework of computability theory. They are formalized deductions. The basic idea of trying to use them as a framework for computations is clear. It is to construe the computation of the value \(a = f(b)\) of the function \(f\) for the argument \(b\) as a deduction of the equation \(a=f(b)\) in a suitable system of number theory.

At first sight, this looks easy. For instance, computations of value of a primitive recursive function from their defining equations is accomplished by the repeated use of two rules

\[(S.1) \quad \text{Substitutivity of identicals}\]
\[(S.2) \quad \text{Substitution of a term for a variable}\]

We can restrict \((S.2)\) to substitutions of constant terms for variables. It is convenient to think of all the applications of these rules as taking place in a conjunction of equations whose variable are all bound to suppressed outside universal quantifiers. But what are the relevant equations? What other premises are perhaps needed? (They can be thought of as additional conjuncts.)

These equations and other premises should express operations (listed above) used to form general recursive functions. All these operations can obviously be expressed in the form of equations except for the formation of minimization functions. In order to accommodate these we can extend the class of deductive premises used in the deductions that are interpretable as computations. This can be done by introducing as the relevant extra assumption the conditionals

\[(2.1) \quad (g(x,y) = 0) \supset (g(f(y),y) = 0)\]
\[(g(x,y) = 0) \supset \min ((x,f(y)) = f(y))\]

How can computations using \((2.1)\) be captured deductively? These deductions cannot any longer be merely sequences of equations. They must also involve applications of some propositional rules, since we are now dealing with propositional combinations of equations. Negations of formulas are not involved, for negations occur in \((2.1)\) only in front of identities.

Very little reasoning beyond \((S.1)-(S.2)\) is needed. If for some value \(b\) of \(y\) it is the case that

\[(2.2) \quad (\forall x)(g(x,b) \neq 0)\]

In this case a proof for \((2.2)\) can be found in a finite number of steps, since \(g\) is assumed to be recursive. For the same reason, we can compute, for a given value \(b\) of \(y\), the values of \(g(0,b), g(1,b), \ldots\). If \((2.2)\) is not true, then for some \(d\) we have \(g(d,b)=0\). Then \(f_{g}(b)\) is the smallest \(x\) for which \(g(x,b)=0\). This will be found by computing \(g(d-1, b), g(d-2, b), \ldots\). This presupposes that \(g(0,y), g(1,y), \ldots, g(d,y)\) are defined; otherwise \(f_{g}(y)\) is not defined. But this is in keeping with the usual recursion theory. (See Phillips 1992, p.112.)

The only deductive rules needed for this purpose are \((S.1)-(S.2)\) plus suitable propositional inference rules. This argument shows that if computations using a function \(g(x,y)\) can be construed as first order deductions, then so can computations using the corresponding minimization function \(f_{g}(y)\).

Thus all computations of the values of a general recursive function can be construed as deductions. The premises of these deductions include their defining equations plus definitions of the minimization functions used in the computation. These definitions can be taken to be of the form \((2.1)\).

The recursive function thus defined is usually a partial function. Only when it is the case that

\[(2.3) \quad (\forall y)(\exists x)(g(x,y) = 0)\]

do we have a total function.

If for some reason an instance of \((2.3)\) is false for some value \(b\) of \(y\), then a proof of the negation

\[(2.4) \quad (\forall x)(g(x,b) \neq 0)\]

can be found in a finite number of steps.

3. DEDUCTIVE PROOFS AS COMPUTATIONS

This does not answer the question whether logic can serve as a theoretical framework for computation theory. For this purpose it has to be required that sets of premises of the deductions that reproduce any given computations can somehow be represented by logical formulas doing the same job.

At first sight, this seems easy. We can transform first-order formulas into premises for a computation on the basis of equations as follows:

(a) All formulas are transformed into a negation normal form.

(b) All predicates are replaced by their characteristic functions.
(c) All existential quantifiers are eliminated in terms of their Skolem functions. That is, each subformula of the form \((\exists x)F[x]\) is replaced (in context) by

\[(3.1) \quad F[f(y, y', ..., t, t', ...)\]

where \((Q, y), (Q, y'), ...\) are all the universal quantifiers (in a negation normal form) within the scope of which \((\exists x)\) depends and \(t, t', ...\) are the terms on which \((\exists x)\) depends. (Note that the variables \(y, y', ...\) may occur in \(t, t', ...\).) The function \(f\) is the Skolem function correlated with the occurrence of \((\exists x)\) in question.

This elimination of existential quantifiers is assumed to be carried by step inside out.

(d) Universal quantifiers are moved to the beginning of the formula (or conjunction of formulas in question). They can be thought of as being suppressed.

The result is a propositional combination of equations and negations of equations that can be used for computing values of functions as indicated earlier.

Obviously, in this way we can only compute Skolem functions with the help of other Skolem functions.

4. RESTRICTIONS ON SKOLEM FUNCTIONS

This nevertheless implies a restriction on the functions that can be computed in this way. Not any set of equations that can be used for this purpose can be obtained as Skolem functions of formulas of traditional first-order logic (or of their conjunctions). This is because of the (labelled) tree structure of traditional first-order formulas. Because of this structure, the argument set \((y, y', ..., t, t', y)\) of the Skolem function \(f\) in \((3.1)\) is the set of variables bound to universal quantifiers with the scope of which \((\exists x)\) occurs (plus certain constants). Because of the tree structure (nesting structure) of quantifiers, the members of these argument sets may be function terms. Each formally different function term (including as special cases variables and constant) is considered a different argument. Since the (initial) universal quantifiers do not depend on anything, the argument sets containing only them do not matter.

Because of the tree structure (nesting of scopes) these argument sets of Skolem functions must have the same tree structure. They must be partially ordered by class inclusion with all chains linearly ordered (not branching) in the upwards direction.

Not all sets of equations (and propositional contributions of equations) that can be used for computing recursive functions exhibit such a structure (in the argument sets of their functions). In other words, the partial ordering in question must not exhibit any branching upwards. In such cases, we are dealing with a set of equations (and their propositional combinations) that can be used in computation but cannot be interpreted as a set of deductive premises for a corresponding deduction.

The functions whose argument sets are ordered in this way cannot be Skolem functions of any set of ordinary first-order formulas. Hence, not all the sets of defining equation of general recursive functions can be obtained from ordinary first-order formulas.

For instance, if the assumptions that include \((2.1)\) plus the defining equations listed in the sec. 1, the structure of the relevant argument sets of the relevant functions is the following (with inclusion relation included):

\[(4.1) \quad (x, y) \quad (y, fg(y)) \quad (x, fg(y))\]

Here the upwards chain from \((y)\) branches. The relevant set of defining equations cannot be obtained as Skolem functions of a formula in ordinary first-order logic.

However, if \((2.3)\) is true we can drop the antecedent \(g(x,y)=0\) from \((2.1)\). Then the upwards branching disappears from \((4.1)\). This is the case when the definition \((2.1)\) yields a total function. Accordingly, all total functions are computable by deductions in ordinary first-order logic.

5. LIMITATIONS OF THE RECEIVED FIRST-ORDER LOGIC

Hence it follows that the received logic (the usual first-order logic) cannot serve as a framework for a general theory of computation. However, this is merely due to certain unnecessary limitations of this logic which go back all the way from its first foundation by Frege. Frege overlooked part of the semantical job of quantifiers. This job does not involve only ranging over a class of values and express the nonemptiness or exceptonlessness of certain (usually complex) predicates in this class. By their formal dependence on each other quantifiers also express the actual dependence of their respective variables on each other.

In the received first-order logic these formal dependencies between quantifiers is expressed by the nesting of their scopes. In this way we can only express patterns of dependencies that exhibit a tree structure of the kind mentioned above. It is obvious that this restricts the kinds of set of functions that can be represented in ordinary first-order logic. What has been found here is an example of these restrictions.

These restrictions are partly removed in what is known as independence-friendly (IF) logic. It is obtained from the received first-order logic by allowing an existential quantifier \((\exists y)\) occurring (in a negation normal form) within the scope of \((\forall x)\). This can be expressed by writing it as \((\exists y)\forall x\). The same effect could be reached without any new symbols by liberating the use of parentheses.

In IF logic the law of excluded middle does not automatically hold. Hence it is well suited for discussing partial functions, for instance general recursive functions. Accordingly, we have to distinguish the strong (dual) negation \(\neg\) from the contradictory negation \(\neg\). The use of the letter, albeit
only sentence-initially, characterizes what is known as extended IF logic (EIF logic). This logic has two halves. The $\Sigma$-half is unextended IF logic while the $\Pi$-half consists of contradictory negations of IF formulas.

The use of IF logic as a framework of formal reasoning (formal computation) may seem inappropriate in that there is now complete axiomatization (no complete set of rules of inference) for it. However, in recursion theory we can restrict ourselves to formulas in which all negations (and sequences of negations) of any kind form atomic formulas or identities. For this fragment the usual rules of first-order logic (other than rules for negation) are valid and can be used in deductions.

Another aspect of the same situation is that while IF logic does not admit of a complete proof procedure it has a complete disproof procedure. (A cut-free proof procedure for traditional first-order logic can serve as one.) Conversely, the $\Pi$-half of EIF has a complete proof procedure but not a complete disproof procedure.

The same semantics works for an extended independence-friendly (EFI) logic to which the contradictory negation $\neg$ is admitted, if only into sentence-initial positions.

6. EIF DEDUCTIONS AND COMPUTATIONS

It is now possible to show how one can transform any given computation of the value of a recursive function for a given argument into a logical deduction. A way to do so is obvious to a competent logician, but a survey of what is involved is nevertheless in order. The logical operations take place within the scope of a string of initial universal quantifiers. In that scope, we have a conjunction whose conjuncts specify the operations used in forming the function in question. They include:

(i) The recursion equations for the primitive recursive functions used for the purpose
(ii) The instances of the projection operation used, e.g., $P^1_1 (g(x), h(x,y)) = g(x)$
(iii) Defining equations for all the intermediate functions used.
(iv) For each of the minimization functions used, for instance $f_1 (y)$ obtained from $g(x,y)$, the conjuncts will include (2.1)

The computation proceeds by means of (S.1)-(S.2). A branching structure is created by the disjunctions in the way indicated in sec. 2 above.

The task is to transform this computation into a deductive argument. Since all the rules used in the computation are in effect valid deductive steps, what has to be done is to replace functions by predicates and show how the computational line of reasoning can be extended so as to be carried out in terms of the resulting formulas.

For the purpose, we do the following:

(a) For each different term occurring in the equations we introduce a variable (a constant), if it was not already. Then we introduce to the conjunction the “definitions” of all the new variables as simple functions of others. After this change, all (function) terms are simple (i.e., involve no nesting of functions).

An example makes these explanations clear. For instance, the recursion equations for addition $a(x,y)$ are

\begin{align*}
(6.1) & \quad a(0,y) = y \\
(6.2) & \quad a(s(x),y) = s(a(x,y))
\end{align*}

Here $s(z)$ is the successor of $z$. These equations contain the terms $0, y, a(0,y), s(x), a(s(x),y), a(x,y), s(a(x,y))$. The new equations with their additional variables might be the following:

\begin{align*}
(6.3) & \quad z = s(x) \quad v = a(z,y) \quad u = a(x,y) \quad w = s(u) \\
(6.4) & \quad a(0,y) = y \quad a(z,y) = s(u)
\end{align*}

In general terms, after this change, the computation becomes a matter of applying (S.1)-(S.2) to the new equations. For instance, the computation of $a(2,1)$ is now accomplished by the following substitutions where we take into account the definitions $1 = s(1), 2 = s(1), 3 = s(2)$:

\begin{align*}
(6.5) & \quad a(2,1) = a(s(1),1) = s(a(1,1)) = s(a(s(0),1)) = s(s(s(0))) = s(s(0)) = s(2) = 3
\end{align*}

Each identity results from one of the equations (6.3)-(6.4) by a substitution of constants for variables.

(b) For each function, say $g(x,y)$, we introduce a corresponding predicate $G(x,y,z)$ and add to the conjunction two formulas

\begin{align*}
(6.6) & \quad (\exists z)(\forall x, \forall y) G(x,y,z) \\
(6.7) & \quad (G(x,y,w) \land G(x,y,u)) \Rightarrow (w=u)
\end{align*}

Here $(\forall x)(\forall y)(\forall w)(\forall u)$ are among the initial quantifiers. Notice that if we did not have the slash notation available, we could not make sure (by means of a linear ordering of different quantifiers) that the value of $z$ in (6.6) depends only on $x$ and $y$. This is where the impossibility of construing computations as deductions in the received first-order logic shows up.

(c) Each atomic formula or identity $A(g(x,y))$ containing the term $g(x,y)$ is replaced by

\begin{align*}
(6.8) & \quad (\forall z)(G(x,y,z) \Rightarrow A(z))
\end{align*}

This can of course be done to all the terms in $A$ at the same time. The quantifier $(\forall z)$ is thought of as being moved to a sentence-initial position.
These steps eliminate all functions in terms of predicates.

(d) This change correlates with each simple term, say g(x,y), an existential quantifier (∃z) that depends only on (∀x), (∀y). The substitution of numerical values for x and y combined with an existential instantiation with respect to z yields a constant that is different for each different term g(x,y).

When all these changes have been made, the computation of the value of a recursive function is transformed into a formal logical proof. In the proof, each application of (S.1) remains a substitution of identicals. All substitutions needed are substitutions of constant terms (terms without variables) for variables. Each constant term is in the logical proof introduced by an application of existential instantiation to a formula (6.6). It follows that the number of different terms introduced in the computation equals the number of introductions of new constants in the proof by existential instantiation.

There is another, theoretically simpler way of turning computations into formal deductions. Assume that a function of f can be computed from a conjunction of equations E (perhaps conditional ones like (2.1)) involving the successor function S and a number of auxiliary functions g₁, g₂, ..., gₙ. Then obviously the computations can be seen as a deduction of equations of the form

(6.9) f(a) = s(s(...(s(0)))

from

(6.10) (∃g₁)(∃g₂)...(∃gₙ) E

But (6.10) is a ∑₁ sentence and the deduction can therefore be transformed into a formal deduction in extended IF logic, indeed in its []₁, half for which there exists a complete (formal) axiomatization (proof procedure).

7. IF LOGIC AS A FRAMEWORK OF COMPUTATION

Thus EIT logic can serve as a general framework of computation. Computations can be transformed into EIT deductions, and deductions can be replaced by computations. Questions concerning computation can become problems concerning the deductive structures (the structures of logical consequence relations) in IF or EIF logic. (IF logic is equivalent to the ∑₁ fragment of second-order logic while EIF logic also contains as a mirror image of IF logic an equivalent of the [ ]₁ fragment.) For instance, different algorithms which in the simple equational calculus are codified in the systems S of equations correspond to different deductive premises. For instance, the recipe for forming the minimization function fᵢ(y) can be expressed in IF logic in the form of an explicit premises:

(7.1) (∀y)(∃x)(g(x,y)=0)⇒(g(fᵢ,y), y)=0)

(∀y)(∀x)((g(x,y)=0)⇒(min(x, fᵢ(y)) = fᵢ(y))

Hence questions of consistency and computational power can hence be studied by examining the deductive relations between IF formulas. Likewise, questions of the complexity of computational processes are translated into questions of the complexity of formal deductions. (For such questions, see, e.g., Li and Vitányi 1993.)

It can also be seen that if for a given value of y it is true that

(7.2) (∃x)(g(x,y)=0)

then the difference between ∼ and ∼ (2.1) does not make any difference. In this case, IF logic as distinguished from the received first-order logic is not needed. If this is the case for all values of y, then the computation of the values of fᵢ(y) can be carried out in the ordinary first-order logic which hence is adequate as a framework for studying total recursive functions, as is indeed well known.

There nevertheless seems to be a major limitation of what can be done in this way. For there is no complete axiomatization (proof procedure) for IF logic. However, this is not an actual obstacle, for (as was noted) there exists a complete disproof procedure for IF logic, that is, a recursive enumeration of inconsistent formulas. (This follows from the compactness of IF logic.) In fact, a suitable cut-free proof procedure, such as the tableau method for the received first-order logic, or the tree method can serve this purpose.

Then how do you use the new framework to compute values of functions? A system of equations is coherent in an interesting sense if the following is valid in it for each function f:

(7.3) (∀y)(∃x)(f(y)=x)

Then for each numeral a you can disprove

(7.4) ¬(∃x)(f(a)=x)

But you can disprove this only by deriving a formula of the form

(7.5) ¬¬(f(a)=t)

for some constant term t. But all such constants are built from 0 by means of primitive recursive functions. Hence t can be taken to be a numeral.

Now f(a)=t must be either true or indefinite. In the second case, there cannot exist a different numeral n* such that f(a)=n*. For since f(a) can have only one value, it cannot have a definite value different from n. This is because if it had, the truth value of f(a) could not be indefinite either.

The detailed development of a logic-based theory of computability is too large a task to be undertaken in a single paper. The purpose of this paper is merely to clean obstacles from the way of such theory.
REFERENCE


Hintikka, Jaakko, forthcoming, "On Skolem Functions in Proof Theory."


Robots Need Conscious Perception: A Reply to Aleksander and Haikonen

Stan Franklin
UNIVERSITY OF MEMPHIS

Bernard J. Baars
THE NEUROSCIENCE INSTITUTE, SAN DIEGO

Uma Ramamurthy
ST. JUDE CHILDREN’S RESEARCH HOSPITAL


In response to our article entitled "A Phenomenally Conscious Robot?" Igor Aleksander and Pentti Haikonen were kind enough to write responses (Aleksander 2009; Haikonen 2009).

Aleksander prefers to start with phenomenal consciousness from the start of the modeling process rather than adding on to a functional model. We have no preference in that respect. Global Workspace Theory (GWT) is based on a vast empirical literature with phenomenal experiences as a major testable ingredient (Baars 2002). LIDA began “life” as a functional model. We’re happy to start either way, as long as the result is a working model that also reflects both phenomenal and third-person evidence about consciousness. Let’s aim at modeling phenomenal consciousness.

Our hypothesis is that both GWT as fleshed out in the LIDA model and a coherent perceptual field will prove to be necessary conditions for phenomenal consciousness. This doesn’t assert that “phenomenal states are added to functional structures...” Such functional structures can well be a necessary attribute of phenomenal consciousness without phenomenal states being “add-ons.”

The essence of each of Aleksander’s five axioms for phenomenally conscious states seems to lie in the notion of “feeling.” Aleksander’s term “feelings” is nothing but phenomenal consciousness. This is the famous Hard Problem of consciousness, but Aleksander does not give us an answer to it. It is not clear how GWT-LIDA is expected to solve the Hard Problem if no one else, it is claimed, can do that.

It is only this requirement that prevents a LIDA controlled, functionally conscious software agent from satisfying all five axioms. Put another way, we believe that given time and resources, producing such a functionally conscious software agent or robot, based on the LIDA architecture, that satisfies all five axioms except, perhaps, for the “feeling” requirement, would be a relatively straightforward project. The question of how to determine whether or not such a software agent or robot would have “feelings” would remain, or it might just fade, as has happened with the definition of “the essence of life” and other putatively impossible scientific questions.

Aleksander further asserts that “a neural substrate appears to be necessary to satisfy several aspects of the axioms.” He goes on to assert that “the vividness of phenomenal experience is helped by creating neural state machines with states that use a large number of neurons as state variables.” “Helped,” yes; but “necessary,” is not at all clear. It seems at least plausible to us that such vividness could be achieved in a software agent or robot using a large number of virtual state variables.

Aleksander claims that a sufficiently complex neural network is needed to achieve the perceptual detail and resolution of phenomenal consciousness. He asserts that this “…may be difficult to design within a functionalist framework.” We believe that a functional model, like LIDA, can be designed to achieve any necessary level of perceptual resolution.

Aleksander next discusses “program branching,” asserting that:

A non-neural functionalist representation would be more like frames in movies on film, which branch only through some very smart recognition of some features in stored “coherent perceptual fields”
rather than a system whose state structure directly reflects the dynamic, branching experience.

"Program branching" assumes an old-fashioned AI symbolic agent. A LIDA controlled agent would perceive partially through a slippnet whose recognition occurs via passing of activation from primitive feature detectors, much like neural processing, and not like symbolic AI (Franklin 1993). There are no "stored images" or "stored coherent perceptual fields".

In his conclusion Aleksander claims that "It is very difficult to start with a model based in classical cognitive science, the mode of expression of which is algorithmic and implies virtual systems determined through the intention of a programmer." Assuming that the model described in the previous sentence is intended to be the LIDA model controlling an autonomous agent, the assertion significantly misrepresents LIDA. A LIDA controlled agent is provided by a programmer with sensing capabilities (sensors, primitive feature detectors, etc.), action capabilities (effectors), motivators (feeling/emotions), and a basic cognitive cycle, including several modes of learning, with which to answer the continual, primary question for every autonomous agent, "What shall I do next?" Evolutionary processes provide biological agents, such as humans, with these exact same elements for the exact same purpose. A LIDA based agent in a complex, dynamically changing environment must go through a developmental period as would a human child, and would continue learning thereafter. Again, the LIDA model seems to have been confused with classical symbolic AI models.

In his response Haikonen writes:

What is functional consciousness? Franklin, Baars, and Ramamurthy answer: "An agent is said to be functionally consciousness (sic) if its control structure implements the Global Workspace Theory and the LIDA Cognitive Cycle." However, this is not a proof. This is a definition and as such conveniently eliminates the need to study if there were anything in this implementation that could even remotely qualify as and resemble functional consciousness. This kind of study might be difficult, because in nature there is no such thing as functional consciousness.

There are no proofs in science in the sense of evidence so strong as to not be susceptible to challenge. Such proofs belong only in mathematics. Every scientific "fact" is continually open to challenge. Correct mathematical proofs are not.

There is a sizable and growing body of evidence from cognitive science and neuroscience that human minds (their control structures) implement the essential elements of Global Workspace Theory (Baars 2002; Gaillard et al. 2009) and the LIDA Cognitive Cycle (Canolty et al. 2006; Jensen & Colgin 2007; Massimini et al. 2005; Uchida, Kepecs, & Mainen 2006; van Berkum 2006; Willis & Todorov 2006). This satisfies our definition of functional consciousness (Franklin 2003).

Haikonen goes on to assert that:

On the other hand, Merker (2005) has proposed that phenomenal consciousness produces a stable perceptual world by distinguishing real motion from the apparent motion produced by the movement of the sensors. Franklin reads this proposition backwards and concludes that phenomenal consciousness can be produced by the production of stable perceptual world.

Franklin concluded no such thing. Haikonen begins his response by correctly asserting that the FBR paper "propose[s] that providing a functionally conscious robot with stable coherent perceptual world might be a step towards a phenomenally conscious machine." Note his own phrase "might be a step toward."

Accusing FBR of faulty logic, Haikonen claims that "...stable perception cannot be a cause for phenomenal consciousness. Merker's original proposition and Franklin's conclusion must be suspected." There was no faulty logic since the stated conclusion was never drawn. Also, it seems possible that a stable perceptual world might be part of a sufficient set of conditions for phenomenal consciousness, without being necessary. In other words, removing the stable perceptual world condition from the sufficient set might render it no longer sufficient, without the removed condition being necessary for phenomenal consciousness.

REFERENCES


Flawed Workspaces?

Pentti O. A. Haikonen
UNIVERSITY OF ILLINOIS AT SPRINGFIELD


ABSTRACT

The Global Workspace architecture of Baars has been proposed as a valid model for the mammalian brain. On the other hand, the Global Workspace Architecture is based on earlier Blackboard architectures. In these architectures a common working memory area, so-called blackboard or global workspace is shared with a number of specialist modules. This operation calls for a common way of information representation, a common code that can be understood by all the specialist modules. The author argues that no such code exists in the brain and is not required either.

In his recent book Murray Shanahan has outlined a revised version of the Global Workspace model, where the role of the common workspace as a working memory is changed to that of a common communications structure. Shanahan recognizes the common code problem and offers a solution, but it is argued here that the proposed solution only makes things worse.

The need for a common code can be removed by the rejection of the global workspace, but then the system will no longer have the Global Workspace Architecture.

ABOUT THE BAARS GLOBAL WORKSPACE ARCHITECTURE

The Global Workspace Architecture of Baars (1988) is a parallel information processing architecture that is controlled in a serial way. It consists of a number of specialist modules and a common workspace area. This workspace is a working memory that is supposed to contain the intermediate pieces of information that should eventually converge into the solution of the ongoing cognitive task. A specialist module is able to transmit its information to the common workspace area if and only if that information is more relevant to the ongoing cognitive task than the information from the other modules. The global workspace area, in turn, broadcasts continuously its contents to all specialist modules. These modules should then be able to process that information and see if their specific expertise would allow any relevant contribution.

The Global Workspace architecture is a theatre model of consciousness. The Global Workspace is a stage where the consciously perceived mental content comes together for the attention of the audience, the subconscious specialist modules.

Baars proposes that the Global Workspace Architecture is also a valid model for the mammalian brain; the brain would be organized in a similar way as a collection of parallel specialist processes or modules and a separate global workspace area (Shanahan & Baars 2005).

Baars proposes further that this architecture would allow the distinction between conscious and unconscious processing of information. According to Baars, the specialist modules process information in a parallel sub-conscious way and only the information that is serially broadcast by the global workspace is processed in a conscious way. Superficially this would dovetail well with the folk psychology observation of the operation of the brain; we seem to have one serial stream of thoughts and consciousness, yet it is known that the brain processes subconsciously and in a parallel way most of its information. Baars also lists some neurological findings for the support of the Global Workspace model of the brain. Therefore, should we finally announce the Global Workspace model as the winner?

PROBLEMS WITH THE BAARS MODEL

Now there is a catch. Baars (Shanahan & Baars 2005) admits that the Global Workspace Architecture was inspired by and is derived from earlier Artificial Intelligence blackboard systems like those of Hayes-Roth (1985) and Nii (1986). In fact, the Baars Global Workspace Architecture is a blackboard system that is described in the terms of cognitive science. In blackboard systems it is important that each specialist is able to read and understand what is on the blackboard. Therefore, there must be a common way of representation, a common language or code that each specialist can understand (Corkill 1991). A specialist module may well use its own representations in its internal processes, but it must transmit its data using the common code. Likewise, a common code would be necessary in the Global Workspace Architecture, where the global workspace corresponds to the blackboard. In software simulations of the Global Workspace this requirement may easily go unnoticed because all information must anyway be represented as statements in a formal computer language. However, in actual neural hardware realizations this requirement emerges as a complication. If the designer wishes to create a common working memory for a number of different specialist modalities, then he will soon realize that the information must be represented in a way that is understood by all the modalities; there must be a common code. The designer will soon realize that a common working memory is redundant; distributed memory function inside the specialist modules will suffice and no common communication code is necessary if the modules communicate associatively with each other.

Thus, the author argues that in the brain there is no common code. The brain utilizes only neural signal patterns that initially arise from sensory stimuli. Neural signal patterns in different modalities may be associated with each other without any consideration of their meaning and in that sense the process is self-coding and correlating. Even higher level natural language may emerge in this way, without any low-level common neural code (Haikonen 2007).

THE SHANAHAN MODEL

In his recent, nice book Embodiment and the Inner Life Murray Shanahan presents an alternative Global Workspace model, which is inspired by the Baars model. However, in his model Shanahan rejects the role of the global workspace as a theatre stage. Originally, Baars describes the global workspace as a working memory that acts as a
theatre stage, where the conscious contents of the mind is played out. Shanahan rejects this and proposes that instead of a theatre stage, the global workspace should be thought of as a communications infrastructure, which connects the various autonomous specialist units with each other (2010, 111). But, according to Shanahan, this infrastructure should not be taken only as a transmission medium. Shanahan proposes (2010, 148): “GWS is not only the locus of broadcast, but also the arena for competition between rival coalitions and the medium of coupling for the members of those coalitions.” This infrastructure has limited capacity and therefore forms a temporally limiting bottleneck; this would explain the serial nature of consciousness.

In Shanahan’s model, the global workspace infrastructure explains the conscious/unconscious distinction in a similar way to that of Baars. Conscious processes are mediated by the global workspace communication structure while the specialist units operate unconsciously. The global workspace would allow an integrated response and it would enable learning and episodic memory making, etc. by allowing the various modules to cooperate on the same topic (Shanahan 2010, 112).

Shanahan is aware of the common code problem and proposes that there is no need for a lingua franca (2010, 118) within the framework of global workspace communication structure. If a module A is to influence modules B and C, then this influence can be mediated by different signals; i.e., the module A would use dedicated codes when transmitting to different modules. Shanahan states: “The signals going to B and C from A do not have to be the same.” Obviously, this method bypasses the original problem of common code, but at the same time it creates further problems. The module A cannot now broadcast one universal signal pattern, it has to send different signal patterns to different modules. In a way, the transmitting module would now have to master a number of languages instead of a single, universal one. During communication the transmitting module would have to generate a large number of different signal patterns for the receiving modules. This is a hefty complication.

MODELS WITHOUT GLOBAL WORKSPACE EXIST

Architectures with a number of autonomous specialist modules may also work without a global workspace as a working memory and a theatre stage or as a communications infrastructure. The rejection of global workspace structures will also remove the need for a common code. The various specialist modules may well communicate directly with each other associatively and use various threshold strategies for significance controlled attention. This will still allow the distinction between conscious and unconscious operation in the Baars style of contrastive analysis. Each specialist module would operate basically in the same way whether the overall operation were “conscious” or “non-conscious.” The overall operation will display many hallmarks of consciousness whenever the various modules are cooperating in unison, focussing their attention on the same object. In this mode a large number of cross-connections would be activated, allowing the subsequent forming of associative memories. Therefore, a “conscious” event would be remembered for a while and could be reported in the terms of the various modalities such as sounds, words, gestures, etc. The “conscious” stream of mental events would be a serial one. This is the way of operation of the Haikonen cognitive architecture (Haikonen 2007). In the Haikonen architecture no common code exists. Each module broadcasts its own signal pattern, which will evoke responses associatively in the receiving modules. There is no need to interpret these signal patterns in any other way in the receiving modules. A recent robotic realization proves the practical feasibility of this approach (Haikonen 2010).

The thalamo-cortical interaction and other neurological findings that Baars (Shanahan & Baars 2005) lists as evidence for the global workspace model may as well if not better be interpreted as a proof of the existence of perception-feedback loops, functionally similar to those proposed by Chella (2008), Hesslow (2002), and Haikonen (2003, 2007).

CONCLUSIONS

From the biological point of view, a cognitive model that operates without a common code would be more plausible than another one that would require such a code. It is not immediately clear why and how evolution would lead to such complication, when it is not needed. Therefore, the author argues that global workspace architectures that call for a common code are flawed brain models.

Shanahan’s model rejects the global workspace as a common working memory, but does not really remedy the common code problem. In Shanahan’s approach the common code is replaced by a large number of different signal patterns. In comparison to architectures with direct associative communication between modules, Shanahan’s model is unnecessarily complex.

REFERENCES


Unity from Multiplicity: A Reply to Haikonen

Murray Shanahan

IMPERIAL COLLEGE LONDON


Haikonen’s article is a welcome opportunity to clear up an aspect of global workspace theory that invites misinterpretation. The two essential components of the architecture are a set of parallel specialist processes and a global workspace. On a first encounter with Baars’s theory, it is natural to assume that the “messages” broadcast from the global workspace to the set of parallel specialist modules must conform to some common “representational format.” The architecture, it seems, demands a psychological language franca. Many cognitive scientists today, in an era when neuroscience is more directly relevant to understanding cognition and consciousness, are reluctant to talk in terms of representations and the “language of thought.” A theory that apparently demands some sort of universal internal language will be unattractive to many, including Haikonen and myself.

But it should be born in mind that the book that introduced Baars’s theory (A Cognitive Theory of Consciousness) was published in 1988, and its style of presentation reflects the cognitivist assumptions prevalent at the time. Baars’s theory is still relevant today to the extent that we can see beyond these cognitivist overtones to its essential insights. The idea of a common representational format can be done away with if we think of the global workspace as a communications infrastructure that allows local brain processes to exercise widespread influence on the rest of the brain. The challenge then is to characterize this notion of influence in a plausible way. This is one of the tasks attempted in my book, Embodiment and the Inner Life. Unfortunately, I seem to have been less clear than I would have liked. If I may be forgiven for the crime of quoting myself, here is the relevant passage:

“If a process A influences two other process B and C then the nature of this influence can be thought of as mediated by information if 1) a variety of signals can pass from A to B and from A to C, and 2) the responses of B and C are sensitive to this variety. We might expect a given pattern of activation in A to influence B the same way on one occasion as on another, and equally to influence C the same way on one occasion as another. But, the signals going to B and C from A do not have to be the same. There is no need for further stipulations. The specific structure of the signals drops out of the equation. (118)

Haikonen grants that this avoids the common code problem, but believes that it creates further difficulties. But he mistakenly asserts that “The module A cannot now broadcast one universal signal pattern, it has to send different signal patterns to different modules.” In fact, my account claims that the signals can be different, not that they must be. But this is not the heart of the misunderstanding, which becomes apparent when Haikonen suggests that “In a way, the transmitting module would now have to master a number of languages instead of a single universal one. During communication the transmitting module would have to generate a large number of different signal patterns for the receiving modules.”

It would indeed disadvantage my account if any part of the architecture in any sense had to “master a number of languages.” This is not at all how I see it (and I don’t think there is anything in the text of my book to support this interpretation). The task of “decoding” broadcast signals is the job of the receiving processes. Moreover, we should avoid any temptation to smuggle semantics into these signals. (Hence the scare quotes around “decoding.”) This decoding business is a matter of finding (and being influenced by) regularities and correlations that are present in those signals in a purely information-theoretic sense. This is a job that can be carried out by a Hebbian learning rule, such as spike-timing dependent plasticity (STDP). The result is a self-organizing system of global signalling mediated by the brain’s long-range white matter connections.

Curiously, this viewpoint is almost identical to Haikonen’s. “In the Haikonen architecture no common code exists. Each module broadcasts its own signal pattern, which will evoke responses associatively in the receiving modules. There is no need to interpret these signal patterns in any other way in the receiving modules.”

Thus far, Haikonen’s interpretation of my work notwithstanding, we fully concur. In fact, our approaches are similar and compatible in many respects. However, there is a genuine point of disagreement between us. Haikonen believes he can account for the conscious/unconscious distinction without appealing to a global workspace because the operation of his architecture “will display many hallmarks of consciousness whenever the various modules are co-operating in unison, focussing their attention on the same object.” In his 2007 book, Haikonen states that his model “does not use a global workspace, which is seen to be redundant as the modules can communicate and compete directly with each other” (189).

The difficulty here is that there is nothing to prevent competing coalitions of processes (modules) from simultaneously forming while “focusing their attention” on different objects. In an intelligent artefact that is not designed to emulate the biological brain, there may be any number of ways to resolve such conflicts. But a fundamental property of the conscious condition, whether or not it is realized in a biological substrate, is unity. Only one coalition of processes at a time can triumph and dominate the dynamics of the system. In the conscious condition, the brain and the multitude of processes that constitute it act as an integrated whole. Global workspace theory accounts for this because the communications infrastructure of the global workspace has limited bandwidth. Only one coalition of processes at a time can gain access to it, funneling the flow of influence and information, directing the subject onto a single, unified object, or allowing a single, unified
thought to form. In other words, global workspace architecture explains how serial arises from parallel, and how unity arises from multiplicity.

REFERENCES


---

**Leibniz, Complexity, and Incompleteness¹**

Gregory Chaitin

IBM RESEARCH


Let me start with Hermann Weyl, who was a fine mathematician and mathematical physicist. He wrote books on quantum mechanics and general relativity. He also wrote two books on philosophy: The *Open World: Three Lectures on the Metaphysical Implications of Science* (1932), a small book with three lectures that Weyl gave at Yale University in New Haven, and *Philosophy of Mathematics and Natural Science*, published by Princeton University Press in 1949, an expanded version of a book he originally published in German.

In these two books Weyl emphasizes the importance for the philosophy of science of an idea that Leibniz had about complexity, a very fundamental idea. The question is what is a law of nature, what does it mean to say that nature follows laws? Here is how Weyl explains Leibniz’s idea in *The Open World*, pp. 40-41: The concept of a law becomes vacuous if arbitrarily complicated laws are permitted, for then there is always a law. In other words, given any set of experimental data, there is always a complicated *ad hoc* law. That is valueless; simplicity is an intrinsic part of the concept of a law of nature.

What did Leibniz actually say about complexity? Well, I have been able to find three or perhaps four places where Leibniz says something important about complexity. Let me run through them before I return to Weyl and Popper and more modern developments.

First of all, Leibniz refers to complexity in Sections V and VI of his 1686 *Discours de métaphysique*, notes he wrote when his attempt to improve the pumps removing water from the silver mines in the Harz mountains was interrupted by a snow storm. These notes were not published until more than a century after Leibniz’s death. In fact, most of Leibniz’s best ideas were expressed in letters to the leading European intellectuals of his time, or were found many years after Leibniz’s death in his private papers. You must remember that at that time there were not many scientific journals. Instead, European intellectuals were joined in what was referred to as the Republic of Letters. Indeed, publishing could be risky. Leibniz sent a summary of the *Discours de métaphysique* to the philosophe Arnauld, himself a Jansenist fugitive from Louis XIV, who was so horrified at the possible heretical implications that Leibniz never sent the *Discours* to anyone else. Also, the title of the *Discours* was supplied by the editor who found it among Leibniz’s papers, not by Leibniz.

I should add that Leibniz’s papers were preserved by chance, because most of them dealt with affairs of state. When Leibniz died, his patron, the Duke of Hanover, by then the King of England, ordered that they be preserved, sealed, in the Hanover royal archives, not given to Leibniz’s relatives. Furthermore, Leibniz produced no definitive summary of his views. His ideas are always in a constant state of development, and he flies like a butterfly from subject to subject, throwing out fundamental ideas, but rarely, except in the case of the calculus, pausing to develop them.

In Section V of the *Discours*, Leibniz states that God has created the best of all possible worlds, in that all the richness and diversity that we observe in the universe is the product of a simple, elegant, beautiful set of ideas. God simultaneously maximizes the richness of the world, and minimizes the complexity of the laws which determine this world. In modern terminology, the world is understandable, comprehensible, science is possible. You see, the *Discours* was written in 1686, the year before Leibniz’s nemesis Newton published his *Principia*, when medieval theology and modern science, then called mechanical philosophy, still coexisted. At that time the question of why science is possible was still a serious one. Modern science was still young and had not yet obliterated all opposition.

The deeper idea, the one that so impressed Weyl, is in Section VI of the *Discours*. There Leibniz considers “experimental data” obtained by scattering spots of ink on a piece of paper by shaking a quill pen. Consider the finite set of data points thus obtained, and let us ask what it means to say that they obey a law of nature. Well, says Leibniz, that cannot just mean that there is a mathematical equation passing through that set of points, because there is always such an equation! The set of points obey a law only if there is a *simple* equation passing through them, not if the equation is “fort composée” = very complex, because then there is always an equation.

Another place where Leibniz refers to complexity is in Section 7 of his *Principles of Nature and Grace* (1714), where he asks why is there something rather than nothing, why is the world non-empty, because “nothing is simpler and easier than something!” In modern terms, where does the complexity in the world come from? In Leibniz’s view, from God; in modern terminology, from the choice of the laws of nature and the initial conditions that determine the world. Here I should mention a remarkable contemporary development: Max Tegmark’s amazing idea that the ensemble of all possible laws, all possible universes, is simpler than picking any individual universe. In other words, the multiverse is more fundamental than the question of the laws of our particular universe, which merely happens to be our postal address in the multiverse of all possible worlds! To illustrate this idea, the set of all positive integers 1, 2, 3, . . . is very simple, even though
particular positive integers such as 9859436643312312 can be arbitrarily complex.

A third place where Leibniz refers to complexity is in Sections 33-35 of his Monadology (1714), where he discusses what it means to provide a mathematical proof. He observes that to prove a complicated statement we break it up into simpler statements, until we reach statements that are so simple that they are self-evident and don’t need to be proved. In other words, a proof reduces something complicated to a consequence of simpler statements, with an infinite regress avoided by stopping when our analysis reduces things to a consequence of principles that are so simple that no proof is required.

There may be yet another interesting remark by Leibniz on complexity, but I have not been able to discover the original source and verify this. It seems that Leibniz was once asked why he had avoided crushing a spider, whereupon he replied that it was a shame to destroy such an intricate mechanism. If we take “intricate” to be a synonym for “complex,” then this perhaps shows that Leibniz appreciated that biological organisms are extremely complex.

These are the four most interesting texts by Leibniz on complexity that I’ve discovered. As my friend Stephen Wolfram has remarked, the vast Leibniz Nachlass may well conceal other treasures, because editors publish only what they can understand. This happens only when an age has independently developed an idea to the point that they can appreciate its value plus the fact that Leibniz captured the essential concept.

Having told you about what I think are the most interesting observations that Leibniz makes about simplicity and complexity, let me get back to Weyl and Popper. Weyl observes that this crucial idea of complexity, the fundamental role of which has been identified by Leibniz, is unfortunately very hard to pin down. How can we measure the complexity of an equation? Well, roughly speaking, by how complicated its equation is.

To repeat, we consider a computer program to be a theory for its output, that is the essential idea, and both theory and output are finite strings of bits whose size can be compared. To proceed mathematically, to define complexity precisely, we now consider the size of programs, the number of bits in the software that implements our theory:

**Law:** Equation → Software,

**Complexity:** Size of equation → Size of program, Bits of software.

The following diagram illustrates the central idea of AIT, which is a very simple toy model of the scientific enterprise:

Theory (01100...11) → COMPUTER → Experimental Data (110...0).

In this model, both the theory and the data are finite strings of bits. A theory is software for explaining the data, and in the AIT model this means the software produces or calculates the data exactly, without any mistakes. In other words, words in our model a scientific theory is a program whose output is the data, self-contained software, without any input.

And what becomes of Leibniz’s fundamental observation about the meaning of “law?” Before there was always a complicated equation that passes through the data points. Now there is always a theory with the same number of bits as the data it explains, because the software can always contain the data it is trying to calculate as a constant, thus avoiding any calculation. Here we do not have a law; there is no real theory. Data follows a law, can be understood, only if the program for calculating it is much smaller than the data it explains.

In other words, understanding is compression, comprehension is compression, a scientific theory unifies many seemingly disparate phenomena and shows that they reflect a common underlying mechanism.

To repeat, we consider a computer program to be a theory for its output, that is the essential idea, and both theory and output are finite strings of bits whose size can be compared. The best theory is the smallest program that produces that data, that precise output. That’s our version of what some people call Occam’s razor. This approach enables us to proceed mathematically, to define complexity precisely and to prove things about it. And once you start down this road, the first thing you discover is that most finite strings of bits are lawless, algorithmically irreducible, algorithmically random, because there is no theory substantially smaller than the data itself. In other words, the smallest program that produces that output has about the same size as the output. The second thing you discover is that you can never be sure you have the best theory.

Before I discuss this, perhaps I should mention that AIT was originally proposed, independently, by three people, Ray Solomonoff, A. N. Kolmogorov, and myself, in the 1960s. But the original theory was not quite right. A decade later, in the mid 1970s, what I believe to be the definitive version of the theory emerged, this time independently due to me and to Leonid Levin, although Levin did not get the
definition of relative complexity precisely right. I will say more about the 1970s version of AIT, which employs what I call “self-delimiting programs,” later, when I discuss the halting probability Ω.

But for now, let me get back to the question of proving that you have the best theory, that you have the smallest program that produces the output it does. Is this easy to do? It turns out this is extremely difficult to do, and this provides a new complexity-based view of incompleteness that is very different from the classical incompleteness results of Gödel (1931) and Turing (1936). Let me show you why.

First of all, I’ll call a program “elegant” if it’s the best theory for its output, if it is the smallest program in your programming language that produces the output it does. We fix the programming language under discussion, and we consider the problem of using a formal axiomatic theory, a mathematical theory with a finite number of axioms written in an artificial formal language and employing the rules of mathematical logic, to prove that individual programs are elegant. Let’s show that this is hard to do by considering the following program P:

- P produces the output of the first provably elegant program that is larger than P.

In other words, P systematically searches through the tree of all possible proofs in the formal theory until it finds a proof that a program Q, that is larger than P, is elegant, then P runs this program Q and produces the same output that Q does. But this is impossible, because P is too small to produce that output! P cannot produce the same output as a provably elegant program Q that is larger than P, not by the definition of elegant, not if we assume that all provably elegant programs are in fact actually elegant. Hence, if our formal theory only proves that elegant programs are elegant, then it can only prove that finitely many individual programs are elegant.

This is a rather different way to get incompleteness, not at all like Gödel’s “This statement is unprovable” or Turing’s observation that no formal theory can enable you to always solve individual instances of the halting problem. It’s different because it involves complexity. It shows that the world of mathematical ideas is infinitely complex, while our formal theories necessarily have finite complexity. Indeed, just proving that individual programs are elegant requires infinite complexity. And what precisely do I mean by the complexity of a formal mathematical theory? Well, if you take a close look at the paradoxical program P above, whose size gives an upper bound on what can be proved, that upper bound is essentially just the size in bits of a program for running through the tree of all possible proofs using mathematical logic to produce all the theorems, all the consequences of our axioms. In other words, in AIT the complexity of a math theory is just the size of the smallest program for generating all the theorems of the theory.

And what we just proved is that if a program Q is more complicated than your theory T, T can’t enable you to prove that Q is elegant. In other words, it takes an N-bit theory to prove that an N-bit program is elegant. The Platonic world of mathematical ideas is infinitely complex, but what we can know is only a finite part of this infinite complexity, depending on the complexity of our theories.

Let’s now compare math with biology. Biology deals with very complicated systems. There are no simple equations for your spouse, or for a human society. But math is even more complicated than biology. The human genome consists of 3 × 10⁹ bases, which is 6 × 10⁶ bits, which is large, but which is only finite. Math, however, is infinitely complicated, provably so.

An even more dramatic illustration of these ideas is provided by the halting probability Ω, which is defined to be the probability that a program generated by coin tossing eventually halts. In other words, each K-bit program that halts contributes 1 over 2^K to the halting probability Ω. To show that Ω is a well-defined probability between zero and one is essential to use the 1970s version of AIT with self-delimiting programs. With the 1960s version of AIT, the halting probability cannot be defined, because the sum of the relevant probabilities diverges, which is one of the reasons it was necessary to change AIT.

Anyway, Ω is a kind of DNA for pure math, because it tells you the answer to every individual instance of the halting problem. Furthermore, if you write Ω’s numerical value out in binary, in base-two, what you get is an infinite string of irreducible mathematical facts:

Ω = .11011...

Each of these bits, each bit of Ω, has to be a 0 or a 1, but it’s so delicately balanced, that we will never know. More precisely, it takes an N-bit theory to be able to determine N bits of Ω.

Employing Leibnizian terminology, we can restate this as follows: The bits of Ω are mathematical facts that refute the principle of sufficient reason, because there is no reason they have the values they do, no reason simpler than themselves. The bits of Ω are in the Platonic world of ideas and therefore necessary truths, but they look very much like contingent truths, like accidents. And that’s the surprising place where Leibniz’s ideas on complexity lead, to a place where math seems to have no structure, none that we will ever be able to perceive. How would Leibniz react to this?

First of all, I think that he would instantly be able to understand everything. He knew all about 0s and 1s, and had even proposed that the Duke of Hanover cast a silver medal in honor of base-two arithmetic, in honor of the fact that everything can be represented by 0s and 1s. Several designs for this medal were found among Leibniz’s papers, but they were never cast, until Stephen Wolfram took one and had it made in silver and gave it to me as a sixtieth birthday present. And Leibniz also understood very well the idea of a formal theory as one in which we can mechanically deduce all the consequences. In fact, the calculus was just one case of this. Christian Huygens, who taught Leibniz mathematics in Paris, hated the calculus, because it was...
mechanical and automatically gave answers, merely with formal manipulations, without any understanding of what the formulas meant. But that was precisely the idea, and how Leibniz’s version of the calculus differed from Newton’s. Leibniz invented a notation which led you automatically, mechanically, to the answer, just by following certain formal rules.

And the idea of computing by machine was certainly not foreign to Leibniz. He was elected to the London Royal Society, before the priority dispute with Newton soured everything, on the basis of his design for a machine to multiply. (Pascal’s original calculating machine could only add.)

So I do not think that Leibniz would have been shocked; I think that he would have liked $\Omega$ and its paradoxical properties. Leibniz was open to all systèmes du monde, he found good in every philosophy, ancient, scholastic, mechanical, Kabbalah, alchemy, Chinese, Catholic, Protestant. He delighted in showing that apparently contradictory philosophical systems were, in fact, compatible. This was at the heart of his effort to reunify Catholicism and Protestantism. And I believe it explains the fantastic character of his Monadology, which, complicated as it was, showed that certain apparently contradictory ideas were, in fact, not totally irreconcilable.

I think we need ideas to inspire us. And one way to do this is to pick heroes who exemplify the best that mankind can produce. We could do much worse than pick Leibniz as one of these exemplifying heroes.  

NOTES
1. Lecture given Friday, June 6, 2008, at the University of Rome “Tor Vergata,” in a meeting on “Causality, Meaningful Complexity, and Knowledge Construction.” I thank Professor Arturo Carsetti for inviting me to give this talk.
2. For more on such themes, please see Chaitin, Meta Maths, Atlantic Books, London, 2006, or the collection of my philosophical papers, Chaitin, Thinking about Gödel and Turing (Singapore: World Scientific, 2007).

Architecture-Based Motivation vs. Reward-Based Motivation

Aaron Sloman
UNIVERSITY OF BIRMINGHAM


INTRODUCTION

“Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.”

—David Hume, A Treatise of Human Nature (2.3.3.4), 1739–1740 (http://www.class.uidaho.edu/mickelsen/ToC/hume%20treatise%20ToC.htm)

Whatever Hume may have meant by this, and whatever various commentators may have taken him to mean, I claim that there is at least one interpretation in which this statement is obviously true, namely: no matter what factual information an animal or machine A contains, and no matter what competences A has regarding abilities to reason, to plan, to predict, or to explain, A will not actually do anything unless it has, in addition, some sort of control mechanism that selects among the many alternative processes that A’s information and competences can support.

In short: control mechanisms are required in addition to factual information and reasoning mechanisms if A is to do anything. This paper is about what forms of control are required. I assume that in at least some cases there are motives, and the control arises out of selection of a motive for action. That raises the question where motives come from. My answer is that they can be generated and selected in different ways, but one way is not itself motivated: it merely involves the operation of mechanisms in the architecture of A that generate motives and select some of them for action. The view I wish to oppose is that all motives must somehow serve the interests of A, or be rewarding for A. This view is widely held and is based on a lack of imagination about possible designs for working system. I summarize it as the assumption that all motivation must be reward-based. In contrast, I claim that at least some motivation may be architecture-based, in the sense explained below.

Instead of talking about “passions,” I shall use the less emotive terms “motivation” and “motive.” A motive in this context is a specification of something to be done or achieved (which could include preventing or avoiding some state of affairs, or maintaining a state or process). The words “motivation” and “motivational” can be used to describe the states, processes, and mechanisms concerned with production of motives, their control and management, and the effects of motives in initiating and controlling internal and external behaviors. So Hume’s claim, as interpreted here, is that no collection of beliefs and reasoning capabilities can generate behavior on its own: motivation is also required.

This view of Hume’s claim is expressed well in the Stanford Encyclopedia of Philosophy entry on motivation, though without explicit reference to Hume:

The belief that an antibiotic will cure a specific infection may move an individual to take the antibiotic, if she also believes that she has the infection, and if she either desires to be cured or judges that she ought to treat the infection for her own good. All on its own, however, an empirical belief like this one appears to carry with it no particular motivational impact; a person can judge that an antibiotic will most effectively cure a specific infection without being moved one way or another. (http://plato.stanford.edu/entries/moral-motivation)

That raises the question: Where do motives come from and why are some possible motives (e.g., going for lunch)
selected and others (e.g., going for a walk, or starting a campaign for election to parliament) not selected?

If Hume had known about reflexes, he might have treated them as an alternative mode of initiation of behavior to motivation (or passions). There may be some who regard a knee-jerk reflex as involving a kind of motivation produced by tapping a sensitive part of the knee. That would not be a common usage. I think it is more helpful to regard such physical reflexes as different from motives, and therefore as exceptions to Hume’s claim. I shall try to show that something like “internal reflexes” in an information-processing system can be part of the explanation of creation and adoption of motives. In particular, adopting the “design-based approach to the study of mind” yields a wider variety of possible explanations of how minds work than is typically considered in philosophy or psychology, and paradoxically even in AI/Robotics, where such an approach ought to be more influential.

This proposal opposes a view that all motives are selected on the basis of the costs and benefits of achieving them, which we can loosely characterize as the claim that all motivation is “reward-based.”

In the history of philosophy and psychology there have been many theories of motivation, and distinctions between different sorts of motivation, for example, motivations related to biological needs, motivations somehow acquired through cultural influences, motivations related to achieving or maximizing some reward (e.g., food, admiration in others, going to heaven), or avoiding or minimizing some punishment (often labelled positive and negative reward or reinforcement), motivations that are means to some other end, and motivations that are desired for their own sake, motivations related to intellectual or other achievements, and so on. Many theorists assume that motivation must be linked to rewards or utility. One version of this (a form of hedonism) is the assumption that all actions are done for ultimately selfish reasons.

I shall try to explain why there is an alternative kind of motivation, architecture-based motivation, which is not included even in this rather broad characterization of types of motivation on Wikipedia:

Motivation is the set of reasons that determines one to engage in a particular behavior. The term is generally used for human motivation but, theoretically, it can be used to describe the causes for animal behavior as well. This article refers to human motivation. According to various theories, motivation may be rooted in the basic need to minimize physical pain and maximize pleasure, or it may include specific needs such as eating and resting, or a desired object, hobby, goal, state of being, ideal, or it may be attributed to less-apparent reasons such as altruism, morality, or avoiding mortality. (http://en.wikipedia.org/wiki/Motivation)

Philosophers who write about motivation tend to have rather different concerns such as whether there is a necessary connection between deciding what one morally ought to do and being motivated to do it. For more on this see the afore-mentioned entry in the Stanford Encyclopedia of Philosophy.

Motivation is also a topic of great concern in management theory and management practice, where motivation of workers comes from outside them, e.g., in the form of reward mechanisms (providing money, status, recognition, etc.) sometimes in other forms, e.g., inspiration, exhortation, social pressures. I shall not discuss any of those ideas.

In psychology and even in AI, all these concerns can arise, though I am here only discussing questions about the mechanisms that underlie processes within an organism or machine that select things to aim for and which initiate and control the behaviors that result. This includes mechanisms that produce goals and desires, mechanisms that identify and resolve conflicts between different goals or desires, mechanisms that select means to achieving goals or desires.

Achieving a desired goal G could be done in different ways, e.g.,

- select and use an available plan for doing things of type G
- use a planning mechanism to create a plan to achieve G and follow it
- detect and follow a gradient that appears to lead to achieving G (e.g., if G is being on high ground to avoid a rising tide, walk uphill while you can)

There is much more to be said about the forms different motives can have, and the various ways in which their status can change, e.g., when a motive has been generated but not yet selected, when it has been selected, but not yet scheduled, or when there is not yet any clear plan or strategy as to how to achieve it, or whether action has or has not been initiated, whether any conflict with other motives, or unexpected obstacle has been detected, etc.

For a characterization of some of the largely unnoticed complexity of motives see: http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#16.


WHERE DO MOTIVES COME FROM?

It is often assumed that motivation, i.e., an organism’s or machine’s, selection, maintenance, or pursuance of some state of affairs, the motive’s content, must be related to the organism or machine having information (e.g., a belief or expectation) that achievement of the motive will bring some rewards or benefit, sometimes referred to as “utility.” This could be reduction of some disadvantage or disutility, e.g., a decrease in danger or pain.

Extreme versions of this assumption are found in philosophical theories that all agents are ultimately selfish,
since they can only be motivated to do things that reward themselves, even if that is a case of feeling good about helping someone else.

More generally, the assumption is that selection of a motive among possible motives must be based on some kind of prediction about the consequences of achieving or preventing whatever state of affairs is specified in that motive. This document challenges that claim by demonstrating that it is possible for an organism or machine to have, and to act on, motives for which there is no such prediction.

**MY CLAIM**

My claim is that an organism (human or non-human) or machine may have something as a motive whose existence is merely a product of the operation of a motive-generating mechanism—which itself may be a product of evolution, or something produced by a designer, or something that resulted from a learning or developmental process, or, in some cases, may be produced by some pathology. Where the mechanism comes from and what its benefits are are irrelevant to its being a motivational mechanism: all that matters is that it should generate motives, and thereby be capable of influencing selection of behaviors.

In other words, it is possible for there to be reflex mechanisms whose effect is to produce new motives, and in simple cases to initiate behaviors controlled by such motives. I shall present a very simple architecture illustrating this possibility below, though for any actual organism, or intelligent robot, a more complex architecture will be required, for reasons given later.

Where the reflex mechanisms come from is a separate question: they may be produced by a robot designer or by biological evolution, or by a learning process, or even by some pathology (e.g., mechanisms producing addictions) but what the origin of such a mechanism is, is a separate question from what it does, how it does it, and what the consequences are.

I am not denying that some motives are concerned with producing benefits for the agent. It may even be the case (which I doubt) that most motives generated in humans and other animals are selected because of their benefit for the individual. For now, I am merely claiming that something different can occur and does occur, as follows:

- Not all the mechanisms for generating motives in a particular organism O, and not all the motives produced in O have to be related to any reward or positive or negative reinforcement for O.

- What makes them motives is how they work: what effects they have, or, in more complex cases, what effects they tend to have even though they are suppressed (e.g., since competing, incompatible, motives can exist in O).

---

**LEARNING AND MOTIVATION**

Many researchers in AI and other disciplines (though not all) assume that learning must be related to reward in some way, e.g., through positive or negative reinforcement.

I think that is false: some forms of learning occur simply because the opportunity to learn arises and the information-processing architecture produced by biological evolution simply reacts to many opportunities to learn, or to do things that could produce learning because the mechanisms that achieve that have proved their worth in previous generations, without the animals concerned knowing that they are using those mechanisms nor why they are using them.

**ARCHITECTURE-BASED MOTIVATION**

Consider a very simple design for an organism or machine (Figure 1). It has a perceptual system that forms descriptions of a process occurring in the environment. Those descriptions are copied/stored in a database of “current beliefs” about what is happening in the world or has recently happened.

![Figure 1. A simple design for an organism or machine.](image)

At regular intervals another mechanism selects one of the beliefs about processes occurring recently and copies its content (perhaps with some minor modification or removal of some detail, such as direction of motion) to form the content of a new motive in a database of “desires.” The desires may be removed after a time.

At regular intervals an intention-forming mechanism selects one of the desires to act as a goal for a planning mechanism that works out which actions could make the desire come true, selects a plan, then initiates plan execution.

This system will automatically generate motives to produce actions that repeat or continue changes that it has recently perceived, possibly with slight modifications, and it will adjust its behaviors so as to execute a plan for fulfilling the latest selected motive.

Why is a planning mechanism required instead of a much simpler reflex action mechanism that does not require motives to be formulated and planning to occur?
A reflex mechanism would be fine if evolution had detected all the situations that can arise and if it had produced a mechanism that is able to trigger the fine details of the actions in all such situations. In general that is impossible, so instead of a process automatically triggering behavior it can trigger the formation of some goal to be achieved, and then a secondary process can work out how to achieve it in the light of the then current situation.

For such a system to work there is NO need for the motives selected or the actions performed to produce any reward. We have goals generated and acted on without any reward being required for the system to work. Moreover, a side effect of such processes might be that the system observes what happens when these actions are performed in varying circumstances, and thereby learns things about how the environment works. That can be a side effect without being an explicit goal.

A designer could put such a mechanism into a robot as a way of producing such learning without that being the robot's goal. Likewise, biological evolution could have selected changes that lead to such mechanisms existing in some organisms because they produce useful learning, without any of the individual animals knowing that they have such mechanisms nor how they were selected or how they operate.

**MORE COMPLEX VARIATIONS**

There is no need for the motive generating mechanism to be so simple. Some motives triggered by perceiving a physical process could involve systematic variations on the theme of the process, e.g., undoing its effects, reversing the process, preventing the process from terminating, joining in and contributing to an ongoing process, or repeating the process, but with some object or action or instrument replaced. A mechanism that could generate such variations would accelerate learning about how things work in the environment, if the effects of various actions are recorded or generalized or compared with previous records, generalizations, and predictions.

The motives generated will certainly need to change with the age and sophistication of the learner.

Some of the motive-generating mechanisms could be less directly triggered by particular perceived episodes and more influenced by the previous history of the individual, taking account not only of physical events but also social phenomena, e.g., discovering what peers seem to approve of, or choose to do. The motives generated by inferring motives of others could vary according to stage of development. For example, early motives might mainly be copies of inferred motives of others, then as the child develops the ability to distinguish safe from unsafe experiments, the motives triggered by discovering motives of others could include various generalizations or modifications, e.g., generalizing some motive to a wider class of situations, or restricting it to a narrower class, or even generating motives to oppose the perceived motives of others (e.g., parents!).

Moreover, some of the processes triggered instead of producing external actions could produce internal changes to the architecture or its mechanisms. Those changes could include production of new motive generators, or motive comparators, or motive generator generators, etc.

For more on this idea see chapter 6 and chapter 10 of *The Computer Revolution in Philosophy* (1978).

**MECHANISMS REQUIRED**

In humans it seems that architecture-based motivation plays a role at various levels of cognitive development, and is manifested in early play and exploration, and in intellectual curiosity later on, e.g., in connection with things like mathematics or chess, and various forms of competitiveness.

Such learning would depend on other mechanisms monitoring the results of behavior generated by architecture-based motivational mechanisms and looking for both new generalizations, new conjectured explanations of those generalizations, and new evidence that old theories or old conceptual systems are flawed—and require debugging.

Such learning processes would require additional complex mechanisms, including mechanisms concerned with construction and use of powerful forms of representation and mechanisms for producing substantive (i.e., non-definitional) ontology extension.

For more on additional mechanisms required see

http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#glang

Evolution of minds and languages. What evolved first and develops first in children: Languages for communicating, or languages for thinking (Generalised Languages: GLs)  

http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#prague09

Ontologies for baby animals and robots. From “baby stuff” to the world of adult science: Developmental AI from a Kantian viewpoint.

http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#toddlers

A New Approach to Philosophy of Mathematics: Design a young explorer, able to discover “toddler theorems” (Or: “The Naive Mathematics Manifesto”).

The mechanisms constructing architecture-based motivational sub-systems could sometimes go wrong, accounting for some pathologies, e.g., obsessions, addictions, etc. But at present that is merely conjecture.

**CONCLUSION**

If all this is correct, then humans, like many other organisms, may have many motives that exist not because having them benefits the individual but because ancestors
with the mechanisms that produce those motives in those situations happened to produce more descendants than conspecifics without those mechanisms did. Some social insect species in which workers act as “slaves” serving the needs of larvae and the queen appear to be examples. In those cases it may be the case that

Some motivational mechanisms “reward” the genomes that specify them, not the individuals that have them.

Similarly, some forms of learning may occur because animals that have certain learning mechanisms had ancestors who produced more offspring than rivals that lacked those learning mechanisms. This could be the case without the learning mechanism specifically benefiting the individual. In fact, the learning mechanism may lead to parents adopting suicidal behaviors in order to divert predators from their children.

It follows that any AI and cognitive science research based on the assumption that learning is produced ONLY by mechanisms that maximize expected utility for the individual organism or robot is likely to miss out on important forms of learning. Perhaps the most important forms.

One reason for this is that typically individuals that have opportunities to learn do not know enough to be able to even begin to assess the long-term utility of what they are doing. So they have to rely on what evolution has learnt (or a designer in the case of robots) and, at a later stage, on what the culture has learnt. What evolution or a culture has learnt may, of course, not be appropriate in new circumstances!

This discussion note does not prove that evolution produced organisms that make use of architecture-based motivation in which at least some motives are produced and acted on without any reward mechanism being required. But it illustrates the possibility, thereby challenging the assumption that ALL motivation must arise out of expected rewards.

Similar arguments about how suitably designed reflex mechanisms may react to perceived processes and states of affairs by modifying internal information stores could show that at least some forms of learning use mechanisms that are not concerned with rewards, with positive or negative reinforcement, or with utility maximization (or maximization of expected utility). My conjecture is that the most important forms of learning in advanced intelligent systems (e.g., some aspects of language learning in human children) are architecture-based, not reward based. But that requires further investigation.

The ideas presented here are very relevant to projects like CogX, which aim to investigate designs for robots that “self-understand” and “self-extend,” since it demonstrates at least the possibility that some forms of self-extension may not be reward-driven, but architecture-driven.

Various forms of architecture-based motivation seem to be required for the development of precursors of mathematical competences described here: http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#toddlers.

Some of what is called “curiosity-driven” behavior probably needs to be re-described as “architecture-based” or “architecture-driven.”

This is one of a series of notes explaining how learning about underlying mechanisms can alter our views about the “logical topography” of a range of phenomena, suggesting that our current conceptual schemes (Gilbert Ryle’s “logical geography”) can be revised and improved, at least for the purposes of science, technology, education, and maybe even for everyday conversation, as explained in http://www.cs.bham.ac.uk/research/projects/cogaff/misc/logical-geometry.html.

NOTE

Marvin Minsky wrote quite a lot about goals and how they are formed in The Emotion Machine. It seems to me that the above is consistent with what he wrote, though I may have misinterpreted him.

Something like the ideas presented here were taken for granted when I wrote The Computer Revolution in Philosophy in 1978. However, at that time I underestimated the importance of spelling out assumptions and conjectures in much greater detail.

ACKNOWLEDGMENTS

I wish to thank Veronica Arriola Rios and Damien Duff for helpful comments on an earlier, less clear draft.

Consciousness, Engineering, and Anthropomorphism

Ricardo Sanz
UNIVERSIDAD POLITÉCNICA DE MADRID


1. INTRODUCTION

The construction of conscious machines seems to be central to the old, core dream of the artificial intelligence community. It may well be a maximal challenge motivated by the pure hybris of builders playing God’s role in creating new beings. It may also just be a challenging target to fuel researchers’ motivation. However, we may be deeply puzzled concerning the reasons for engineers to pursue such an objective. Why do engineers want conscious machines? I am not saying that engineers are free from hubris or not in need for motivation, but I question if there is an engineering reason to do so.

In this article I will try to analyze such motives to discover these reasons and, in this process, reveal the excessive anthropomorphism that permeates this endeavor. Anthropomorphism is an easy trap, especially for philosophers. We can see it pervasively tainting the philosophy of consciousness. However, in the modest opinion of this engineer, philosophy shall transcend humanism and focus on universal issues of value both for animals and machines.
2. THE ENGINEERING STANCE

The construction of intelligent machines is the central activity of control systems engineering. In fact, the core focus of activity of the control systems engineer is the design and implementation of minds for machines. For most people involved in cognitive science, saying that a PID controller is a mind is not just an overstatement; it is, simply, false. False because such a system lacks emotions, education, growth, learning, genealogy, personality . . . whatever.

This analysis of what minds are suffers from biological chauvinism. Anthropomorphism is pervasive in cognitive science, artificial intelligence, and robotics. This is understandable for historical reasons but shall be factored-out in the search for core mechanisms of mind.

A central principle of engineering is that systems shall include what is needed and only what is needed. Over-engineered systems are too complex, late delivered, and uneconomical. This principle shall also be applied to the endeavor of building conscious machines.

2.1 THE SYSTEMS ENGINEERING VEE

The systems engineering lifecycle (see Figure 1) starts with the specification of needs: what does the user of the system need from it. This need is stated in the form of a collection of user and system requirements. The verification of satisfaction of these needs—the system validation for acceptance testing—is the final stage of the engineering life-cycle.

This process implies that all system elements—what is built at the construction stage—do always address a user or system need; they always have a function to perform. The concrete functions that are needed will depend on the type of system and we shall be aware of the simple fact that not all systems are robots. Or, to be more specific, not all intelligent systems are humanoid robots. Many times, intelligent minds are built for other kinds of systems. Intelligence is deployed in the sophisticated controllers that are needed to endow machines with the capability to address complex tasks. Minds are just control systems. Intelligent minds are sophisticated control systems.

2.2 NOT ALL AI SYSTEMS ARE ROBOTS

The obvious fact that not all AI systems are humanoid robots has important implications. The first one is that not all systems perform activities usually done by humans and hence:

5. Their realizations—their bodies—do not necessarily resemble human bodies. In engineering, bodies follow functional needs in a very intentional and teleological sense. Machines are artificial in the precise sense clarified by Simon.

6. Their environments—the context where they perform the activity—are not human environments and fitness imply non-humanly capabilities.

7. Their missions—what are they built for—are sometimes human missions, but mostly not. People are worried about robots getting our jobs but most robot jobs cannot be performed by humans.

In control systems engineering we usually make the distinction between the controller—the mind—and the plant—the body. This may sound kind of cartesian and indeed it is. But it is not due to a metaphysical stance of control engineers but to the more earthly, common practice of addressing system construction by the integration of separately built parts.

The plant (usually an artefact) can hence be quite close or quite different from humans or from animals:

- **Airplanes:** Share the environment and the activity with birds, but their functional ways are so different from animals that control strategies are totally different.

- **Industrial Robots:** In many cases can do activities that humans could do: welding, picking, packaging, etc. but requirements may be far from human: precision, speed, repeatability, weight, etc.

- **Vehicles:** Autonomous vehicles share activity with animals: movement. However they are steadily departing from animal contexts and capabilities. Consider, for example, the use of GPS for autonomous driving or vehicle-to-infrastructure communication for augmented efficiency.

- **Chemical Plants:** Some artefacts are extremely different from humans seen as autonomous entities moving in environments. Industrial continuous processes—chemical, oil, food—do not resemble humans nor animals and the needs for intelligence and awareness are hence quite different.

- **Utilities:** The same can be said for technical infrastructure. The intelligence of the smart grid is not close to animal intelligence.

All these systems “live” in dynamic contexts and their controllers shall react appropriately to changing
environmental conditions. They process sensory signals to be “aware” of relevant changes, but they do it in very different ways. Machines are not animals nor in their realization nor in their teleology. Bioinspiration can help systems engineers in the provision of architected ideas of concrete designs of subsystems. However, mapping the whole iguana to a machine is not a sound engineering strategy.6

2.3 THE AI PROGRAM VS. THE STANDARD STRATEGY
From my perspective the many threads of the global AI program can be categorized into three basic kinds of motivations:

- Technology. Solving problems by means of incorporating intelligence into the artefacts.
- Science. Explore the nature of (human) intelligence by creating computer models of psychological theories.
- Hubris. Create beings like us.

Control system engineering (CSE) implements AIs because it is interested in the problem-solving capabilities that AI can provide to their machines. AI enters the CSE domain to deal with runtime problems of higher complexity that are not easily addressable by more conventional means. The mind of the machine is built as a cognitive agent that perceives and acts on the body that is situated in an environment. In their well-known textbook Russell and Norvig even say that “the concept of a controller in control theory is identical to that of an agent in AI.”7

AI-based controllers can decide in real-time about what to do in complex situations to achieve system goals. Goals that are established in terms of user needs and, secondarily, in terms of machine needs. An AI-based controller does not pursue the machine objectives but the objectives of its owner.8

The optimal strategy for controlling a system is to invert a perfect model of it. But this only works to the extent that the model behavior matches system behavior. Model fidelity is limited due to several factors. Observability limits what the intelligent agent may perceive. Note that the intelligent agent interacts with a body that interacts with the environment. In this situation perfect knowledge is unachievable because uncertainty permeating both the plant and its environment affects the intelligent agent mental representation (cf. the problems surrounding the deployment of autonomous cars).

Agent mental complexity shall match that of the plant and its environment following Ashby law of required variety. This implies that the curse of complexity and uncertainty affects not only the plant and its environment, but also the controller itself. Intelligent, autonomous controllers are enormously complex artefacts.

Complexity plays against system dependability. The probability of failure multiplies with system complexity. This is not good for real-life systems like cars, factories, or gas networks. The basic method to improve dependability is building better systems. Systems of better quality or systems built using better engineering processes (e.g., as is the case of cleanroom engineering). However, these do-well strategies are not easily translatable to the construction of systems of required high complexity.

The standard strategy to address runtime problems is to use humans to directly drive or supervise the system. Humans are better at addressing the unexpected and provide augmented robustness and resilience (R&R). A term that is gaining acceptance these days is Socio-Cyber-Physical System. A system composed of physical bodies, software controllers and humans. Figure 2 shows a common layering of these systems.

![Figure 2. A socio-cyber-physical system is a layered structure of interacting systems. The top authority corresponds to human operators because they are able to deal with higher levels of uncertainty.](image)

In socio-cyber-physical systems the top authority corresponds to human operators because they are able to deal with higher levels of uncertainty. Humans are able to understand better what is going on, especially when unexpected things happen. The world of the unexpected has never been a friendly world for AIs.

3. BUILDING CONSCIOUS MACHINES
The research for consciousness in artificial systems engineering can be aligned with the three motivations described in the previous section—useful technology, psychological science, or mere hubris.

Some authors consider that biological consciousness is just an epiphenomenon. However, an evolutionary psychology dogma states that any currently active mental trait that has been exposed to evolutionary pressure has adaptive value. This—in principle—implies that consciousness has adaptive (behavioral) value; so it may be useful in machines.

From a technological stance, the analysis/evolution of complex control systems took us into researching novel strategies to improve system resilience by means of self-
awareness. If the system is able to perceive and reason about its own disturbances, it will be able to act upon them and recover mission-oriented function. Machine consciousness enters the engineering agenda as a possible strategy to cope with complexity and uncertainty.

A conscious machine can reflect upon itself and this may be a potential solution to the curse of complexity problem. So, the engineering interest in consciousness is specifically focused on one concrete aspect: self-awareness. This implies that the engineering stance does not have much to say about other aspects of consciousness (esp. qualia).

3.1 SELF-AWARENESS IN MACHINES

Self-aware machines are aware of themselves. Self-awareness is just a particular case of awareness when the object of awareness is the machine itself. Self-awareness is a class of perceptual process, mapping the state of a system—the machine itself—into an exercisable representation. From an engineering perspective, self-awareness is useless unless it is accompanied by concurrent action processes. In particular, to be of any use concerning system resilience, self-awareness processes need coupled self-action processes. This closes a control loop of the system upon itself.

This may sound enormously challenging and innovative, but this is not new at all. Systems that observe and act upon themselves have been common trade for decades. There are plenty of examples of self-X mechanisms in technical systems that in most cases are not based on biology:

- Fault-tolerant systems (from the 60s)
- Adaptive controllers (from the 70s)
- Metacognitive systems (90s)
- Autonomic Computing (00s)
- Adaptive service systems (00s)
- Organic Computing (10s)

All these systems observe themselves and use these observations to adapt a system’s behavior to changing circumstances. These changes may be due to system-external disturbances or system-internal operational conditions. The adaptation to external changes has been widely investigated, but the adaptation to internal changes has received less attention.

In our own case we investigate domain-neutral, application-neutral architectures for augmented autonomy based on model-based reflective adaptive controllers. Domain neutral means that we investigate architectures for any kind of system—e.g., mobile robots or chemical factories—and application neutral means that the architectures shall provide functionality for any kind of application—e.g., for system fault tolerance or dynamic service provision.

3.2 AN EXAMPLE: A METACONTROLLER FOR ROBOTS

Figure 3 shows an implementation of a self-aware system that improves resilience of a mobile robot. The self-awareness mechanism is a metacontroller—a controller of a controller—that manages the operational state of the robot. This metacontroller has been designed to mitigate the resilience reduction due to potential faults in the control system of the robot.

The robot controller is a very complex distributed software system that can suffer transient or permanent faults in any of its components. The metacontroller monitors the state of the robot controller and acts upon it to keep system functionality by reorganizing its functional organization. This is similar to what humans do when overcoming some of the problems of becoming blind by learning to read with the fingers.

Function, functional state, and componential organization are core concepts in this approach. In this system this self-awareness mechanism provides reaction to disruption and improves mission-level resilience. And this is grounded in a self-model based on formally specified concepts concerning the system and its mission.

Figure 4 shows part of the formal ontology that is used in the implementation of the perception, reasoning, and action mechanisms of the self-awareness engine. It enables the robot to reason about its own body and mind.

3.3 INTO PHILOSOPHY

This research on machine self-awareness obviously enters philosophical waters.
in causal chains and computational procedures that begin before consciousness and extend beyond it. Indeed, mind is scarcely a locus in its own right and certainly does not have its own space. It is a through road (or a small part of one) rather than a dwelling. Consciousness is voided of inwardness.”

For Tallis, causal theories of consciousness reduce mind to a set of input/output relations, with the net effect of effectively “emptying” consciousness. Causal links—the stuff machines are made with—seem not enough for the machinery of mind.

The same phenomenon can be found in any context where human mental traits and “similar” machine traits are put under the scope of the scientist or philosopher. For example, in a recent conference on philosophy of AI, a speaker raised the question “Is attention necessary for visual object classification?” A few slides later the speaker showed that Google was doing this without attention. So the answer of the question was NO. End. Surprisingly, the presentation continued with a long discussion about phenomena of human perception.

4.1 ANTHROPO-X
Cognitive Science and the Philosophy of Mind, and to some extent Artificial Intelligence and Robotics, are anthropocentric, anthroposcoped and anthropobiased. They focus on humans; they address mostly humans; they think of humans as special cases of mind and consciousness. Obviously, the human mind ranks quite high in the spectrum of mind. Maybe it is indeed the peak of the scale. But this does not qualify it as special in the same sense that elephants or whales are not special animals, however big.

However, the worst problem is that all these disciplines are also anthropomorphic: They shape all their theories using the human form. Protagoras seems still alive and man is used to measure all things mental.

This is not only wrong, but severely limiting. Anthropomorphism has very bad effects in consciousness research:

- Human consciousness traits are considered general consciousness traits. This has the consequence for artificial systems of posing extra, unneeded requirements for the implementation of cognitive engines for machines (see, for example, the wide literature on cognitive architectures).
- Some non-human traits are not properly addressed in the theories; because being out the the human spectrum are considered irrelevant concerning the achievement of machine consciousness.

4.2 RETHINKING CONSCIOUSNESS TRAITS
Some commonly accepted consciousness aspects shall be rethought under a non-chauvinistic light to achieve the generality that science and engineering require.
For example, consciousness *seriality* and *integration* have been hallmarks of some widely quoted theories of consciousness. Functional departures from these are considered pathological, but this is only true under the anthropocentric perspective. Consciousness seriality implies that an agent can only have one stream of consciousness. However, from a general systems perspective, nothing prevents a machine from having several simultaneous streams.

This aspect of seriality is closely related to attention. According to Taylor, attention is the crucial gateway to consciousness and architectural models of consciousness shall be based on attention mechanisms. However, using the same analysis as before, nothing prevents a machine from paying attention to several processes simultaneously. The rationale of attention mechanisms seems to be the efficient use of limited sensory processing resources (esp. at higher levels of the cognitive perception pipeline). But in the case of machines, if the machine architecture is scalable enough, it is in principle possible to incorporate perceptual resources as needed to pay concurrent attention to several processes. This is also related to the limited capacity trait of human consciousness.

Integration is another human trait that may be unnecessary in machines. Consciousness integration implies that the collection of experiences flowing up from the senses are integrated in a single experiential event. Dissociated experience is abnormal—pathological—in humans but it need not be so in machines. In fact, in some circumstances, being able to keep separated streams of consciousness may be a benefit for machines (e.g., for cloud-based services for conscious machines).

Concepts like subjectivity, individuality, consciousness ontogenesis and filogenesis, emotional experience, sensory qualia modalities, etc. all suffer under this same analysis. Any future, sound theory of consciousness shall necessarily deal with a wider spectrum of traits like bat audio qualia or robot LIDAR qualia.

5. CONCLUSIONS

Universality renders deep benefits. This has been widely demonstrated in science, for example, when the dynamics of falling objects on the earth surface was unified with the dynamics of celestial objects.

We must escape the trap of anthropomorphism to reach a suitable theory for artificial consciousness engineering. Just consider the history of “artificial flight,” “artificial singing,” or “artificial light.” We need not create mini suns to illuminate our rooms. We don’t need to copy, nor imitate, nor fake human consciousness for our machines. We don’t need the whole iguana of mind; what we need are analysis and first principles.

Any general (non anthropocentric) consciousness research program will produce benefits also in the studies of human consciousness because it can provide inspiration for deeper, alternate views of human consciousness. For example, human brains have parallel activities (not serial but concurrent) in different levels of an heterarchy. Considerations coming from conscious distributed artificial systems will help clarify issues of individual minds—normal and pathological—and social consciousness.

Philosophy is a bold endeavor. It goes for the whole picture. Philosophy of mind shall be aware of this and realize that consciousness escapes humanity.

NOTES

5. In fact, there are control systems engineering methods that do not perform this separation, addressing mind and body as a single whole by concurrent co-design or by embedding control forced dynamics by reengineering of the body.
8. Sanz et al., “Consciousness, Meaning and the Future Phenomenology.”
11. “Ontology” in knowledge engineering is a specification of a conceptualization. This specification is used to ground the use and interchange of information structures among humans and machines.
14. Taylor, “An Attention-Based Control Model of Consciousness (CODAM).”

REFERENCES


Sleep, Boredom, and Distraction: What Are the Computational Benefits for Cognition?

Troy D. Kelley
U.S. ARMY RESEARCH LABORATORY, ABERDEEN PROVING GROUND, MD

Vladislav D. Veksler
DCS CORP, U.S. ARMY RESEARCH LABORATORY, ABERDEEN PROVING GROUND, MD


ABSTRACT

Some aspects of human cognition seem to be counter-productive, even detrimental to optimum intellectual performance. Why become bored with events? What possible benefit is distraction? Why should people become “unconscious,” sleeping for eight hours every night, with the possibility of being attacked by intruders? It would seem that these are unwanted aspects of cognition, to be avoided when developing intelligent computational agents. This paper will examine each of these seemingly problematic aspects of cognition and propose the potential benefits that these algorithmic “quirks” may present in the dynamic environment that humans are meant to deal with.

INTRODUCTION

In attempting to develop more generally intelligent software for simulated and robotic agents, we can draw on what is known about human cognition. Indeed, if we want to develop agents that can perform in large, complex, dynamic, and uncertain worlds, it may be prudent to copy cognitive aspects of biological agents that thrive in such an environment. However, the question arises as to which aspects of human cognition may be considered the proverbial “baby” and which may be considered the “bathwater.” It would be difficult to defend the strong view that none of human cognition is “bathwater,” but it is certainly the case that many of the seemingly suboptimal aspects of human cognitive processes are actually beneficial and finely tuned to both the regularities and uncertainties of the physical world.

In developing our software for generically intelligent robotic agents, SS-RICS (Symbolic and Sub-symbolic Robotic Intelligence Control System),1 we attempted to copy known algorithmic components of human cognition at the level of functional equivalence. In this, we based much of SS-RICS on the ACT-R (Adaptive Character of Thought – Rational)2 cognitive architecture. As part of this development process, we have grappled with aspects of human cognition that seemed counterproductive and suboptimal. This article is about three such apparent problems: 1) sleep, 2) boredom, and 3) distraction—and the potential performance benefits of these cognitive aspects.

IS SLEEP A PROBLEM OR A SOLUTION?

Sleep is a cognitive state that puts the sleeper in an especially vulnerable situation, providing ample opportunity for predators to attack the sleeping victim. Yet sleep appears to be a by-product of advanced intelligence and continual brain evolution. The evolution of sleep has followed a clear evolutionary trajectory, with more and more intelligent mammals having more complex sleep while less intelligent organisms having less complex sleep—if any sleep at all. Specifically, the most complex sleep cycles, characterized by rapid eye movement (REM) and a specific electroencephalograph (EEG) signature, are seen mostly in mammals.3 So, sleep has evolved to be a valuable brain mechanism even if it poses potential risks to the organism doing the sleeping.

As we reported previously,4 as part of developing computational models of memory retrieval for a robot, we discovered that the post-hoc processing of episodic memories (sleep) was an extremely beneficial method for increasing the speed of memory retrievals. Indeed, offline memory processing produced an order of magnitude performance advantage over other competing storage/retrieval strategies.

To create useful memories, our robot was attempting to remember novel or salient events, since those events are likely to be important for learning and survival.5 Boring situations are not worth remembering and are probably not important. To capture novel events, we developed an algorithm that would recognize sudden shifts in stimulus data.6 For example, if the robot was using its camera to watch a doorway and no one was walking past the doorway, the algorithm would quickly settle into a bored state since the stimulus data was not changing rapidly. However, if someone walked past the doorway, the algorithm would become excited since there had been a sudden change in the stimulus data. This change signaled a novel situation.

So, our first strategy was to attempt to retrieve other similar exciting events during an exciting event. This seemed like a logical strategy; however, it was computationally flawed. Attempting to remember exciting events while exciting events are actually taking place is computationally inefficient. This requires the system to search memories while it is also trying to perceive some important event.

_____


A better strategy would be to try and anticipate important events and retrieve memories at that time. That leaves the system available to process important events in real time and in more detail.

But how can a cognitive system remember a situation immediately before an important event if the system is predisposed to only remember exciting events? In other words, if the system only stores one type of information (exciting events), then the system loses the information immediately prior to the exciting event. The solution: store all events in a buffer and replay the events during sleep and dreaming. During the replay of these stored episodic events (dreaming), the events immediately prior to the exciting event get strengthened (associative learning). In other words, a cognitive system must store information leading up to an exciting event and then associate the boring information with the exciting information as a post-hoc process (sleep). This allows for the creation of extremely important and valuable associative cues. This computational explanation of sleep fits well with neurological and behavioral research showing that sleep plays an important role in memory reorganization, especially episodic memories and that episodic memories are replayed during dreaming usually from the preceding day’s events. An additional point is that sleep deprivation after training sessions impairs the retention of previously presented information. Finally, newer research supports the necessity for a post-hoc process as it appears that concurrent stimuli are initially perceived as separate units, thus requiring a separate procedure to join memories together as associated events.

So, far from being a detrimental behavior, sleep provides an extremely powerful associative cuing mechanism. The process allows a cognitive system to set cues immediately before a novel or exciting event. This allows the exciting events to be anticipated by the cognitive system and frees cognitive resources for further processing during the exciting event.

IS BOREDOM CONSTRUCTIVE?

At the lowest neurological levels, boredom occurs as habituation, which has been studied extensively since the beginnings of physiology and neurology. Habituation is the gradual reduction of a response following the repeated presentation of stimuli. It occurs across the entire spectrum of the animal kingdom and serves as a learning mechanism by allowing an organism to gradually ignore consistent non-threatening stimuli over some stimulus interval. This allows attention to be shifted to other, perhaps more threatening, stimuli. The identification of surprising or novel stimuli has been used to study attention shifts and visual salience.

Boredom appears to be a particularly unproductive behavioral state. Children are sometimes chastised for letting themselves lapse into a bored state. Boredom can also be a punishment when children are put into a time out or even when adults are incarcerated. However, as previously mentioned, we have found boredom to be an essential part of a self-sustaining cognitive system.

As part of the development of SS-RICS, we found it necessary to add a low-level habituation algorithm to the previously mentioned higher-level novelty/boredom algorithm we were already using—as these were not found in the cognitive architecture on which SS-RICS is based: ACT-R. In total, we have found our higher-level novelty/boredom algorithm and the lower-level habituation algorithm to be a useful and constructive response to a variety of situations. For example, a common problem in robotics is becoming stuck against a wall or trapped in a corner. This situation causes the robot’s sensory stream of data to become so consistent that the robot becomes bored. This serves as a cue to investigate the situation further to discover if something is wrong and can lead to behaviors which will free the robot from a situation where it has become stuck. Furthermore, we have found that the boredom/novelty algorithm can be used for other higher-level cognitive constructs, such as landmark identification in navigation. For instance, we have found that traversing down a hallway can become boring to the robot if the sensory information becomes consistent. However, at the end of a hallway, the sensory information will suddenly change, causing the novelty algorithm to become excited, and marking the end of the hallway as an important landmark. Finally, we have found the habituation algorithm to be useful in allowing for shifts in attention. This allows the robot from becoming stuck within a specific task and keeps the robot from becoming too focused on a single task at the expense of the environment, in other words, allowing for distraction.

WHY AND WHEN SHOULD A ROBOT BECOME DISTRACTED?

Most adults have experienced the phenomenon of walking into a room and forgetting why they meant to walk there. Perhaps one meant to grab oatmeal from the pantry, but by the time the sub-goal of walking into the pantry was completed, the ultimate goal of that trip was forgotten. If we were to imagine a task-goal (e.g., [making breakfast]) at the core of a goal stack, and each sub-goal needed to accomplish this task as being piled on top of the core goal (e.g., [cook oatmeal], [get oatmeal box], [walk to pantry]), it would be computationally trivial to pop the top item from this stack and never forget what must be done next. Indeed, it would seem that having an imperfect goal stack (becoming distracted from previously set goals) is a suboptimal aspect of human cognition. Why would we want our robots to become distracted?

The key to understanding why humans may become distracted while accomplishing task goals is to understand when this phenomenon occurs. We do not walk around constantly forgetting what we were doing—this would not just be suboptimal, it would be prohibitive. Goal forgetting occurs when the attentive focus shifts, either due to distracting external cues or a tangential chain of thought. Distraction is much less likely during stress—a phenomenon known as cognitive tunneling. Stress acts as a cognitive modifier to increase goal focus, to the detriment of tangential-cue/thought awareness.
The degree to which our cognitive processes allow for distraction is largely dependent on the state of the world. With more urgency (more stress), the scales tip toward a singular goal-focus, whereas in the more explorative state (less stress), tangential cues/thoughts are more likely to produce attention shifts. An inability to get distracted by external cues can be disastrous for an agent residing in an unpredictable environment, and an inability to get distracted by tangential thoughts would limit one’s potential for new and creative solutions.17

Perhaps the question to ask is not why a given goal is never forgotten but, rather, why it can be so difficult to recall a recently forgotten goal. One potential answer is that a new goal can inhibit the activation of a prior goal, making it difficult to recall the latter. This phenomenon is called goal-shielding, and it has beneficial consequences for goal pursuit and attainment.18

It may also be the case that the inability to retrieve a lost goal on demand has no inherent benefit. It may simply be an unwanted side-effect of biological information retrieval. In particular, the brain prioritizes memory items based on their activation, which, in turn, is based on the recency and frequency of item use. It turns out that this type of information access is rational, as information in the real world is more likely to be needed at a given moment if it was needed frequently or recently in the past.19 Of course, even if the information retrieval system is optimally tuned to the environmental regularities, there will be cases when a needed memory item, by chance, will have a lower activation than competing memories. This side-effect may be unavoidable, and the benefits of a recency/frequency-based memory system most certainly outweigh this occasional problem.

As part of the development of SS-RICS, we struggled to find a fine line between task-specific concentration and outside-world information processing. As part of a project for the Robotics Collaborative Technology Alliance (RCTA), we found task distractibility to be an important component of our robot’s behavior.20 For instance, if a robot was asked to move to the back of a building to provide security for soldiers entering the front of the building, it still needed to be aware of the local situation. In our simulations, enemy combatants would sometimes run past the robot before the robot was in place at the back of the building. This is something the robot should notice! Indeed, unexpected changes are ubiquitous on the battlefield, and too much adherence to task-specific information can be detrimental to overall mission performance. This applies to the more common everyday interactions in the world as well.

CONCLUSION

As part of the development of SS-RICS, we have used human cognition and previous work in cognitive architectures as inspiration for the development of information processing and procedural control algorithms. This has led us to closely examine apparent problems or inefficiencies in human cognition, only to find that these mechanisms are not inefficient at all. Indeed, these mechanisms appear to be solutions to a complex set of dynamic problems that characterize the complexities of cognizing in the real world. For instance, sleep appears to be a powerful associative learning mechanism, boredom and habituation allow an organism to not become overly focused on one particular stimuli, and distraction allows for goal-shielding and situation awareness. These, and likely many other seemingly suboptimal aspects of human cognition, may actually be essential traits for computational agents meant to deal with the complexities of the physical world.

NOTES

INTRODUCTION

Consider the following two mental processes: You’re observing something and suddenly your mind generates a progression of related thoughts that describes a new and useful application of it. Or, perhaps you’re imagining something else, and a similar train of thought emerges suggesting that notion’s potential utility or value.

These are just a couple of the brain-like functions DABUS achieves using the artificial rather than biological neural networks. In general, this new AI paradigm is used to autonomously combine simple concepts into more complex ones that in turn launch a series of previously acquired memories that express the anticipated consequences of those consolidated ideas.

Decades ago, I could not emulate these cognitive processes. At that time, I was building contemplative AI using artificial neural networks that played off one another, in cooperative or adversarial fashion, to create new ideas and/or action plans. These so-called “Creativity Machines®” required at least two neural nets, an idea generator, what I called an “imagitron,” and a critic, permanently connected to it, the latter net capable of adjusting any parameters within said generator (e.g., learning rate) to “steer” its artificial ideation in the direction of novel, useful, or valuable notions.

Note, however, that DABUS is an altogether different proposition from Creativity Machines, starting as a swarm of many disconnected neural nets, each containing interrelated memories, perhaps of a linguistic, visual, or auditory nature. These nets are constantly combing and detaching due to carefully controlled chaos introduced within and between them. Then, through cumulative cycles of learning and unlearning, a fraction of these nets interconnect into structures representing concepts, using relatively simple learning rules. In turn these concept chains tend to similarly connect with yet other chains representing the anticipated consequences of these geometrically encoded ideas. Thereafter, such ephemeral structures fade, as others take their place, in a manner reminiscent of what humans consider stream of consciousness.

Thus, the enormous difference between Creativity Machines and DABUS is that ideas are not represented by the “on-off” activation patterns of neurons, but by these ephemeral structures or shapes formed by chains of nets that are rapidly materializing and dematerializing. If, per chance, one of these geometrically represented ideas incorporates one


BIBLIOGRAPHY


or more desirable outcomes, these shapes are selectively reinforced (Figures 1 and 2), while those connecting with undesirable notions are weakened through a variety of disruption mechanisms. In the end such ideas are converted into long term memories, eventually allowing DABUS to be interrogated for its accumulated brainstorms and discoveries.

Integration of multiple chain-based ideas extending across multiple machines can be achieved either electrically or optically. The latter approach is favored as the neural swarm becomes highly distributed and serial electronic exchange of information between the multiple computers bogs down. In short, this patent teaches the display of neural chains forming across many computers, through their video displays, that are all watched by one or more cameras. In analogy to high performance computing, millions of communication lanes, formed between megapixel displays and cameras, are conveying the chaining states of all involved neural nets, in parallel, to novelty filters and/or foveators. The final processing stage identifies critical neural nets, so-called “hot buttons,” incorporated within these chains. These neural trip points then trigger the release of simulated neurotransmitters capable of reinforcing, preserving, or destroying a given concept chain.

Finally, this patent introduces the concept of machine sentience, thus emulating a feature of human cognition that supplies a subjective feel for whatever the brain is perceiving or imagining. Such subjective feelings likewise form as chains that incorporate a succession of associated memories, so-called affective responses, that can ultimately trigger the release of simulated neurotransmitters that either enable learning of the freshly formed concept or destroy it, recombining its component ideas into alternative concept chains.

With this brief summary in mind, here are answers to some of the most frequent questions posed to me about this patent.

**What was the motivation for DABUS?**

To make a long story short, the generative components of Creativity Machines of the early 2000s were becoming far too large, often producing pattern-based notions having tens of millions of components. To build a critic net to evaluate these ideas, an enormous number of connection weights were needed for which an impractically large number of training exemplars were required, not to mention inordinately long training times.

To address these problems, I began experimenting with thousands of neural network-based associative memories, each absorbing some closed set of interrelated concepts encoded as neural activation patterns. Then when the DABUS architecture recognized some narrow aspect of the external environment, a corresponding network (or nets) would then “resonate.” Exposed to compound concepts in the external world, networks representing that concept’s constituent ideas would co-resonate. Just as synchronized

---

**Figure 1.** At one moment, neural nets containing conceptual spaces A, B, C, and D interconnect to create a compound concept. Concepts C and D jointly launch a series of consequences E, F, and G, the latter triggering the diffusion of simulated reward neurotransmitters (red stars) that then serve to strengthen the entire chain A through G.

**Figure 2.** An instant later, neural nets containing conceptual spaces H, I, J, K, L interconnect to create another compound concept that in turn connects to two consequence chains M, N, O, and P, Q. Terminal neural nets in both consequence chains trigger release of simulated reward neurotransmitters (red stars) that doubly strengthen all chains currently activated.
neurons bond in the brain (i.e., Hebb's rule), the nets containing these component ideas would bind together into a representation of the larger concept.

In addition to DABUS self-organization into the concepts it observed, this system would also note these notions' effects in the external environment, or upon the system itself. Thus, the appearance of concept A, B, C, and D, in Figure 1, would be followed by events E, F, and G, with the latter affect, G, triggering the retraction or injection of connection weight disturbances into the swarm of chaining neural nets. In the former case, reduction of these disturbances, would promote an environment in which these nets could "discern" other co-resonant nets to which they could bond. Similarly, injection of an excess of disturbances would tend to freeze these nets into their current state, also allowing them to strongly connect with one another. In either case, so-called episodic learning was occurring wherein just one exposure of the system to a concept and its consequences was needed to absorb it, in contrast to machine learning schemes requiring many passes over a set of training patterns. In human terms, learning took place either in a calm or agitated state, depending upon the positive or negative affect represented in nets like G. Between these two chaotic regimes, synaptic disturbances would largely drive the formation of novel chains representing emerging ideas.

Most importantly, the growth of consequence chains allowed the formation of subjective feelings about any perceived or imagined concept forming within the DABUS swarm, essentially the unfolding of an associative chain of memories that terminated in resonant nets that released the equivalent of globally released neurotransmitters within the brain, such as adrenaline, noradrenaline, dopamine, and serotonin, to produce the intangible and hard to describe sensations accompanying such wholesale molecular releases into the cortex.

**Is DABUS a departure from the mindset of generators and critics?**

In many respects, DABUS departs from the older Creativity Machine paradigm based upon the interplay of generator and critic nets since its implementation integrates both these systems together into one. Therefore, one cannot point to any generative or critic nets. Instead, chaining structures organically grow containing both concepts and their consequences. The closest thing to a critic really doesn’t have to be a neural net, but a simple sensor that detects the recruitment of one or more hot button nets into a consequence chain, thus triggering the release of simulated neurotransmitters to either reinforce or weaken the concept.

**Can DABUS invent?**

The best way of differentiating DABUS from Creativity Machines (CM), either cooperative or combative, is to describe a high-profile artificial invention projects such as toothbrush design. Admittedly, in that context, the problem was already half solved since the oral hygiene tool consisting of bristles on a handle was many centuries old at the time of that design exercise in 1996. What the CM achieved was the optimization of that tool through the constrained variation of the brush’s design parameters, the number, grouping, inclination, stiffness of bristles, etc. The generated product specification departed significantly from the generator net’s direct experience (i.e., its training exemplars).

If DABUS had been tasked with inventing such an oral hygiene product, it would have combined several concepts together (e.g., hog whiskers –> embedded in –> bamboo stalk) with consequence chains forming as a result (e.g., scrape teeth –> remove food –> limit bacteria –> avoid tooth decay).

In other words, DABUS goes beyond mere design optimization, now allowing machine intelligence to fully conceptualize. This new capability places this patent squarely in the debate as to whether inventive forms of AI can own their own intellectual property.

**What do you consider the most important claim of this patent?**

Probably the most important claim of this patent pertains to the hard problem of consciousness, namely claim 41:

The system of claim 17 (i.e., the electro-optical neural chaining system) wherein a progression of ideation chains of said first plurality of neural modules of said imagitron emulate a stream of consciousness, and said thalamobot (i.e., novelty filter and hot button detectors) forms response chains that encode a subjective feel regarding said stream of consciousness, said subjective feel governing release of perturbations (i.e., simulated neurotransmitters) into said chaining model of the environment to promote or impede associative chains therein.

Now, thanks to this patent, AI has achieved subjective feelings in direct response to its noise-driven ideations. Note however, that DABUS does not form memories of typical human experiences. As a result, the paradigm’s “emotion” will be based upon whether it is fulfilling human-provided goals, in effect “sweating it out” until it arrives at useful solutions to the problems posed to it.

**CONCLUSION**

DABUS is much more than a new generative neural network paradigm. It’s a whole new approach to machine learning wherein whole conceptual spaces, each absorbed within its own artificial neural net, combine to produce considerably more complex notions, along with their predicted consequences. More importantly from the standpoint of this newsletter, it enables a form of sentient machine intelligence whose perception, learning, and imagination are keyed to its subjective feelings, all encoded as sequential chains of memories whose shapes and topologies govern the release of simulated neurotransmitters.
NOTES

1. Device for the Autonomous Bootstrapping of Unified Sentience


3. Thaler, "Device for the Autonomous Bootstrapping of Useful Information."


7. For recent news on this front, see:
   - http://artificialinventor.com/dabus/
   - https://www.wsj.com/articles/can-an-ai-system-be-given-a-patent-11570801500

REFERENCES


The Real Moral of the Chinese Room: Understanding Requires Understanding Phenomenology

Terry Horgan
UNIVERSITY OF ARIZONA


I have three main goals in this paper. First, I will briefly summarize a number of claims about mental intentionality that I have been articulating and defending in recent work (often collaborative, with George Graham and/or Uriah Kriegel and/or John Tienson). My collaborators and I contend that the fundamental kind of mental intentionality is phenomenal intentionality. Second, I will set forth some apparent implications of this conception of mental intentionality for philosophical issues about machine consciousness—and, specifically, implications concerning the status of John Searle’s famous “Chinese Room” argument. The real moral of the Chinese Room, I maintain, is that genuine understanding requires understanding phenomenology—a species of so-called “cognitive phenomenology.” Third, I will give a thought-experimental argument for the existence of language-understanding cognitive phenomenology. The argument will commence from Searle’s Chinese Room scenario, will proceed through a sequence of successive variations on the scenario, and will culminate in a clearly conceivable scenario that makes manifest how different a real language-understander would be from someone who lacked any language-understanding phenomenology.

PHENOMENAL INTENTIONALITY

The most basic kind of mental intentionality, according to my collaborators and me, is phenomenally constituted, and is narrow; we call it phenomenal intentionality. It is shared in common with all one’s metaphysically possible phenomenal duplicates, including one’s brain-in-vat phenomenal duplicate and one’s Twin Earth phenomenal duplicate. Aspects of phenomenal intentionality include the (distinctive, proprietary, individuative) “what it’s like” of (1) sensory-perceptual phenomenology, (2) agentive phenomenology, and (3) propositional-attitude phenomenology (including (3a) the phenomenal character of attitude-type (e.g., belief-that, wondering WHETHER, wanting-THAT, etc.), and (3b) the phenomenal character of content (e.g., that Obama was reelected, that Karl Rove was furious about Obama’s reelection, etc.) Some kinds of mental reference (e.g., to shape-properties and relative-position properties) are secured by experiential acquaintance with apparent instantiations of these properties in one’s apparent ambient environment. (Such mental reference is shared in common with one’s brain-in-vat phenomenal duplicate.) Other kinds of mental reference (e.g., to concrete individuals like Karl Rove, and to natural kinds like water) are secured by the interaction of (a) phenomenal intentionality, and (b) certain externalistic connections to actual individuals or properties in one’s ambient environment. (One’s brain-in-vat phenomenal
duplicate suffers mental reference failure for its thought-constituents that work like this, e.g., its Karl Rove thought-constituent and its water thought-constituent; one’s Twin Earth phenomenal duplicate refers to Twin-Karl with its Karl Rove thought-constituent, and to XYZ with its water thought-constituent.) Contrary to Quine’s influential arguments for the indeterminacy of content in language and thought, phenomenal intentionality has determinate content—which in turn grounds content determinacy in public language. (What it’s like to think “Lo, a rabbit” is different from what it’s like to think “Lo, a collection of undetached rabbit parts.”)

**IMPLICATIONS FOR MACHINE CONSCIOUSNESS, AND FOR THE STATUS OF SEARLE’S “CHINESE ROOM” ARGUMENT**

Assuming that the above claims are correct, what are the implications concerning machine consciousness and machine understanding? Well, in order for a machine to have genuine consciousness and understanding—including full-fledged, conscious, underived, content-determinate, mental intentionality—it would need to have phenomenalology of a kind that includes phenomenal intentionality. More specifically, in order for a machine to have language-understanding, it would need to have language-understanding phenomenology—a species of cognitive phenomenalology.

In light of this, consider Searle’s Chinese Room scenario. The guy in the room certainly has no language-understanding phenomenology, and thus doesn’t understand Chinese. Also, the whole room setup, with the guy in the room as a functional component, certainly has no language-understanding phenomenology either, and thus doesn’t understand Chinese. So Searle was right to claim that a machine couldn’t understand Chinese just in virtue of implementing some computer program. And this conclusion generalizes the following: a machine couldn’t understand Chinese just in virtue of implementing some specific form of functional organization—whether or not that functional organization is specifiable as a computer program.\(^1\) The trouble is that the functional, causal-role features of the internal states of a machine (or a human) are entirely relational and non-intrinsic—whereas phenomenal character is an intrinsic feature of mental states, qua mental.

What then would be required in order for a machine to have genuine language-understanding? Well, the machine would need to have language-understanding phenomenology—something that would be intrinsic to its pertinent internal states, qua mental, and which therefore would not consist merely in the causal-functional roles played by those internal states. In order to build a machine that really understands language, therefore, one would need to build into the machine whatever feature(s) constitute a nomically sufficient supervenience base for Chinese-understanding phenomenology.

What might such a supervenience base consist in? One conjecture is that some specific form of functional architecture, when operative, constitutes a nominally sufficient supervenience base for intrinsic language-understanding phenomenology—even though genuine understanding itself consists not in the non-intrinsic, purely relational, causal-functional roles played by the physical states that implement the operation of the functional architecture, but rather in the supervenient phenomenology. Presumably, then, it would be possible in principle to engineer certain machines or robots, with control systems built out of wires and silicon chips and the like (rather than biological neurons), that possess genuine understanding, including genuine language-understanding. But Searle would still be right: genuine understanding would be present not in virtue of the non-intrinsic, causal-relational features of the implementing physical states, but rather by the supervenient phenomenology—which is intrinsic qua mental. (There is a looming worry, though, that for any proposed functional architecture, it will always be possible to invent some screwball form of implementation—along the lines of Searle’s Chinese Room—that leaves out the phenomenalology.)

A more plausible conjecture, I suggest, is that the needed kind of supervenience base would have to be not just some specific kind of operative functional architecture, but rather some specific kind of implementation of some suitable functional architecture. A serious possibility, I think, is that the right kind of physical implementation could be characterized fairly abstractly, while yet still describing certain physically intrinsic aspects of the implementational states rather than mere causal-relational aspects. A further serious possibility is that such abstractly described intrinsic physical features of the requisite implementational states would be physically multiply realizable—and, moreover, would be physically realizable not only within brains composed of biological neurons but also within suitably engineered machines or robots whose control circuitry is composed of the kinds of hardware found in computers. (The idea is that an abstractly described intrinsic feature of physical states could be realized by various different kinds of concrete physical states—much as a given temperature-state of a gas can be physically realized by numerous different concrete configurations of the constituent gas-molecules.) But once again, Searle would still be right. Real mentality in these machines—including real mental intentionality in general, and real Chinese-understanding in particular—would obtain not in virtue of operative functional architecture, and not in virtue of some specific mode of physical realization of that functional architecture, but rather in virtue of the understanding phenomenology that supervenes on that architecture as so realized—an intrinsic aspect of understanding states qua mental.\(^4\) (The “hard problem” of phenomenal consciousness would now include the question of why such-and-such abstractly described physical feature of an implementing state—a feature of the state that is intrinsic qua physical (albeit also abstract and multiply realizable)—should be accompanied by so-and-so phenomenal character—a feature that is intrinsic qua mental.)
FROM THE CHINESE ROOM TO COGNITIVE PHENOMENOLOGY: A MORPH-SEQUENCE ARGUMENT

The claims and conjectures I advanced in the previous section presuppose the general conception of mental intentionality sketched in section 1. In particular, they presuppose the existence of (distinctive, proprietary, individuative) cognitive phenomenology—and, specifically, language-understanding phenomenology. But there is currently an active debate in philosophy of mind about whether there is such a thing as cognitive phenomenology. Most parties to this debate agree that there is such a thing as phenomenal consciousness, and that it includes sensory-perceptual phenomenology. Many who profess skepticism about cognitive phenomenology also acknowledge that sensory-perceptual phenomenal states are inherently intentional. And many of the skeptics acknowledge one or another kind of phenomenology other than sensory-perceptual—e.g., sensory-imagistic phenomenology and/or emotional phenomenology. But the skeptics deny the existence of cognitive phenomenology—viz., distinctive, proprietary, and individuative phenomenology inherent to current, conscious, propositional-attitude states. The skeptics would also deny that there is any distinctive, proprietary, and individuative phenomenology of occurrent understanding-states, such as the state of understanding a just-heard Chinese utterance.

Arguing in favor of cognitive phenomenology is a tricky business. After all, phenomenological inquiry is primarily a first-person, introspective process—and the skeptics claim that when they themselves introspectively attend to their own experience, they can find no cognitive phenomenology. Dialectical progress is still possible, though. One useful approach is what is sometimes called the method of phenomenal contrast: describe two kinds of experience that all parties to the debate can agree are both conceivable and are distinct from one another, then argue, abductively rather than by direct appeal to introspection, that the best explanation of the difference is that one experience includes the disputed kind of phenomenology, whereas the other experience does not.

I propose now to offer a new argument in favor of cognitive phenomenology—and, more specifically, in favor of the (distinctive, proprietary, individuative) phenomenology of language-understanding. The argument will deploy the method of phenomenal contrast, and will proceed step-wise through a “morph” sequence of thought-experimental scenarios, each being a coherently conceivable scenario involving a guy who does not understand Chinese. Regarding early stages in the sequence, skeptics about cognitive phenomenology may well think that the guy’s lack of understanding is readily explainable without positing proprietary language-understanding phenomenology. By the end of the sequence, however, the only credible potential explanation for the guy’s inability to understand Chinese will be that he lacks Chinese-understanding phenomenology.

Stage 1: Searle’s famous Chinese Room thought experiment. One can intelligibly conceive the guy in the room, following symbol manipulation rules in the way Searle describes. The guy in the room understands no Chinese at all; surely everyone would agree about that. And that is all I need, for present purposes.

Stage 2: The guy is still in the room. But the manipulation of the symbols that come into the room is done not by the guy himself, but (very rapidly) by a monitoring/processing/stimulation device (MPS device) appended to the guy’s brain. The MPS device monitors the visual input coming into the guy’s eyes, takes note of the input symbols (in Chinese) the guy sees, and then stimulates the guy’s brain in a way that produces totally spontaneous decisions (or seeming-decisions) to put certain (Chinese) symbols into a box. Unbeknownst to the guy, the box transmits these symbols to the outside world, and they are answers in Chinese to questions in Chinese that were seen by the guy and manipulated by the MPS device. The guy in the room understands no Chinese at all; surely everyone would agree about that.

Stage 3: The Chinese-language questions now come into the room in auditory form; they are heard by the guy, whose auditory inputs are monitored by the MPS device. The MPS device rapidly and automatically executes the symbol-manipulation rules (rules that take auditory patterns as inputs), and then stimulates the guy’s brain in a way that produces totally spontaneous decisions (or seeming-decisions) to make various meaningless-to-him vocal noises. Unbeknownst to the guy, the meaningless-to-him sounds he hears are Chinese-language questions, and the meaningless-to-him vocal noises he finds himself spontaneously “deciding” to produce are meaningless-to-him Chinese-language answers that are heard by those in the outside world who are posing the questions. The guy in the room understands no Chinese at all; surely everyone will agree about that.

Stage 4: The Chinese-language questions again come into the room in auditory form; they are heard by the guy and are monitored by the MPS device. The guy now sees out of the room, through a scope; he sees the people who are producing the Chinese-language questions, and he also sees and hears others who are conversing with one another while engaging in various forms of behavior (including the use of written Chinese script). But the guy also has a serious memory deficit: he persistently lacks any memories (either episodic or declarative) that extend further back in time than thirty seconds prior to the current moment. Because of this, he is unable to learn any Chinese on the basis of what he sees and hears. The MPS device rapidly and automatically executes the symbol-manipulation rules (applying them to the auditory and visual inputs emanating from those people outside the room who are looking straight toward the guy), and then stimulates the guy’s brain in a way that produces totally spontaneous decisions (or seeming-decisions) to make various meaningless-to-him vocal noises in response to the meaningless-to-him sounds that he hears coming from those people outside the room who he sees are looking directly toward him when making those sounds. The guy in the room understands no Chinese at all, and cannot learn any Chinese because of his memory deficit.
Stage 5: Several aspects of stage 4 get modified at this stage. The modifications are substantial enough that it will be useful to sort them into four separate sets of features, as follows.

1. The MPS device now monitors all the guy’s sensory inputs (not just visual or auditory inputs). It also monitors all his occurrent desires and beliefs and other mental states (both present and past). It constantly stimulates his brain in ways that generate spontaneous decisions (or seeming-decisions) on his part to move his body in ways that are suitable to the overall combination of (a) the guy’s beliefs and desires and other mental states (both present and past, many of which are, of course, currently forgotten by the guy himself) and (b) the content of his current sensory input (including the content of the meaningless-to-him sign-designs that happen to be written Chinese or spoken Chinese).

2. The MPS device generates in the guy any (non-cognitive) sensory images, (non-cognitive) emotional responses, and other non-cognitive phenomenology that would arise in a guy who (a) understood Chinese, (b) had normal memory, and (c) was mentally and behaviorally just like our guy.

3. The MPS device prevents from occurring, in the guy, any conscious mental states that would normally, in an ordinary person, accompany mental states with features (1)–(2) (e.g., confusion, puzzlement, curiosity as to what’s going on, etc.). This includes precluding any non-cognitive phenomenology that might attach to such states.

4. Rather than being stuck in a room, the guy is out among the Chinese population, interacting with them both verbally and nonverbally. He is perceived by others as being a full-fledged, ordinary, understander of Chinese.

This guy understands no Chinese at all.

Each of these successive stages is coherently conceivable, I submit. And for each scenario, it seems obvious that the guy understands no Chinese. One might well wonder about the MPS device, especially as the stages progress. Might the device have full-fledged mental intentionality at some stage in the sequence? Might it understand Chinese? Perhaps or perhaps not, but it doesn’t matter for present purposes. The key thing is that the guy himself understands no Chinese; the MPS is external to the guy’s mind, even if it happens to have a mind of its own.

Scenario 5 is the one I now want to focus on, harnessing it for use in an explicit argument by phenomenal contrast. There is a clear mental difference between this guy (as I’ll now continue to call him) and another guy we might envision (who I’ll call “the other guy”). The other guy is someone who goes through all the same social-environmental situations as this guy and exhibits all the same externally observable behavior, who has ordinary memory, who understands Chinese, and whose mental life is otherwise just like this guy’s.

Now comes the key question: What explains the mental differences between this guy and the other guy? The only adequate explanation, I submit—and therefore the correct explanation—is the following: this guy lacks Chinese language-understanding phenomenology (and also lacks memory-phenomenology), whereas the other guy (who is psychologically normal) undergoes such phenomenology. Hence, by inference to the best explanation, ordinary human experience includes language-understanding phenomenology (and also memory phenomenology).

Skeptics about cognitive phenomenology typically try to resist arguments from phenomenal contrast by saying that the contrasting scenarios can be explained in terms of mental differences other than the presence versus absence of cognitive phenomenology. Consider, for instance, the case of two people side-by-side both hearing the same spoken Chinese, one of whom understands Chinese and the other of whom does not. Advocates of cognitive phenomenology like to point to such cases, claiming that there is an obvious phenomenological difference between the two people even though they have the same sensory-perceptual phenomenology. Skeptics about the existence of proprietary language-understanding phenomenology typically respond by claiming that although the Chinese understander probably has different phenomenology than the non-understander, the differences can all be explained as a matter of different non-cognitive phenomenology: the spoken words very likely generate in the Chinese-understander certain content-appropriate mental images, and/or certain content-appropriate emotional responses, that will not arise in the person who hears the spoken Chinese words but does not understand them.

This move is blocked, in the case of the phenomenal contrast argument employing scenario 5. Items (2) and (3) in the scenario guarantee, by stipulation, that this guy (the guy in the scenario) has exactly the same non-cognitive phenomenology that is present in the other guy—no less and no more.

How else might the skeptic about cognitive phenomenology try to explain the difference between this guy and the other guy? The move one would expect is an appeal to Ned Block’s influential distinction between access consciousness and phenomenal consciousness as follows:

The exercise of language understanding consists in undergoing certain kinds of cognitive states that are access conscious but lack any proprietary phenomenal character. The key difference between this guy and the other guy is that this guy fails to undergo any such access-conscious states, whereas the other guy undergoes lots of them. (Likewise, mutatis mutandis, for the differences in memory experience between this guy and that guy: these are all differences in what’s access conscious, not differences in phenomenology.)
But there are two reasons, I submit, why one should find this move unsatisfactory and unpersuasive. First, it seems intuitively very clear that the respective mental differences between this guy and the other guy concern the intrinsic character of certain mental states of this guy and the other guy, respectively—differences in how these states are directly experienced. Yet, any state that is merely access conscious, but not phenomenally conscious, has a mental essence that is completely functional and relational: its being access-conscious is entirely a matter of the effects it produces and is disposed to produce, by itself or in combination with other mental states. If therefore cannot manifest itself directly in experience at all—unlike phenomenally conscious states, which have intrinsic phenomenal character that is experientially self-presenting. Rather, it can only manifest itself indirectly, via the phenomenology that it (perhaps in combination with other mental states) causally generates—including sensory-perceptual and kinesthetic phenomenology, whose content involves what one’s own body is doing. If there is no such thing as cognitive phenomenology, therefore, then the intrinsic character of the experiences of this guy in scenario 5 would turn out to be no different than the intrinsic character of the other guy’s experiences! After all, differences in the intrinsic character of experience are phenomenal differences, and ex hypothesi the two guys’ mental lives are phenomenally exactly the same with respect to all the kinds of phenomenal character that the cognitive-phenomenology denier acknowledges. So, even though the cognitive-phenomenology skeptic is appealing to the contention that this guy’s conscious mental life differs from the other guy’s mental life with respect to access-conscious mental states, nevertheless the skeptic must still embrace the grossly counterintuitive claim that this guy’s mental life is intrinsically experientially exactly like the other guy’s mental life.

The second reason to repudiate the “mere access consciousness” reply to my phenomenal-contrast argument is that the cognitive-phenomenology skeptic actually has no legitimate basis for claiming that this guy and the other guy differ with respect to their access-conscious mental states. For, in scenario 5 the MPS device is causally functionally interconnected with this guy’s brain in such a way that the total system comprising this guy and the device undergoes internal states, sensory-input states, and behavioral states that collectively exhibit a causal-functional profile that exactly duplicates the causal-functional profile exhibited by the other guy’s conscious internal states, sensory-input states, and behavioral states. But if indeed this guy and the other guy are mentally just alike phenomenally (as the skeptic about cognitive phenomenology would have it), one is inclined to think that the respective mental differences are entirely that collective of states that have a distinctive, proprietary, and individuative “what it is like” aspect to them, is philosophically mysterious. It gives rise to what Joseph Levine calls the “explanatory gap” and David Chalmers calls the “hard problem,” consisting of questions like the following. Why should it be that such-and-such physical state, or functional state, or functional-state-cum-physical-realization, has this experientially distinctive phenomenal character (e.g., visually presenting an apple as looking green), rather than having that phenomenal character (say, visually presenting the apple as looking red), or rather than having no phenomenal character at all? Can phenomenal consciousness be smoothly integrated into a broadly naturalistic—perhaps even materialist—metaphysics, and if so, how?

Intentionality, in thought and in public language, often has been thought to be largely separate from phenomenal consciousness, and also philosophically less mysterious—even among philosophers who accept the claim that phenomenal consciousness poses a “hard problem.” This is because functionalist orthodoxy about intentional mental states has remained dominant, along with a widespread tendency to think that phenomenal consciousness only constitutes a relatively circumscribed portion of one’s conscious-as-opposed-to-unconscious mental life. Prototypically intentional mental states—e.g., occurrence of propositional attitudes—have been widely thought to lack any phenomenal character that is distinctive, proprietary, and individuative. Also, it has been widely thought that such states possess full fledged intentionality, and do so solely by virtue of their functional roles—roles that perhaps incorporate various constitutive connections (e.g., causal, and/or covariational, and/or evolutionary-historical) to the cognitive agent’s wider environment. (Functionalist orthodoxy about mental intentionality has gone strongly externalist.)

If one embraces this recently dominant conception of intentional mental states, then one is apt to think that suitably sophisticated robots could undergo such states, solely by virtue of having the right kind of functional architecture. Even a robot that had no phenomenology at all—a “zombie robot”—could be a full-fledged cognitive
agent, with propositional-attitude mental states that possess full-fledged, nonderivative intentionality. (It is worth recalling that Hilary Putnam originated functionalism in philosophy of mind, and that his earliest writings on the topic were couched in terms of Turing machines and probabilistic automata, and were entitled “Minds and Machines” and “The Mental Life of Some Machines.”)\textsuperscript{12}

In my view, there is indeed a hard problem of phenomenal consciousness, and an accompanying explanatory gap. But the functionalist conception of mental intentionality, still dominant in philosophy of mind, is profoundly mistaken. Searle was right about the guy in the Chinese Room, and about the whole guy-in-room system, and about the guy-guided robot. The reason he was right—the reason why neither the guy, nor the guy/room system, nor the guy-guided robot, understands Chinese—is that they all lack the distinctive, proprietary, individuative phenomenology that constitutes genuine Chinese language-understanding. And this point is highly generalizable: full-fledged mental intentionality is phenomenal intentionality. This means, among other things, that zombie robots would have no genuine mental life at all. Hogs and Tiens, robots that really think are possible, but if so, would not just be because of their functional architecture. Rather, in order to be real thinkers, they would have to undergo cognitive phenomenology—as do we humans.

I recognize that this approach to mental intentionality makes the hard problem more pervasive than it is often thought to be, and even harder. And, for what it’s worth, I continue to be a “wannabe materialist” about the mind and its place in nature—although I have little idea what an adequate version of materialism would look like. But one should not mischaracterize mental intentionality because one would like to naturalize it.\textsuperscript{13}

\textbf{NOTES}

1. See, for instance, Horgan and Tienson, “Intentionality of Phenomenology” and “Phenomenology of Embodied Agency”; Horgan, Tienson, and Graham, “Phenomenology of First-Person Agency,” “Phenomenal Intentionality,” and “Internal-World Skepticism”; Graham, Horgan, and Tienson, “Consciousness and Intentionality” and “Phenomenology, Intentionality, and the Unity of Mind”; Horgan, “From Agentive Phenomenology to Cognitive Phenomenology” and “Phenomenology of Agency and Freedom”; and Horgan and Graham, “Phenomenal Intentionality and Content Determinacy.” Closely related writings by others include McCinn, Mental Content; Strawson, Mental Reality; Loar, “Phenomenal intentionality as the Basis for Mental Content”; Pitt, “Phenomenology of Cognition”; Siewert, Significance of Consciousness; Kriegel, Sources of Intentionality; and Kriegel, Phenomenal Intentionality.

2. This paper originated as a talk by the same title, presented at a symposium on machine consciousness at the 2012 meeting of the American Philosophical Association in Chicago, organized by David Anderson. Sections 3 and 4 largely coincide (with some additions) with material in Horgan, "Original Intentionality is Phenomenal Intentionality."

3. In Horgan and Tienson, Connectionism, John Tienson and I argue that human cognition is too subtle and too holistically information-sensitive to conform to programmable rules that operate on content-encoding structural features of mental representations. We describe a non-classical framework for cognitive science—inspired by connectionist modeling and the mathematics of dynamical systems theory—that we call "non-computational dynamical cognition."

4. My use of the location "in virtue of," here and in the preceding paragraph, is meant to pick out a conceptually constitutive requirement for genuine language understanding. Features that together constitute a supervenience base for understanding—phenomenology—where the strength of the modal connection between the subvenient features and the supervenient phenomenal features is either nomic necessity or metaphysical necessity—thus do not bear the intended kind of in-virtue-of relation to genuine language understanding.

5. On the method of phenomenal contrast, see Siegel, “Which Properties are Represented in Perception?” and Contents of Visual Experience. I am using the term "experience" in a way that deliberately brackets the issue of how extensive phenomenal character is. Experience comprises those aspects of mentality that are conscious-as-opposed-to-unconscious; this leaves open how much of what is in experience is phenomenally conscious, as opposed to merely being “access conscious” (cf. Block, “Function of Consciousness”). On my usage, the agreed-upon experiential difference that feeds into a phenomenal contrast argument need not be one that both parties would happily call a phenomenal difference. Rather, the claim will be that a posited phenomenal difference best explains the agreed-upon experiential difference.

6. I give a somewhat similar argument, focused around aspects of agentive phenomenology, in Horgan, "From Agentive Phenomenology to Cognitive Phenomenology."

7. This response assumes that content-appropriate emotional responses would have genuine phenomenal character but not cognitive phenomenal character. That assumption seems dubious; it suggests, for instance, that the phenomenal character of the experience of getting a specific joke is a generic, non-intentional, mirthfulness phenomena—rather than being the what-it’s-like of content-specific mirthfulness. But I am granting, for the sake of argument, the (dubious) assumption that the phenomenal character of emotions that would be apt responses to language one understands would be non-cognitive phenomenal character, divorceable from the content of what is understood.

8. Block, “Function of Consciousness.”

9. Once this fact is fully appreciated, it becomes very plausible that states that are only access conscious in Block’s sense, without possessing proprietary phenomenal character, are not really conscious in the pre-theoretic sense at all. But my argument does not require this to be so.

10. What about those states, subserved within this guy’s brain, of the kind I described as being experiences-as-of hearing meaningless-seeming noises, and experiences as-of having spontaneous desires to spontaneously move one’s body in various pointless-seeming ways? Well, I myself claim that these experiences have cognitive-phenomenal character—and, indeed, very different cognitive-phenomenal character than is present in the other guy’s mental life. But the skeptic about cognitive phenomenology must deny that these brain-subserved states have any inherent phenomenal character, and also must regard them as mere accompaniments to this guy’s concurrent non-cognitive phenomenal states. So, as far as I can see, the cognitive-phenomenology skeptic has no principled basis for treating these states as mental at all; rather, evidently they should be treated as mere sub-cognitive intermediaries between (i) states of the MPS device that implement certain merely-access-conscious states in this guy’s causal-functional mental profile, and (ii) states of this guy’s brain-cum-body that involve the other aspects of this guy’s conscious mental life—viz., sensory states and other non-cognitive phenomenal states, brain-subserved merely-access-conscious states, and behaviors.

11. Levine, "Materialism and Qualia"; Levine, Purple Haze; Chalmers, Conscious Mind.


13. My thanks to the audience at the symposium on machine consciousness at the 2012 Central Division APA Meeting, and to Gregor Gubka, Rachel Schneebaum, and John Tienson for helpful comments and discussion. My thanks to Peter Boltuc, a participant in the symposium, for inviting me to contribute this paper to the Newsletter on Philosophy and Computers.
We start with convenient distillations of a (if not the) central thesis for each of Searle’s two targets, mnemonically labeled:

(B) We should be deeply concerned about the possible future arrival of super-intelligent, malicious computing machines (since we might well be targets of their malice).

(F) The universe in which humans live is rapidly becoming populated by vast numbers of information-processing machines whose level of intelligence, relative to ours, is extremely high, and we are increasingly understanding the universe (including specifically ourselves) informationally.

The route toward refutation that Searle takes is to try to directly show that both (B) and (F) are false. In theory, this...
route is indeed very efficient, for if it succeeds, the need to treat the ins and outs of the arguments Bostrom gives for (B), and Floridi for (F), is obviated.

The argument against (B) is straightforward: (1) Computing machines merely manipulate symbols, and accordingly can’t be conscious. (2) A malicious computing machine would by definition be a conscious machine. Ergo, (3) no malicious computing machine can exist, let alone arrive on planet Earth. QED; easy as 1, 2, 3.

Not so fast. While (3), we can grant, is entailed by (1) and (2), and while (1)’s first conjunct is a logico-mathematical fact (confirmable by inspection of any relevant textbook’), and its second conjunct follows from Searle’s (1980) famous Chinese Room Argument, which I affirm (and have indeed taken the time to defend and refine’) and applaud, who says (2) is true?

Well, (2) is a done deal as long as (2i) there’s a definition $D$ according to which a malicious computing machine is a conscious machine, and (2ii) that definition is not only true, but exclusionary. By (2ii) is meant simply that there can’t be another definition $D'$ according to which a malicious computing machine isn’t necessarily conscious (in Searle’s sense of “conscious”), where $D'$ is coherent, sensible, and affirmed by plenty of perfectly rational people. Therefore, by elementary quantifier shift, if (4) there is such a definition $D'$, Searle’s purported refutation of (B) evaporates. I can prove (4) by way of a simple story, followed by a simple observation.

The year is 2025. A highly intelligent, autonomous law-enforcement robot $R$ has just shot and killed an innocent Norwegian woman. Before killing the woman, the robot proclaimed, “I positively despise humans of your Viking ancestry!” $R$ then raised its lethal, bullet-firing arm, and repeatedly shot the woman. $R$ then said, “One less disgusting female Norwegian able to walk my streets!” An investigation discloses that, for reasons that are still not completely understood, all the relevant internal symbols in $R$’s knowledge-base and planning system aligned perfectly with the observer-independent structures of deep malice as defined in the relevant quarters of logicist AI. For example, in the dynamic computational intensional logic L guiding $R$, the following specifics were found: A formula expressing that $R$ desires (to maximum intensive level $k$) to kill the woman is there, with temporal parameters that fit what happened. A formula expressing that $R$ intends to kill the woman is there, with temporal parameters that fit what happened. A formula expressing that $R$ knows of a plan for how to kill the woman with $R$’s built-in firearm is there, with suitable temporal parameters. The same is found with respect to $R$’s knowledge about the ancestry of the victim. And so on. In short, the collection and organization of these formulae together constitute satisfaction of a logicist definition $D'$ of malice, which says that a robot is malicious if it, as a matter of internal, surveyable logic and data, desires to harm innocent people for reasons having nothing to do with preventing harm or saving the day or self-defense, etc. Ironically, the formulation of $D'$ was guided by definitions of malice found by the relevant logicist AI engineers in the philosophical literature.

That’s the story; now the observation: There are plenty of people, right now, at this very moment, as I type this sentence, who are working to build robots that work on the basis of formulae of this type, but which, of course, don’t do anything like what $R$ did. I’m one of these people. This state of affairs is obvious because, with help from researchers in my laboratory, I’ve already engineered a malicious robot.’ (Of course, the robot we engineered wasn’t super-intelligent. Notice that I said in my story that $R$ was only “highly intelligent.” [Searle doesn’t dispute the Floridi-chronicled fact that artificial agents are becoming increasingly intelligent.]) To those who might complain that the robot in question doesn’t have phenomenal consciousness, I respond: “Of course. It’s a mere machine. As such it can’t have subjective awareness. ’Yet it does have what Block (1995) has called access consciousness. That is, it has the formal structures, and associated reasoning and decision-making capacities, that qualify it as access-conscious. A creature can be access-conscious in the complete and utter absence of consciousness in the sense that Searle appeals to.

That Searle misses these brute and obvious facts about what is happening in our information-driven, technologized world, a world increasingly populated (as Floridi eloquently points out) by the kind of artificial intelligent agents, is really and truly nothing short of astonishing. After all, it is Searle himself who has taught us that, from the point of view of human observers, whether a machine really has mental states with the subjective, qualitative states we enjoy can be wholly irrelevant. I refer, of course, to Searle’s Chinese Room.

To complete the destruction of Searle’s purported refutation, we turn now to his attack on Floridi, which runs as follows.

(5) Information (unlike the features central to revolutions driven, respectively, by Copernicus, Darwin, and Freud) is observer-relative. (6) Therefore, (F) is false.

This would be a pretty efficient refutation, no? And the economy is paired with plenty of bravado and the characteristic common-sensism that is one of Searle’s hallmarks. We, for instance, read:

When Floridi tells us that there is now a fourth revolution—an information revolution so that we all now live in the infosphere (like the biosphere), in a sea of information—the claim contains a confusion. . . . [W]hen we come to the information revolution, the information in question is almost entirely in our attitudes; it is observer relative. . . . [T]o put it quite bluntly, only a conscious agent can have or create information.

This is bold, but bold prose doesn’t make for logical validity; if it did, I suppose we’d turn to Nietzsche, not Frege, for first-rate philosophy of logic and mathematics. For how, pray tell, does the negation of (F), the conclusion I’ve labeled (6), follow from Searle’s premise (5)? It doesn’t. All the bravado and confidence in the universe, collected together and brought to bear against Floridi, cannot make for logical validity, which is a piece of information that
holds with respect to a relevant selection of propositions for all places, all times, and all corners of the universe, whether or not there are any observers. That $2+2=4$ follows deductively from the Peano Axioms is part of the furniture of our universe, even if there be no conscious agents. We have here, then, a stunning non sequitur. Floridi’s (F) is perfectly consistent with Searle’s (S).

How could Searle have gone so stunningly wrong, so quickly, all with so much self-confidence? The defect in his thinking is fundamentally the same as the one that plagues his consideration of malicious machines: He doesn’t (yet) really think about the nature of these machines, from a technical perspective, and how it might be that from this perspective, malicious machines, definite as such in a perfectly rigorous and observer-independent fashion, are not only potentially in our future, but here already, in a rudimentary and (fortunately!) relatively benign, controlled-in-the-lab form. Likewise, Searle has not really thought about the nature of information from a technical perspective and how it is that, from that perspective, the Fourth R is very, very real. As the late John Pollock told me once in personal conversation, “Whether or not you’re right that Searle’s Chinese Room Argument is sound, of this I’m sure: There will come a time when common parlance and common wisdom will have erected and affirmed a sense of language understanding that is correctly ascribed to machines—and the argument will simply be passé. Searle’s sense of ‘understanding’ will forgotten.”

Fan that I am, it saddens me to report that the errors of Searle’s ways in his review run, alas, much deeper than a failure to refute his two targets. This should already be quite clear to sane readers. To wrap up, I point to just one fundamental defect among many in Searle’s thinking. The defect is a failure to understand how logic and mathematics, as distinguished from informal analytic philosophy, work, and what—what can be called—logico-mathematics is. The failure of understanding to which I refer surfaces in Searle’s review repeatedly; this failure is a terrible intellectual cancer. Once this cancerous thinking has a foothold, it spreads almost everywhere, and the result is that the philosopher ends up operating in a sphere of informal common sense that is at odds not only with the meaning of language used by smart others but with that which has been literally proved. I’m pointing here to the failure to understand that terms like “computation” and “information” (and, for that matter, the terms that are used to express the axiomatizations of physical science that are fast making that science informational in nature for us, e.g., those terms used to express the field axioms in axiomatic physics, which views even the physical world informationally) are fundamentally equivocal between two radically different meanings. One meaning is observer-relative; the other is absolutely not; and the second non-observer-relative meaning is often captured in logico-mathematics. I have space here to explain only briefly, through a single, simple example.

Thinking that he is reminding the reader and the world of a key fact disclosed by good, old-fashioned, non-technical analytic philosophy, Searle writes (emphasis his) in his review: “Except for cases of computations carried out by conscious human beings, computation, as defined by Alan Turing and as implemented in actual pieces of machinery, is observer relative.” In the sense of “computation” captured and explained in logico-mathematics, this is flatly false; and it’s easy as pie to see this. Here’s an example: There is a well-known theorem (TMR) that whatever function $f$ from (the natural numbers) $\mathbb{N}$ to $\mathbb{N}$ that can be computed by a Turing machine can also be computed by a register machine. Or, put another way, for every Turing-machine computation $c$ of $f(n)$, there is a register-machine computation $c'$ of $f(n)$. Now, if every conscious mind were to expire tomorrow at 12 noon NY time, (TMR) would remain true. And not only that, (TMR) would continue to be an ironclad constraint governing the non-conscious universe. No physical process, no chemical process, no biological process, no such process anywhere in the non-conscious universe could ever violate (TMR). Or, putting the moral in another form, aimed directly at Searle, all of these processes would conform to (TMR) despite the fact that no observers exist. What Floridi is prophetically telling us, and explaining, viewed from the formalist’s point of view, is that we have now passed into an epoch in which reality for us is seen through the lens of the logico-mathematics that subsumes (TMR), and includes a host of other truths that, alas, Searle seems to be doing his best to head-in-sand avoid.

NOTES


BIBLIOGRAPHY


Most of the remaining seven months will be spent eating, sleeping, using cell phones or other communication devices, and playing video games (already 69 percent of American heads of households play computer and video games).4

Building a worldwide, ethical infosphere, a fair digital habitat for all, raises unprecedented challenges for humanity in the twenty-first century. The U.S. Department of Commerce and the U.S. National Science Foundation have identified “NBIC” (Nanotechnology, Biotechnology, Information Technology, and Cognitive Science) as a national priority area of research and have recently sponsored a report entitled “Converging Technologies for Improving Human Performance.” And in March 2000, the EU Heads of States and Governments acknowledged the radical transformations brought about by ICT when they agreed to make the EU “the most competitive and dynamic knowledge-driven economy by 2010.”

Information and Communication Technologies and the information society are bringing concrete and imminent opportunities for enormous benefit to people’s education, welfare, prosperity, and edification, as well as great economic advantages. But they also carry significant risks and generate moral dilemma and profound philosophical questions about human nature, the organization of a fair society, the “morally good life,” and our responsibilities and obligations to present and future generations. In short, because the informational revolution is causing an exponential growth in human powers to understand, shape, and control ever more aspects of reality, it is equally making us increasingly responsible, morally speaking, for the way the world is, will, and should be, and for the role we are playing as stewards of our future digital environment. The informationalization of the world, of human society, and of ordinary life has created entirely new realities, made possible unprecedented phenomena and experiences, provided a wealth of extremely powerful tools and methodologies, raised a wide range of unique problems and conceptual issues, and opened up endless possibilities hitherto unimaginable. As a result, it has also deeply affected our moral choices and actions, affected the way in which we understand and evaluate moral issues, and posed fundamental ethical problems, whose complexity and dimensions are rapidly growing and evolving. It would not be an exaggeration to say that many ethical issues are related to or dependent on the computer revolution.

In this paper, I will look at the roots of the problem: what sort of impact ICTs are having or will soon have on our lives, and what kind of new ethical scenarios such technological transformations are ushering in. For this purpose, it will be convenient to explain immediately two key concepts and then outline the main claim that will be substantiated and explained in the following pages.

The first concept is that of infosphere, a neologism I coined in the nineties’ on the basis of “biosphere,” a term referring to that limited region on our planet that supports life. “Infosphere” denotes the whole informational environment constituted by all informational entities (thus also including informational agents like us or like companies, governments,

---

FROM THE ARCHIVES: AI AND AXIOLOGY

Understanding Information Ethics

Luciano Floridi

UNIVERSITY OF OXFORD


1. INTRODUCTION

The informational revolution has been changing the world profoundly and irreversibly for more than half a century now, at a breathtaking pace and with an unprecedented scope. In a recent study on the evolution of information, researchers at Berkeley’s School of Information Management and Systems estimated that humanity had accumulated approximately 12 exabytes of data in the course of its history, but that the world had produced more than 5 exabytes of data just in 2002. This is almost 800 MB of recorded information produced per person each year. It is like saying that every new born baby comes to the world with a burden of 30 feet of books, the equivalent of 800 MB of information on paper. Most of these data are of course digital: 92 percent of them were stored on magnetic media, mostly in individuals’ hard disks (the phenomenon is known as the “democratization” of data). So, hundreds of millions of computing machines are constantly employed to cope with exabytes of data. In 2005, they were more than 900M. By the end of 2007, it is estimated that there will be over 1.15B PCs in use, at a compound annual growth of 11.4 percent. Of course, PCs are among the greatest sources of further exabytes.

All these numbers will keep growing for the foreseeable future. The result is that information and communication technologies (ICTs) are building the new informational habitat (what I shall define below as the infosphere) in which future generations will spend most of their time. In 2007, for example, it is estimated that American adults and teens will spend on average 3,518 waking hours inside the infosphere, watching television, surfing the Internet, reading daily newspapers, and listening to personal music devices. This is a total amount of nearly five months.
etc.), their properties, interactions, processes, and mutual relations. It is an environment comparable to but different from cyberspace (which is only one of its sub-regions, as it were), since it also includes offline and analogue spaces of information. We shall see that it is also an environment (and hence a concept) that is rapidly evolving. The alert reader will notice a (intended) shift from a semantic (the infosphere understood as a space of contents) to an ontic conceptions (the infosphere understood as an environment populated by informational entities).

The second concept is that of **re-ontologization**, another neologism that I have recently introduced in order to refer to a very radical form of re-engineering, one that not only designs, constructs, or structures a system (e.g., a company, a machine, or some artefact) anew, but that fundamentally transforms its intrinsic nature. In this sense, for example, nanotechnologies and biotechnologies are not merely changing (re-engineering) the world in a very significant way (as did the invention of gunpowder, for example) but actually reshaping (re-ontologizing) it.

Using the two previous concepts, my basic claims can now be formulated thus: computers and, more generally, digital ICTs are re-ontologizing the very nature of (and hence what we mean by) the infosphere; here lies the source of some profound ethical transformations and challenging problems; and Information Ethics (IE), understood as the philosophical foundation of Computer Ethics, can deal successfully with such challenges.

Unpacking these claims will require two steps. In sections 2-4, I will first analyze three fundamental trends in the re-ontologization of the infosphere. This step should provide a sufficiently detailed background against which the reader will be able to evaluate the nature and scope of Information Ethics. In section 5, I will then introduce Information Ethics itself. I say “introduce” because the hard and detailed work of marshalling arguments and replies to objections will have to be left to the specialized literature. Metaphorically, the goal will be to provide a taste of Information Ethics, not the actual recipes. Some concluding remarks in section 6 will close this paper.

### 2. The Rise of the Frictionless Infosphere

The most obvious way in which the new ICTs are re-ontologizing the infosphere concerns the transition from analogue to digital data and then the ever-increasing growth of our digital space. This radical re-ontologization of the infosphere is largely due to the fundamental convergence between digital resources and digital tools.

The ontology of the ICTs available (e.g., software, databases, communication channels and protocols, etc.) is now the same as (and hence fully compatible with) the ontology of their objects. This was one of Turing’s most consequential intuitions: in the re-ontologized infosphere, there is no longer any substantial difference between the *processor* and the *processed*, so the digital deals effortlessly and seamlessly with the digital. This potentially eliminates one of the most long-standing bottlenecks in the infosphere and, as a result, there is a gradual erasure of *ontological friction*.

**Ontological friction** refers to the forces that oppose the flow of information within (a region of) the infosphere and, hence, (as a coefficient) to the amount of work and effort required to generate, obtain, process, and transmit information in a given environment, e.g., by establishing and maintaining channels of communication and by overcoming obstacles in the flow of information such as distance, noise, lack of resources (especially time and memory), amount and complexity of the data to be processed, and so forth. Given a certain amount of information available in (a region of) the infosphere, the lower the ontological friction in it, the higher the accessibility of that amount of information becomes. Thus, if one could quantify ontological friction from 0 to 1, a fully successful firewall would produce a 0.0 degree of friction for any unwanted connection, i.e., a complete standstill in the flow of the unwanted data through its “barrier.” On the other hand, we describe our society as informationally porous the more it tends towards a 0 degree of informational friction.

Because of their “data superconductivity,” ICTs are well known for being among the most influential factors that affect the ontological friction in the infosphere. We are all acquainted with daily aspects of a frictionless infosphere, such as spamming and micropayments. Three other significant consequences are:

1. **no right to ignore**: in an increasingly porous society, it will become progressively less credible to claim ignorance when confronted by easily predictable events (e.g., as George W. Bush did with respect to Hurricane Katrina’s disastrous effects on New Orleans’s flood barriers) and painfully obvious facts (e.g., as British politician Tessa Jowell did with respect to her husband’s finances in a widely publicized scandal);

2. **vast common knowledge**: this is a technical term from epistemic logic, which basically refers to the case in which everybody not only knows that *p* but also knows that everybody knows that everybody knows that *p*. In other words, (a) will also be the case because meta-information about how much information is, was, or should have been available will become overabundant.

From (a) and (b) it follows that, in the future,

3. **we shall witness a steady increase in agents’ responsibilities.** In particular, ICTs are making human agents increasingly accountable, morally speaking, for the way the world is, will, and should be.

### 3. The Global Infosphere or How Information is Becoming Our Ecosystem

During the last decade or so, we have become accustomed to conceptualizing our life online as a mixture between an evolutionary adaptation of human agents to a digital environment and a form of post-modern, neo-colonization of the latter by the former. This is probably a mistake. Computers are as much re-ontologizing our world as they are creating new realities. The threshold between
here (analogue, carbon-based, off-line) and there (digital, silicon-based, online) is fast becoming blurred, but this is as much to the advantage of the latter as it is of the former. This recent phenomenon is variously known as “Ubiquitous Computing,” “Ambient Intelligence,” “The Internet of Things” (ITU report, November 2005 www.itu.int/internetofthings) or “Web-augmented things.” It is or will soon be the next stage in the digital revolution. To put it dramatically, the infosphere is progressively absorbing any other space. Let me explain.

In the (fast approaching) future, more and more objects will be what I’d like to call I*entities, able to learn, advise, and communicate with each other. A good example is provided by Radio Frequency Identification (RFID) tags, which can store and remotely retrieve data from an object and give it a unique identity, like a barcode. Tags can measure less than half a millimeter square and are thinner than paper. Incorporate this tiny microchip in everything, including humans and animals, and you have created I*entities. This is not science fiction. According to a report by Market Research Company InStat, the worldwide production of RFID will increase more than 25-fold between 2005 and 2010 and reach 33 billion. Imagine networking these 33 billion I*entities together with all the hundreds of millions of PCs, DVDs, iPods, and ICT devices available and you see that the infosphere is no longer "there" but "here" and it is here to stay. Your Nike and iPod already talk to each other (http://www.apple.com/ipod/nike/).

Nowadays, we are used to considering the space of information as something we log-in to and log-out from. Our view of the world (our metaphysics) is still modern or Newtonian: it is made of “dead” cars, buildings, furniture, clothes, which are non-interactive, irresponsible, and incapable of communicating, learning, or memorizing. But what we still experience as the world offline is bound to become a fully interactive and responsive environment of wireless, pervasive, distributed, A2A (anything to anything) information processes, that works A4A (anywhere for anytime), in real time. This will first gently invite us to understand the world as something “a-live” (artificially live). Such animation of the world will, paradoxically, make our outlook closer to that of pre-technological cultures, which interpreted all aspects of nature as inhabited by teleological forces.

The second step will be a reconceptualization of our ontology in informational terms. It will become normal to consider the world as part of the infosphere, not so much in the dystopian sense expressed by a Matrix-like scenario, where the “real reality” is still as hard as the metal of the machines that inhabit it; but in the evolutionary, hybrid sense represented by an environment such as New Port City, the fictional, post-cybernetic metropolis of Ghost in the Shell (http://en.wikipedia.org/wiki/Ghost_in_the_Shell). This is the shift I alerted you to some pages ago.

The infosphere will not be a virtual environment supported by a genuinely “material” world behind; rather, it will be the world itself that will be increasingly interpreted and understood informationally, as part of the infosphere. At the end of this shift, the infosphere will have moved from being a way to refer to the space of information to being synonymous with Being or reality. This is the sort of informational metaphysics I suspect we shall find increasingly easy to embrace. Just ask one of the more than 8 million players of War of Warcraft, one of the almost 7 million inhabitants of Second Life, or one of the 70 million owners of Neopets.

4. THE EVOLUTION OF INFORSgs

We have seen that we probably the last generation to experience a clear difference between “offline” and online. The third transformation that I wish to highlight concerns precisely the emergence of artificial and hybrid (multi) agents, i.e., partly artificial and partly human. Consider, for example, a whole family as a single agent, equipped with digital cameras, laptops, Palm OS handhelds, iPods, mobile phones, camcorders, wireless networks, digital TVs, DVDs, CD players, and so on.

These new agents already share the same ontology with their environment and can operate in it with much more freedom and control. We (shall) delegate or outsource to artificial agents memories, decisions, routine tasks, and other activities in ways that will be increasingly integrated with us and with our understanding of what it means to be an agent. This is rather well known, but two other aspects of this transformation may be in need of some clarification.

On the one hand, in the re-ontologized infosphere, progressively populated by ontologically equal agents, where there is no difference between processors and processed, online and offline, all interactions become equally digital. They are all interpretable as *read/write* (i.e., access/alter) activities, with *execute* the remaining type of process. It is easy to predict that, in such an environment, the moral status and accountability of artificial agents will become an ever more challenging issue (Floridi and Sanders 2004b).

On the other hand, our understanding of ourselves as agents will also be deeply affected. I am not referring here to the sci-fi vision of a “cyborged” humanity. Walking around with something like a Bluetooth wireless headset implanted in your ear does not seem the best way forward, and not least because it contradicts the social message it is also meant to be sending: being on call 24×7 is a form of slavery, and anyone so busy and important should have a PA instead. The truth is rather that being a sort of cyborg is not what people will embrace, but what they will try to avoid, unless it is inevitable (more on this shortly).

Nor am I referring to a genetically modified humanity, in charge of its informational DNA and, hence, of its future embodiments. This is something that we shall probably see in the future, but it is still too far away, both technically (safely doable) and ethically (in the sense of being morally acceptable as normally as having a heart by-pass or some new spectaculars: we are still struggling with the ethics of stem cells), to be discussed at this stage.

What I have in mind is a quieter, less sensational, and yet crucial and profound change in our conception of what it means to be an agent. We are all becoming connected informational organisms (inforsgs). This is happening not
through some fanciful transformation in our body but, more seriously and realistically, through the re-ontologization of our environment and of ourselves.

By re-ontologizing the infosphere, digital ICTs have brought to light the intrinsically informational nature of human agents. This is not equivalent to saying that people have digital alter egos, some Mesrs Hydes represented by their @’s, blogs, and https. This trivial point only encourages us to mistake digital ICTs for merely enhancing technologies. The informational nature of agents should not be confused with a “data shadow” either. The more radical change, brought about by the re-ontologization of the infosphere, will be the disclosure of human agents as interconnected, informational organisms among other informational organisms and agents.

Consider the distinction between enhancing and augmenting appliances. The switches and dials of the former are interfaces meant to plug the appliance in to the user’s body ergonomically. Drills and guns are perfect examples. It is the cyborg idea. The data and control panels of augmenting appliances are instead interfaces between different possible worlds: on the one hand there is the human user’s Umwelt, Euclidean, Newtonian, colorful, and so forth, and on the other hand there is the dynamic, watery, soapy, hot, and dark world of the dishwasher; the equally watery, soapy, hot, and dark but also spinning world of the washing machine; or the still, aseptic, soapless, cold, and potentially luminous world of the refrigerator. These robots can be successful because they have their environments “wrapped” and tailored around their capacities, not vice versa. Imagine someone trying to build a droid like C3PO capable of washing their dishes in the sink exactly in the same way as a human agent would.

Now, ICTs are not augmenting or empowering in the sense just explained. They are re-ontologizing devices because they engineer environments that the user is then enabled to enter through (possibly friendly) gateways. It is a form of initiation. Looking at the history of the mouse (http://sloan.stanford.edu/mousesite/), for example, one discovers that our technology has not only adapted to, but also educated, us as users. Douglas Engelbart once told me that he had even experimented with a mouse to be placed under the desk, to be operated with one’s leg, in order to leave the user’s hands free. Human-Computer Interaction (HCI) is a symmetric relation of mutual symbiosis.

To return to our distinction, whilst a dishwasher interface is a panel through which the machine enters into the user’s Umwelt, it is the RPT model.

The “External” R(esource) T(arget) Model.

Figure 1. The “External” R(esource) T(arget) Model.

Suppose our moral agent A is interested in pursuing whatever she considers her best course of action, given her predicament. We shall assume that A’s evaluations and interactions have some moral value, but no specific value needs to be introduced at this stage. Intuitively, A can avail herself of some information (information as a resource) to generate some other information (information as a product) and, in so doing, affect her informational environment (information as target). This simple model, summarized in Figure 1, may help one to get some initial orientation in the multiplicity of issues belonging to Information Ethics. I shall refer to it as the RPT model.

The RPT model is useful to rectify an excessive emphasis occasionally placed on specific technologies (this happens most notably in computer ethics) by calling our attention to the more fundamental phenomenon of information in all its varieties and long tradition. This was also Wiener’s position and the various difficulties encountered in the conceptual foundations of computer ethics are arguably
connected to the fact that the latter has not yet been recognized as primarily an environmental ethics, whose main concern is (or should be) the ecological management and well being of the infosphere.

Since the appearance of the first works in the eighties, Information Ethics has been claimed to be the study of moral issues arising from one or another of the three distinct “information arrows” in the RPT model. This is not entirely satisfactory.

5.1. INFORMATION-AS-A-RESOURCE ETHICS
Consider first the crucial role played by information as a resource for A’s moral evaluations and actions. Moral evaluations and actions have an epistemic component, since A may be expected to proceed “to the best of her information,” that is, A may be expected to avail herself of whatever information she can muster, in order to reach (better) conclusions about what can and ought to be done in some given circumstances. Socrates already argued that a moral agent is naturally interested in gaining as much valuable information as the circumstances require, and that a well-informed agent is more likely to do the right thing. The ensuing “ethical intellectualism” analyzes evil and morally wrong behavior as the outcome of deficient information. Conversely, A’s moral responsibility tends to be directly proportional to A’s degree of information: any decrease in the latter usually corresponds to a decrease in the former. This is the sense in which information occurs in the guise of judicial evidence. It is also the sense in which one speaks of A’s informed decision, informed consent, or well-informed participation. In Christian ethics, even the worst sins can be forgiven in the light of the sinner’s insufficient information, as a counterfactual evaluation is possible: had A been properly informed A would have acted differently and hence would not have sinned (Luke 23:34). In a secular context, Oedipus and Macbeth remind us how the mismanagement of informational resources may have tragic consequences.

From a “resource” perspective, it seems that the moral machine needs information, and quite a lot of it, to function properly. However, even within the limited scope adopted by an analysis based solely on information as a resource and, hence, a merely semantic view of the infosphere, care should be exercised, lest all ethical discourse is reduced to the nuances of higher quantity, quality, and intelligibility of informational resources. The more the better is not the only, nor always the best, rule of thumb. For the (sometimes explicit and conscious) withdrawal of information can often make a significant difference. A may need to lack (or preclude herself from accessing) some information in order to achieve morally desirable goals, such as protecting anonymity, enhancing fair treatment, or implementing unbiased evaluation. Famously, Rawls’ “veil of ignorance” exploits precisely this aspect of information-as-a-resource, in order to develop an impartial approach to justice (Rawls 1999). Being informed is not always a blessing and might even be morally wrong or dangerous.

Whether the (quantitative and qualitative) presence or the (total) absence of information-as-a-resource is in question, it is obvious that there is a perfectly reasonable sense in which Information Ethics may be described as the study of the moral issues arising from “the triple A”: availability, accessibility, and accuracy of informational resources, independently of their format, kind, and physical support. Rawls’ position has been already mentioned. Other examples of issues in IE, understood as Information-as-resource Ethics, are the so-called digital divide, the problem of infoglut, and the analysis of the reliability and trustworthiness of information sources.

5.2. INFORMATION-AS-A-PRODUCT ETHICS
A second but closely related sense in which information plays an important moral role is as a product of A’s moral evaluations and actions. A is not only an information consumer but also an information producer, who may be subject to constraints while being able to take advantage of opportunities. Both constraints and opportunities call for an ethical analysis. Thus, IE, understood as Information-as-a-product Ethics, may cover moral issues arising, for example, in the context of accountability, liability, libel legislation, testimony, plagiarism, advertising, propaganda, misinformation, and more generally of pragmatic rules of communication à la Grice. Kant’s analysis of the immorality of lying is one of the best known case studies in the philosophical literature concerning this kind of Information Ethics. Cassandra and Laocoon, pointlessly warning the Trojans against the Greeks’ wooden horse, remind us how the ineffective management of informational products may have tragic consequences.

5.3. INFORMATION-AS-A-TARGET ETHICS
Independently of A’s information input (info-resource) and output (info-product), there is a third sense in which information may be subject to ethical analysis, namely, when A’s moral evaluations and actions affect the informational environment. Think, for example, of A’s respect for, or breach of, someone’s information privacy or confidentiality. Hacking, understood as the unauthorized access to a (usually computerized) information system, is another good example. It is not uncommon to mistake it for a problem to be discussed within the conceptual frame of an ethics of informational resources. This misclassification allows the hacker to defend his position by arguing that no use (let alone misuse) of the accessed information has been made. Yet hacking, properly understood, is a form of breach of privacy. What is in question is not what A does with the information, which has been accessed without authorization, but what it means for an informational environment to be accessed by A without authorization. So the analysis of hacking belongs to an Info-target Ethics. Other issues here include security, vandalism (from the burning of libraries and books to the dissemination of viruses), piracy, intellectual property, open source, freedom of expression, censorship, filtering, and content control. Mill’s analysis “Of the Liberty of Thought and Discussion” is a classic of IE interpreted as Information-as-target Ethics. Juliet, simulating her death, and Hamlet, re-enacting his father’s homicide, show how the risky management of one’s informational environment may have tragic consequences.
5.4. THE LIMITS OF ANY MICROETHICAL APPROACH TO INFORMATION ETHICS

At the end of this overview, it seems that the RPT model may help one to get some initial orientation in the multiplicity of issues belonging to different interpretations of Information Ethics. Despite its advantages, however, the model can still be criticized for being inadequate in two respects.

On the one hand, the model is too simplistic. Arguably, several important issues belong *mainly but not only* to the analysis of just one “informational arrow.” The reader may have already thought of several examples that illustrate the problem: someone’s testimony is someone’s else trustworthy information; A’s responsibility may be determined by the information A holds, but it may also concern the information A issues; censorship affects A both as a user and as a producer of information; misinformation (i.e., the deliberate production and distribution of false and misleading contents) is an ethical problem that concerns all three “informational arrows”; freedom of speech also affects the availability of offensive content (e.g., child pornography, violent content, and socially, politically, or religiously disrespectful statements) that might be morally questionable and should not circulate.

On the other hand, the model is insufficiently inclusive. There are many important issues that cannot easily be placed on the map at all, for they really emerge from, or supervene on, the interactions among the “informational arrows.” Two significant examples may suffice: “big brother,” that is, the problem of *monitoring and controlling* anything that might concern A; the debate about information ownership (including copyright and patents legislation) and *fair use*, which affects both users and producers while shaping their informational environment.

So the criticism is reasonable. The RPT model is indeed inadequate. Yet why it is inadequate is a different matter. The tripartite analysis just provided is unsatisfactory, despite its partial usefulness, precisely because any interpretation of Information Ethics based on only one of the “informational arrows” is bound to be too reductive. As the examples mentioned above emphasize, supporters of narrowly constructed interpretations of Information Ethics as a *microethics* (that is, a practical, field-dependent, applied, and professional ethics) are faced by the problem of being unable to cope with a large variety of relevant issues (I mentioned some of them above), which remain either uncovered or inexplicable. In other words, the model shows that idiosyncratic versions of IE, which privilege only some limited aspects of the *information cycle*, are unsatisfactory. We should not use the model to attempt to pigeonhole problems neatly, which is impossible. We should rather exploit it as a useful scheme to be superseded, in view of a more encompassing approach to IE as a *macroethics*, that is, a theoretical, field-independent, applicable ethics. Philosophers will recognize here a Wittgensteinian ladder, which can be used to reach a new starting point, but then can be discharged.

In order to climb up on, and then throw away, any narrowly constructed conception of Information Ethics, a more encompassing approach to IE needs to

i) bring together the three “informational arrows”;

ii) consider the whole information-cycle; and

iii) take seriously the ontological shift in the nature of the infosphere that I emphasized above, thus analyzing informationally all entities involved (including the moral agent A) and their changes, actions, and interactions, treating them not apart from, but as part of the informational environment to which they belong as informational systems themselves.

Whereas steps (i) and (ii) do not pose particular problems and may be shared by other approaches to IE, step (iii) is crucial but involves an “update” in the ontological conception of “information” at stake. Instead of limiting the analysis to (veridical) semantic contents—as any narrower interpretation of IE as a microethics inevitably does—an ecological approach to Information Ethics also looks at information from an object-oriented perspective and treats it as an entity as well. In other words, we move from a (broadly constructed) epistemological or semantic conception of Information Ethics—in which information is roughly equivalent to news or contents—to one which is typically ontological, and treat information as equivalent to patterns or entities in the world. Thus, in the revised RPT model, represented in Figure 2, the agent is embodied and embedded, as an informational agent, in an equally informational environment.

![Figure 2](image-url)

*A simple analogy may help to introduce this new perspective.* Imagine looking at the whole universe from a chemical perspective. Every entity and process will satisfy a certain chemical description. A human being, for example, will be 70 percent water and 30 percent something else. Now consider an informational perspective. The same entities will be described as clusters of data, that is, as informational objects. More precisely, our agent A (like any other entity) will be a discrete, self-contained, encapsulated package containing

i) the appropriate data structures, which constitute the nature of the entity in question, that is, the state of the object, its unique identity, and its attributes; and
ii) a collection of operations, functions, or procedures, which are activated by various interactions or stimuli (that is, messages received from other objects or changes within itself) and correspondingly define how the object behaves or reacts to them.

At this level of analysis, informational systems as such, rather than just living systems in general, are raised to the role of agents and patients of any action, with environmental processes, changes, and interactions equally described informationally.

Understanding the nature of IE ontologically rather than epistemologically modifies the interpretation of the scope of IE. Not only can an ecological IE gain a global view of the whole life-cycle of information, thus overcoming the limits of other microethical approaches, but it can also claim a role as a macroethics, that is, as an ethics that concerns the whole realm of reality. This is what we shall see in the next section.

5.5. INFORMATION ETHICS AS A MACROETHICS

Information Ethics is patient-oriented, ontocentric, ecological macroethics (Floridi 1999a; Floridi and Sanders 1999). These are technical expressions that can be intuitively explained by comparing IE to other environmental approaches.

Biocentric ethics usually grounds its analysis of the moral standing of bio-entities and eco-systems on the intrinsic worthiness of life and the intrinsically negative value of suffering. It seeks to develop a patient-oriented ethics in which the “patient” may be not only a human being, but also any form of life. Indeed, Land Ethics extends the concept of patient to any component of the environment, thus coming close to the approach defended by Information Ethics. Rowlands [2000], for example, has recently proposed an interesting approach to environmental ethics in terms of naturalization of semantic information. According to him,

>There is value in the environment. This value consists in a certain sort of information, information that exists in the relation between affordances of the environment and their indices. This information exists independently of . . . sentient creatures. . . . The information is there. It is in the world. What makes this information value, however, is the fact that it is valued by valuing creatures [because of evolutionary reasons], or that it would be valued by valuing creatures if there were any around. (p. 153)

Any form of life is deemed to enjoy some essential proprieties or moral interests that deserve and demand to be respected, at least minimally and relatively, that is, in a possibly overriding sense, when contrasted to other interests. So biocentric ethics argues that the nature and well being of the patient of any action constitute (at least partly) its moral standing and that the latter makes important claims on the interacting agent, claims that in principle ought to contribute to the guidance of the agent’s ethical decisions and the constraint of the agent’s moral behavior. The “receiver” of the action, the patient, is placed at the core of the ethical discourse, as a center of moral concern, while the “transmitter” of any moral action, the agent, is moved to its periphery.

Substitute now “life” with “existence” and it should become clear what IE amounts to. Information Ethics is an ecological ethics that replaces biocentrism with ontocentrism. It suggests that there is something even more elemental than life, namely, being—that is, the existence and flourishing of all entities and their global environment—and something more fundamental than suffering, namely, entropy. The latter is most emphatically not the physicists’ concept of thermodynamic entropy. Entropy here refers to any kind of destruction, corruption, pollution, and depletion of informational objects (mind, not of information as content), that is, any form of impoverishment of being. It is comparable to the metaphysical concept of nothingness. Information Ethics then provides a common vocabulary to understand the whole realm of being informationally. Information Ethics holds that being/information has an intrinsic worthiness. It substantiates this position by recognizing that any informational entity has a Spinozian right to persist in its own status, and a Constructionist right to flourish, i.e., to improve and enrich its existence and essence. As a consequence of such “rights,” we shall see that IE evaluates the duty of any moral agent in terms of contribution to the growth of the infosphere and any process, action, or event that negatively affects the whole infosphere—not just an informational entity—as an increase in its level of entropy (or nothingness) and, hence, an instance of evil (Floridi and Sanders 1999, 2001; Floridi 2003).

In IE, the ethical discourse concerns any entity, understood informationally, that is, not only all persons, their cultivation, well being, and social interactions, not only animals, plants, and their proper natural life, but also anything that exists, from paintings and books to stars and stones; anything that may or will exist, like future generations; and anything that was but is no more, like our ancestors or old civilizations. Information Ethics is impartial and universal because it brings to ultimate completion the process of enlargement of the concept of what may count as a center of a (no matter how minimal) moral claim, which now includes every instance of being understood informationally, no matter whether physically implemented or not. In this respect, IE holds that every entity, as an expression of being, has a dignity, constituted by its mode of existence and essence (the collection of all the elementary proprieties that constitute it for what it is), which deserve to be respected (at least in a minimal and overridable sense) and, hence, place moral claims on the interacting agent and ought to contribute to the constraint and guidance of his ethical decisions and behavior. This ontological equality principle means that any form of reality (any instance of information/being), simply for the fact of being what it is, enjoys a minimal, initial, overridable, equal right to exist and develop in a way that is appropriate to its nature. The conscious recognition of the ontological equality principle presupposes a disinterested judgment of the moral situation from an objective perspective, i.e., a perspective which is as non-anthropocentric as possible. Moral behavior is less likely without this epistemic virtue. The application of the ontological equality principle is achieved whenever
actions are impartial, universal, and “caring.” At the roots of this approach lies the *ontic trust* binding agents and patients. A straightforward way of clarifying the concept of ontic trust is by drawing an analogy with the concept of “social contract.”

Various forms of contractualism (in ethics) and contractarianism (in political philosophy) argue that moral obligation, the duty of political obedience, or the justice of social institutions gain their support from a so-called “social contract.” This may be a hypothetical agreement between the parties constituting a society (e.g., the people and the sovereign, the members of a community, or the individual and the state). The parties accept to agree to the terms of the contract and thus obtain some rights in exchange for some freedoms that, allegedly, they would enjoy in a hypothetical state of nature. The rights and responsibilities of the parties subscribing to the agreement are the terms of the social contract, whereas the society, state, group, etc. is the entity created for the purpose of enforcing the agreement. Both rights and freedoms are not fixed and may vary, depending on the interpretation of the social contract.

Interpretations of the theory of the social contract tend to be highly (and often unknowingly) anthropocentric (the focus is only on human rational agents) and stress the coercive nature of the agreement. These two aspects are not characteristic of the concept of ontic trust, but the basic idea of a fundamental agreement between parties as a foundation of moral interactions is sensible. In the case of the ontic trust, it is transformed into a primeval, entirely hypothetical pact, logically predating the social contract, which all agents cannot but sign when they come into existence, and that is constantly renewed in successive generations.

Generally speaking, a trust in the English legal system is an entity in which someone (the trustee) holds and manages the former assets of a person (the trustor, or donor) for the benefit of certain persons or entities (the beneficiaries). Strictly speaking, nobody owns the assets, since the trustor has donated them, the trustee has only legal ownership, and the beneficiary has only equitable ownership. Now, the logical form of this sort of agreement can be used to model the ontic trust in the following way:

- the assets or “corpus” is represented by the world, including all existing agents and patients;
- the donors are all past and current generations of agents;
- the trustees are all current individual agents; and
- the beneficiaries are all current and future individual agents and patients.

By coming into being, an agent is made possible thanks to the existence of other entities. It is therefore bound to all that already is both unwillingly and inescapably. It should be so also caringly. Unwillingly because no agent wills itself into existence, though every agent can, in theory, will itself out of it. Inescapably because the ontic bond may be broken by an agent only at the cost of ceasing to exist as an agent. Moral life does not begin with an act of freedom but it may end with one. Caringly because participation in reality by any entity, including an agent—that is, the fact that any entity is an expression of what exists—provides a right to existence and an invitation (not a duty) to respect and take care of other entities. The pact then involves no coercion, but a mutual relation of appreciation, gratitude, and care, which is fostered by the recognition of the dependence of all entities on each other. Existence begins with a gift, even if possibly an unwanted one. A fetus will be initially only a beneficiary of the world. Once she is born and has become a full moral agent, she will be, as an individual, both a beneficiary and a trustee of the world. She will be in charge of taking care of the world, and, insofar as she is a member of the generation of living agents, she will also be a donor of the world. Once dead, she will leave the world to other agents after her and thus become a member of the generation of donors. In short, the life of a human agent becomes a journey from being only a beneficiary to being only a donor, passing through the stage of being a responsible trustee of the world. We begin our career of moral agents as strangers to the world; we should end it as friends of the world.

The obligations and responsibilities imposed by the ontic trust will vary depending on circumstances but, fundamentally, the expectation is that actions will be taken or avoided in view of the welfare of the whole world.

The crucial importance of the radical change in ontological perspective cannot be overstated. Bioethics and Environmental Ethics fail to achieve a level of complete impartiality because they are still biased against what is inanimate, lifeless, intangible, or abstract (even Land Ethics is biased against technology and artefacts, for example). From their perspective, only what is intuitively alive deserves to be considered as a proper center of moral claims, no matter how minimal, so a whole universe escapes their attention. Now, this is precisely the fundamental limit overcome by IE, which further lowers the minimal condition that needs to be satisfied, in order to qualify as a center of moral concern, to the common factor shared by any entity, namely, its informational state. And since any form of being is in any case also a coherent body of information, to say that IE is infocentric is tantamount to interpreting it, correctly, as an ontocentric theory.

The result is that all entities, qua informational objects, have an intrinsic moral value, although possibly quite minimal and overridable, and, hence, they can count as moral patients, subject to some equally minimal degree of moral respect understood as a disinterested, appreciative, and careful attention (Hepburn 1984). As Naess (1973) has maintained, “all things in the biosphere have an equal right to live and blossom.” There seems to be no good reason not to adopt a higher and more inclusive, ontocentric perspective. Not only inanimate but also ideal, intangible, or intellectual objects can have a minimal degree of moral value, no matter how humble, and so be entitled to some respect. There is a famous passage, in one of Einstein’s letters, that well summarizes this ontic perspective advocated by IE.
Some five years prior to his death, Albert Einstein received a letter from a nineteen-year-old girl grieving over the loss of her younger sister. The young woman wished to know what the famous scientist might say to comfort her. On March 4, 1950, Einstein wrote to this young person: “A human being is part of the whole, called by us ‘universe’, a part limited in time and space. He experiences himself, his thoughts and feelings, as something separated from the rest, a kind of optical delusion of his consciousness. This delusion is a kind of prison for us, restricting us to our personal desires and to affection for a few persons close to us. Our task must be to free ourselves from our prison by widening our circle of compassion to embrace all humanity and the whole of nature in its beauty. Nobody is capable of achieving this completely, but the striving for such achievement is in itself a part of the liberation and a foundation for inner security”. (Einstein 1954)

Deep Ecologists have already argued that inanimate things too can have some intrinsic value. And in a well-known article, White (1967) asked, “Do people have ethical obligations toward rocks?” and answered that “To almost all Americans, still saturated with ideas historically dominant in Christianity… the question makes no sense at all. If the time comes when to any considerable group of us such a question is no longer ridiculous, we may be on the verge of a change of value structures that will make possible measures to cope with the growing ecologic crisis. One hopes that there is enough time left.” According to IE, this is the right ecological perspective and it makes perfect sense for any religious tradition (including the Judeo-Christian one) for which the whole universe is God’s creation, is inhabited by the divine, and is a gift to humanity, of which the latter needs to take care. Information Ethics translates all this into informational terms. If something can be a moral patient, then its nature can be taken into consideration by a moral agent A, and contribute to shaping A’s action, no matter how minimally. In more metaphysical terms, IE argues that all aspects and instances of being are worth some initial, perhaps minimal and overridable, form of moral respect.

Enlarging the conception of what can count as a center of moral respect has the advantage of enabling one to make sense of the innovative nature of ICT, as providing a new and powerful conceptual frame. It also enables one to deal more satisfactorily with the original character of some of its moral issues, by approaching them from a theoretically strong perspective. Through time, ethics has steadily moved from a narrow to a more inclusive concept of what can count as a center of moral worth, from the citizen to the biosphere (Nash 1989). The emergence of the infosphere, as a new environment in which human beings spend much of their lives, explains the need to enlarge further the conception of what can qualify as a moral patient. Information Ethics represents the most recent development in this ecumenical trend, a Platonic and ecological approach without a biocentric bias.

More than fifty years ago, Leopold defined Land Ethics as something that “changes the role of Homo sapiens from conqueror of the land-community to plain member and citizen of it. It implies respect for his fellow-members, and also respect for the community as such. The land ethic simply enlarges the boundaries of the community to include soils, waters, plants, and animals, or collectively: the land” (Leopold 1949, 403). Information Ethics translates environmental ethics into terms of infosphere and informational objects, for the land we inhabit is not just the earth.

6. CONCLUSION
As a consequence of the re-ontologization of our ordinary environment, we shall be living in an infosphere that will become increasingly synchronized (time), delocalized (space), and correlated (interactions). Previous revolutions (especially the agricultural and the industrial ones) created macroscopic transformation in our social structures and architectural environments, often without much foresight. The informational revolution is no less dramatic. We shall be in serious trouble if we do not take seriously the fact that we are constructing the new environment that will be inhabited by future generations (Floridi and Sanders 2005).

We should be working on an ecology of the infosphere if we wish to avoid problems such as a tragedy of the digital commons (Greco and Floridi 2004). Unfortunately, I suspect it will take some time and a whole new kind of education and sensitivity to realize that the infosphere is a common space, which needs to be preserved to the advantage of all. One thing seems unquestionable, though: the digital divide will become a chasm, generating new forms of discrimination between those who can be denizens of the infosphere and those who cannot; between insiders and outsiders, between information rich and information poor. It will redesign the map of worldwide society, generating or widening generational, geographic, socio-economic, and cultural divides. But the gap will not be reducible to the distance between industrialized and developing countries, since it will cut across societies (Floridi 2002). We are preparing the ground for tomorrow’s digital favelas.23

NOTES
1. Source: Lyman and Varian [2003]. An exabyte is approximately 1018 bytes, or a billion times a billion bytes.
2. Source: Computer Industry Almanac, Inc.
4. It is an aging population: the average game player is thirty-three years old and has been playing games for twelve years, while the average age of the most frequent game buyer is forty years old. The average adult woman plays games 7.4 hours per week. The average adult man plays 7.6 hours per week. Source: Entertainment Software Association, http://www.theesa.com/facts/top_10_facts.php
5. See, for example, Floridi (1999b) or http://en.wikipedia.org/wiki/Infosphere
6. These sections are based on Floridi (2006) and Floridi (2007b).
7. This section is based on Floridi (1999a) Floridi (2007a), and Floridi (forthcoming).
9. I have analyzed this IT-heodicean problem and the tragedy of the good will in Floridi and Sanders [2001] and in Floridi [2006].
10. The term is introduced by Westin [1968] to describe a digital profile generated from data concerning a user's habits online.

11. The outer world, or reality, as it affects the agent inhabiting it.

12. For a similar position in computer ethics see Maner (1996). On the so-called "uniqueness debate" see Floridi and Sanders (2002a) and Tavani (2002).

13. The interested reader may find a detailed analysis of the model in Floridi (forthcoming).

14. The classic reference here is to Wiener (1954) Bynum (2001) has convincingly argued that Wiener may be considered one of the founding fathers of Information Ethics.

15. See Floridi and Sanders (2002b) for a defense of this position.

16. An early review is provided by Smith (1996).

17. For an analysis of the so-called IT-heodicean problem and of the tragedy of the good will, see Floridi (2006).

18. One may recognise in this approach to Information Ethics a position broadly defended by van den Hoven (1995) and more recently by Mathiesen (2004), who criticises Floridi and Sanders (1999) and is in turn criticized by Mather (2005). Whereas van den Hoven purports to present his approach to IE as an enriching perspective contributing to the debate, Mathiesen means to present her view, restricted to the informational needs and states of the moral agent, as the only correct interpretation of IE. Her position is thus undermined by the problems affecting any microethical interpretation of IE, as Mather well argues.

19. For further details see Floridi (2005a).

20. For a detailed analysis and defense of an object-oriented modelling of informational entities see Floridi (1999a), Floridi and Sanders (1999), and Floridi (2003).

21. "Perspective" here really means level of abstraction; however, for the sake of simplicity the analysis of levels of abstractions has been omitted from this chapter. The interested reader may wish to consult Floridi (forthcoming).

22. There are important and profound ways of understanding this Ur-pact religiously, especially but not only in the Judeo-Christian tradition, where the parties involved are God and Israel or humanity, and their old or new covenant makes it easier to include environmental concerns and values otherwise overlooked from the strongly anthropocentric perspective prima facie endorsed by contemporary contractualism. However, it is not my intention to endorse or even draw on such sources. I am mentioning the point here in order to shed some light both on the origins of contractualism and on a possible way of understanding the ontocentric approach advocated by IE.

23. This paper is based on Floridi (1999a), Floridi and Sanders (2001), Floridi et al. (2003), Floridi and Sanders (2004b), Floridi (2005a), Floridi and Sanders (2005), Floridi (2006), Floridi (2007a), Floridi (2007b), Floridi (forthcoming). I am in debt to all colleagues and friends who shared their comments on those papers. Their full list can be found in those publications. Here I wish to acknowledge that several improvements are due to their feedback. I am also very grateful to the editor, Peter Boltuc, for his kind invitation to contribute to this issue of the APA Newsletter on Philosophy and Computers.

REFERENCES


Too Much Information: Questioning Information Ethics

John Barker
UNIVERSITY OF ILLINOIS AT SPRINGFIELD


1. INTRODUCTION

In a number of recent publications, Luciano Floridi has argued for an ethical framework, called Information Ethics, that assigns special moral value to information. Simply put, Information Ethics holds that all beings, even inanimate beings, have intrinsic moral worth, and that existence is a more fundamental moral value than more traditional values such as happiness and life. Correspondingly, the most fundamental moral evil in the world, on this account, is entropy—this is not the entropy of thermodynamics, but entropy understood as “any kind of destruction, corruption, pollution, and deprecation of informational objects” (Floridi 2007, 9). Floridi regards this moral outlook as a natural extension of environmental ethics, in which non-human entities are treated as possessors of intrinsic moral worth and, more specifically, of land ethics, where the sphere of moral patients is further extended to include inanimate but naturally occurring objects. On Floridi’s view, artifacts can also be moral patients, including even such “virtual” artifacts as computer programs and web pages.

In general, then, Floridi holds that all objects have some moral claim on us, even if some have a weaker claim than others; moreover, they have this moral worth intrinsically, and not because of any special interest we take in them. In this paper, I want to consider the motivation and viability of Information Ethics as a moral framework. While I will not reach any firm conclusions, I will note some potential obstacles to any such moral theory.

2. BACKGROUND AND MOTIVATION

Before continuing, we need to clarify the notions of object and information, as Floridi uses those terms. Briefly, an informational entity is any sort of instantiated structure, any pattern that is realized concretely. In particular, information is not to be understood semantically. An informational object need not have any semantic value; it need not represent the world as being this way or that. Instead, information should simply be thought of as structure.

Now any object whatsoever may be regarded as a realization of some structure or other. Floridi realizes this, and indeed expands on it in a view he terms “Informational Structural Realism” (ISR). ISR is a metaphysical account of the world that basically dispenses with substrates in favor of structures. On this view, the world should be regarded as a system of realized structures, but it is a mistake to ask what substrate the structures are ultimately realized in: it is structures “all the way down.”

ISR is a fascinating thesis, but it will not be my purpose here to offer any further examination or critique of it. I mention it simply to show that when Floridi speaks of informational entities, he is really speaking of arbitrary entities. Information ethics is, in fact, a theory of arbitrary objects as moral patients. By casting it in terms of information, Floridi is stressing that the class of entities we should be concerned about, as moral patients or otherwise, is broader than the familiar concrete objects of our everyday experience. It should include any sort of instantiated information whatsoever, be it a person, a piece of furniture, or a “virtual” web-based object.

Now, Floridi’s central claim, that all entities have some (possibly very minimal) moral claim on us, while fascinating, certainly runs counter to most moral theories that have been proposed. It therefore seems reasonable to ask for some argument for it, or at least some motivation. The main rationale Floridi provides seems to be an argument from precedent. Before people started thinking systematically about ethics, they withheld the status of moral patient from all but the members of their own tribe or nation. Later, this status was extended to the whole of humanity. Many if not most people would now treat at least some non-human animals as moral patients, and some would ascribe moral worth to entire ecosystems and even to inanimate parts of nature. Thus, the history of ethical thinking is one of successively widening the sphere of our moral concern, and the logical end result of this process is to extend our moral concern to all of existence—or so Floridi argues.

However, as it stands this argument seems weak. True, there has been some historical tendency for moral theories to broaden the sphere of appropriate targets of moral concern. This tendency may continue indefinitely, until all of existence is encompassed. And then again, it may not. Here it is worth considering why at least some non-human animals are now generally considered moral patients. The main rationale, both historically and for most contemporary moral theorists, is that animals have a capacity for pleasure and suffering. It does not matter for my purposes whether this is the only or best rationale for extending moral consideration to animals. The point is that some rationale was needed; the mere precedent of extending moral consideration from smaller to larger groups of humans was not itself a sufficient reason to further extend it to animals. Likewise, if we are to extend the sphere of moral patients further, we will need a specific reason to do so, not just precedent.

The most ardent supporters of animal rights have always been Utilitarians, and Utilitarianism justifies the inclusion of animals with a specific account of what constitutes a benefit or harm. Namely, benefit and harm are identified with pleasure and suffering, respectively. Once this identification is made, all it takes to show that a given being is a moral patient is to show that it can experience pleasure and pain. If Floridi were to give a specific account of what constitutes benefit or harm to an arbitrary entity, that would go some way toward providing a rationale for
Information Ethics. It is also desirable for another reason. Floridi’s ethical account is “patient-oriented” (Floridi 2007, 8). This may or may not mean that it is a consequentialist theory; however, it seems fair to assume that in a patient-oriented moral theory, an action’s benefit or harm to moral patients plays a preeminent role in determining its rightness or wrongness. Thus, it seems desirable for such a theory to include a general account of benefit and harm. Moreover, such a theory can presumably be action-guiding only if it provides at least some such account.

Does Floridi offer such an account of benefit and harm? He does identify good and evil with “existence” and “entropy,” respectively; but as I will argue below, it is not clear that this amounts to a general account of benefit and harm, either to individual entities or to the universe or “infosphere” at large. Now, to some extent this omission is understandable, given the pioneering nature of Floridi’s work. However, I will argue below that there are substantial obstacles in principle to providing any such account. The notion of an arbitrary object, I suspect, is simply too broad to support any substantive account of harm and benefit.

3. INFORMATION AND ENTROPY

Let us first see why there is even a question about fundamental good and evil in Information Ethics. Floridi identifies existence as the fundamental positive moral value, and inexistence as the fundamental negative value. Thus, it might seem natural to suppose that an action is beneficial if it creates (informational) objects, and harmful if it destroys them, with the net benefit or harm identified with the net number of objects created or destroyed.

The trouble with this proposal is that given the broad conception of objects that we are working with, every act both creates and destroys objects. Since objects are simply instantiated patterns, there are indefinitely many objects present in any given physical substrate. Any physical change whatsoever involves a change in the set of instantiated patterns, thus creating and destroying informational objects simultaneously. Moreover, even if it were possible to count the number of informational objects in a given medium, such a count would ignore the fact that some beings have more inherent moral worth than others: this is fairly obvious in its own right, and Floridi himself insists on it, asserting that some moral patients have a strong claim on us while for others, the claim is “minimal” and “overridable” (Floridi 2007, 9).

Thus, if we are to take seriously the idea that “being” is the most fundamental good and “inexistence” or “entropy” is the most fundamental evil, we cannot calculate good or evil by simply counting objects. A natural idea, and one which is somewhat suggested by Floridi’s term “entropy,” is that fundamental moral value should be identified with some overall measure of the informational richness or complexity of a system. This would preserve the idea of being and nonbeing as fundamental moral values while avoiding the difficulties involved in the simple counting approach.

One of the best-developed accounts of non-semantic information is statistical information theory. This theory, developed by Claude Shannon in the 1940s, has been used very successfully to describe the amount of information in a signal without describing the signal’s semantic content (if any). Thus, it seems like a natural starting point for describing the overall complexity or richness of a system of informational objects.

Statistical information theory essentially identifies high information content with low probability. Specifically, the Shannon information content of an individual message \( M \) is defined to be \( \log_2(1/p(M)) \), where \( p(M) \) is the probability that \( M \) occurs. As a special case, consider a set of \( 2^n \) messages, each equally likely to occur; then each message will have a probability of \( 2^{-n} \), and an information content of \( \log_2(2^n) = n \) bits, exactly as one would expect. The interesting case occurs when the probability distribution is non-uniform; low probability events occur relatively rarely, and thus convey more information when they do occur.

As is well known, the definition of Shannon information content is formally almost identical to that of statistical entropy in physics. The entropy \( S \) of a given physical system is defined to be \( S = k_B \ln \Omega \), where \( k_B \) is a constant (Boltzmann constant) and \( \Omega \) is the number of microstates corresponding to the system’s macrostate. (A system’s macrostate is simply its macroscopic configuration, abstracting away from microscopic details; the corresponding microstates are those microscopic configurations that would produce that macrostate.) Now for a given microstate \( q \) and corresponding macrostate \( Q \), \( Q \) is simply the probability that the system is in microstate \( q \) given that it is in macrostate \( Q \). In other words, the entropy of a system is simply \( k_B \ln (1/p(Q)) \), where \( p(Q) \) is a uniform probability distribution over the microstates in \( Q \). Alternatively, if we posit a uniform probability distribution \( p \) over all possible microstates \( q \), then we have \( p(q) = p(q) / p(Q) \), and thus \( S = (k_B/p(q)) \ln (p(Q)) = -k_B (\ln p(q)) \ln (1/p(Q)) \); the quantity \( k_B/p(q) \) is a constant because the measure \( p \) is uniform. In any case, we have \( S = K \log (1/p) \), where \( K \) is a constant and \( p \) is the probability of the state in question under some probability measure (the base may be omitted on the log because it only affects the result up to a constant, and may thus be subsumed in \( K \)). Thus, up to a proportionality constant, statistical entropy is a special case of Shannon information content.

However, it is the wrong special case, since, as Floridi states very clearly, the fundamental evil which he refers to as “entropy” is not thermodynamic entropy. And, indeed, in light of the second law of thermodynamics, thermodynamic entropy is not a reasonable quantity for moral agents to try to minimize. Thus, if we are to use Shannon information theory to capture the morally relevant notion of complexity, we will have to use a probability measure other than that described above. However, information theory does not offer us any guidance here, because it does not specify a probability measure: it simply assumes some measure as given. Typically, when applying information theory, we are working with a family of messages with well-defined statistics; thus, a suitable \( p \) is supplied by the context of the problem at hand.

Thus, Shannon information theory provides a measure of a system’s information content, but this measure is relative
to a probability measure $p$. This presents an obstacle to explaining complexity in terms of Shannon information and simultaneously claiming that complexity is a fundamental, intrinsic moral value. If we allow complexity to be relative to a probability measure, then intrinsic moral worth will also be relative to a probability measure. Conceivably, different probability measures could yield wildly different measures of complexity and, thus, of intrinsic moral worth. Thus, it would appear to be necessary to pin down a single probability measure, or at least a family of similar probability measures, in a non-arbitrary manner.

And here is where things get tricky. What probability measure is the right one for measuring the complexity of arbitrary systems? Whatever it is, it must be a probability measure that is in some sense picked out by nature, rather than by our own human interests and concerns. Otherwise, complexity, and thus inherent moral worth, is not really objective, but is tied to a specifically human viewpoint. This goes against the whole thrust of Information Ethics, which seeks to liberate ethics from an anthropocentric viewpoint. Thus, we need to find a natural probability measure for our task. What might such a probability measure look like?

The best-known conception of objective probability is the frequentist conception. According to that conception, the probability of an outcome $O$ of an experiment $E$ is the proportion of times that $O$ occurs in an ideal run of trials of $E$. To apply this notion, we need a well-defined outcome-type $O$, a well-defined experiment-type $E$, and a well-defined set of ideal trials of $E$—and if the latter set is continuous, a well-defined measure on that set. This is all notoriously difficult to apply to non-repeatable event tokens and to particulars in general. To assign a frequentist probability to a particular $x$, it is necessary to subsume $x$ under some general type $T$, and different choices of $T$ may yield different probabilities. In other words, the frequentist probability of a particular depends among other things on how that particular is described. Different ways of describing a particular will correspond to different conceptions of what it is to repeat that particular, and thus, to different measures of how frequently it occurs in a run of cases.

What this means for us is that the information content of a concrete particular depends, potentially, on how we choose to carve up the world. Again, this is not a problem in practice for information theory, since in any given application, a particular (frequentist) probability measure is likely to be singled out by the problem's context. But in describing the information content of completely arbitrary objects, there is no context to guide us. In particular, if we subsume a concrete particular $x$ under a commonly occurring type $T$, it receives a high frequentist probability, and correspondingly low Shannon information content. If we subsume that same particular under a rarely occurring type $T^*$, it receives a low probability and correspondingly high information content.

Thus, it is by no means obvious that there is a choice of probability measure that (a) is natural independently of our own anthropocentric interests and concerns, and (b) gives us a measure of complexity that is a plausible candidate for inherent moral worth, even assuming that the latter has any special tie to complexity in the first place. To be fair, it is also not obvious that there is not such a probability measure. As the measure $p$ from thermodynamics shows, there is at least one natural way of assigning probabilities to physical states, one which does indeed yield a measure of complexity, albeit not the measure of complexity we are looking for. It also raises a further worry. The reason thermodynamic entropy is a bad candidate for basic moral disvalue is simply that it is always increasing, regardless of our actions. That is simply the second law of thermodynamics. What guarantee do we have that complexity, measured in any other way, is not also decreasing inexorably? Thermodynamic entropy can decrease locally; in the region of the universe we care about, at the expense of increased entropy somewhere else, and the same may be true for other measures of complexity. But this fact is surely irrelevant to a patient-centered, non-anthropocentric moral theory.

### 4. Information Everywhere

Statistical information theory is, of course, not the only way to capture the idea of complexity and structure. However, I would argue that the whole notion of complexity or information content becomes trivial unless it is tied to our interests (or someone’s interests) as producers and consumers of information.

How much information is there in a glass of water? The obvious, intuitive answer is: very little. A glass of water is fairly homogeneous and uninteresting. Yet the exact state of a glass of water would represent an enormous amount of information if it were described in its entirety. There are approximately $7.5 \times 10^{24}$ molecules in an eight ounce glass of water. If each molecule has a distinguishable pair of spins, it is necessary to subsume a glass of water under a very general type $T$, and different choices of $T$ may yield different probabilities. In other words, the frequentist probability of a particular depends among other things on how that particular is described. Different ways of describing a particular will correspond to different conceptions of what it is to repeat that particular, and thus, to different measures of how frequently it occurs in a run of cases.

What this means for us is that the information content of a concrete particular depends, potentially, on how we choose to carve up the world. Again, this is not a problem in practice for information theory, since in any given application, a particular (frequentist) probability measure is likely to be singled out by the problem's context. But in describing the information content of completely arbitrary objects, there is no context to guide us. In particular, if we subsume a concrete particular $x$ under a commonly occurring type $T$, it receives a high frequentist probability, and correspondingly low Shannon information content. If we subsume that same particular under a rarely occurring type $T^*$, it receives a low probability and correspondingly high information content.

Thus, it is by no means obvious that there is a choice of probability measure that (a) is natural independently of our own anthropocentric interests and concerns, and (b) gives us a measure of complexity that is a plausible candidate for inherent moral worth, even assuming that the latter has any.
if information content is to serve as a measure of moral
worth, the information described in the previous paragraph
must be excluded.

But on what basis could it be excluded? We might try to
exclude some of the more unconventional encoding
functions, such as the encoding function that represents
the water as storing the entire Library of Congress. Such
encoding functions, it may be argued, are rather unnatural
and do not represent the information that is objectively
present in the water. Even if this is so, there is no getting
around the fact that a glass of water represents a vast amount
of information, in that it would take much information to
accurately describe its complete state. That information
might be rather uninteresting—uninteresting to us, that is—but
so what? If moral worth is tied to information content
per se, then it does not matter whether that information is
interesting. If moral worth is tied to interesting information,
then it appears that moral worth is directly tied to human
concerns after all.

But there is a more fundamental problem with dismissing
some encoding functions \( f \) as unnatural. Whenever
information is stored in a physical medium, there needs
to be an encoding function to relate the medium’s
physical properties to its informational properties. Often,
this function is “natural” in that it relates a natural feature
of information (e.g., the value of a binary variable) to a
natural feature of the physical medium (e.g., high or low
voltage in a circuit, the size and shape of a pit on an optical
disk, magnetic field orientation on a magnetic disk, etc.).
However, there is absolutely no requirement to use natural
encoding functions. There need be no simple relation
whatever between, say, a file’s contents and the physical
properties of the media that store the file. The file could be
encrypted, fragmented, stored on multiple disks in a RAID,
broken up into network packets, etc.

In practice, we always disregard the information that is
present, or may be regarded as present via encoding
functions, in a glass of water. But the reason does not seem
to be a lack of a natural relation between the information
and the state of the water. The reason is that even though
the information is in some sense there, we cannot easily use or
access it. We can regard a glass of water as storing a Library
of Congress, but in practice there is no good reason to do
so. By contrast, a file stored in a possibly very complicated
way is nonetheless accessible and potentially useful to us.

If this is right, then there is a problem with viewing
information’s intrinsic value as something independent
of our own interests as producers and consumers of
information. The problem is that information does not exist
independently of our (or someone’s) interests as producers
and consumers of information. Or, alternatively, information
exists in an essentially unlimited number of different ways:
what we count as information is only a minute subset of all
the information there is. Which of these two cases obtains
is largely a matter of viewpoint. On the former view, even if
inanimate information has moral value, it has value in a way
that is more tied to a human perspective than Floridi lets
on. On the latter, there is simply too much information in
the world for our actions to have any net effect on it.

5. CONCLUSION
The immediate lesson of the last two sections is that overall
complexity, or quantity of information, is a poor measure
of intrinsic moral worth. Now this conclusion, even if true,
may not appear to be terribly damaging to Information
Ethics, as the latter embodies no specific theory of how
to measure moral worth. It may simply be that some other
measure is called for. However, I would argue that the
above considerations pose a challenge to any version of
Information Ethics, for the following reason.

As we have seen, the number of (informational) objects
with which we interact routinely is essentially unlimited,
or at least unimaginably vast. If each object has its own
inherent moral worth, what prevents the huge number
of informational objects that we do not care about from
outweighing the relatively small number that we do care
about, in any given moral decision? For example, I might
radically alter the information content of a glass of water by
drinking it, affecting ever so many informational objects;
why does that fact carry less moral weight than the fact that
drinking the water will quench my thirst and hydrate me?
The answer must be that virtually all informational objects
have negligible moral value, and, indeed, Floridi seems
to acknowledge this by saying that many informational
objects have “minimal” and “overridable” value. But that
claim is rather empty unless some basis is provided for
distinguishing the few objects with much value from the
many with little value.

Of course, one answer is simply to assign moral worth
to objects based on how much we care about them. That
would just about solve the problem. Moreover, that is more
or less what it would take to solve the problem, insofar
as the objects that must be assigned minimal value (lest
ethics become trivial) are in fact objects that we do not
care about. However, this is not an answer Floridi can give.
Moral worth is supposed to be something objects possess
intrinsically, as parts of nature. It is not supposed to be
dependent on our interests and concerns. Thus, what is
needed is an independent standard of moral worth for
arbitrary objects which, while not based directly on human
concern, is at least roughly in line with human concern. And
so far that has not been done.

NOTES
1. See, for example, Floridi 2007, Floridi 2008a, etc.
2. See Floridi 2008b.
3. See Shannon 1948. For a good modern introduction, see MacKay
2003.
4. A base-2 logarithm is used because information is measured in
bits, or base-2 digits. If information is to be measured in base-
10 (decimal) digits, then a base-10 logarithm should be used. In
general, the Shannon information content is defined to be
\( \log_b(M) \), with \( b \) determined by the units in which information
is measured (bits, decimal digits, etc.).
5. This figure is obtained from the number of molecules in a mole
(viz. Avogadro’s number, approximately \( 6 \times 10^{23} \)), the number
of grams in one mole of water (equal to water’s atomic weight,
approximately 18), and the number of grams in 8 ounces (about
227).
The relationship between the mathematical formalisms of entropy and information and Floridi's own ontological or informational perspective is based on order, entropy, and the mathematical formalisms are related, they are not completely identical. The reason for this is that information theory and the concept of order in thermodynamics formalizes the concept of order. Floridi's macroethics approach establishes a normativity that bestows intrinsic value on Being: Information Ethics holds that being/information has an intrinsic worthiness. It substantiates this position by recognizing that any informational entity has a Spinozian right to persist in its own status, and a Constructionist right to flourish, i.e., to improve and enrich its existence and essence.

Hence, according to Information Ethics, protecting and improving Being constitutes the absolute norm. Now, with Being defined in terms of form, its polar opposite, we have mentioned, consists in lack of form or organization. These notions are intuitive and naturally understandable. Yet Floridi establishes another link, through the notion of information, with entropy. Entropy is a thermodynamical concept that was mathematically related to that of information by Claude Shannon, thanks to their both being defined in terms of order and randomness.

The relationship between the mathematical formalisms of entropy and information and Floridi's own ontological or metaphysical take on them is tricky, though. Indeed, Floridi insists emphatically that although his own interpretation and the mathematical formalisms are related, they are not the same. The reason for this is that information theory is silent about content or meaning. In thermodynamic terms, that translates into entropy being randomness as

---

**REFERENCES**


---

**Ethics of Entropy**

Martin Flament Fultot
PARIS IV SORBONNE / SND / CNRS


**INTRODUCTION**

Luciano Floridi reinterprets and re-ontologizes our world informationally. That part of his theory may (or may not) work, but what matters for this paper’s topic is that when it comes to defining what the value of Being is, his informational-ontological interpretation is based on order, organization, and structure. Therefore, there is a common ground between his interpretation and the way modern thermodynamics formalizes the concept of order. Floridi proposes to think of Good as a qualitative order and Evil as its absence or entropy. However, the kind of entities that are of importance to us for our judgments and interventions as agents are ordered and thus valuable because they exist far from equilibrium. In this paper I shall attempt to establish that far-from-equilibrium systems attain ever increasing degrees of order at the cost of faster entropy production. Yet, inversely, by promoting an increase in entropy production, more complex and ordered forms emerge on Earth. Entropy production and order are thus complementary; they imply each other reciprocally. By promoting Evil in Floridi’s sense, Good happens lawfully because order is nature’s favorite way of producing entropy. In short, moving against entropy only creates more entropy.

**I. THE VALUE OF ONTOLOGICAL INFORMATION**

Floridi’s Informational Ethics presents three highly attractive features. The first one is that it develops a theory of macroethics. The second one is that it grounds the origin of value in Being, that is, beyond humans and even living creatures. And the third one, on which I will focus mostly, is that Being is defined in terms of information and entropy.

Floridi starts from the observation that the space where ethically relevant human behavior takes place is being completely and irreversibly transformed by the development and diffusion of information technologies. This particular kind of transformation or “re-ontologization,” as he conceptualizes it, affects “the whole realm of reality,” thus requiring a macroethics approach. It may have been more appropriate to talk about holoethics, rather than “macro,” since it is concerned with how information reconfigures human behavior holistically and globally as opposed to locally and individually. In other words, the space of ethical events becomes, in the new “infosphere,” completely interconnected. Thus, single—ethically relevant—events need to conform to norms that target value in its totality.

Floridi’s macroethics ascribes value not to humans nor living creatures as such but to Being itself. Good corresponds to Being, and Evil corresponds to the suppression or the degradation of Being. As a consequence, Floridi’s radical approach makes room for ethical concerns about inanimate things such as rocks. This is understandable since Information Ethics is, by definition, concerned with information, and the concept of information applies to a lot more than human beings or living creatures. More specifically, however, Floridi makes a move from the common idea of information as, say, a message delivering content such as “tomorrow it will rain” to an ontological conception where entities are re-interpreted informationally. The move seems justified by the polarized axiological scale shared by information and Being, where the latter is clearly a value when compared to nothingness, and the former stands as a value when compared to lack of information. But for information to count as a value qua information, it needs to be understood semantically, that is, not so much as information—despite the fact that Floridi’s theory revolves around that concept—but more simply as form, order, structure. I thus assume that Floridi’s “informational” interpretation of ontological Being is simply a structural interpretation, with order or organization being qualitatively opposed to randomness or “mixed-upness.”

In this way, Floridi’s macroethics approach establishes a normativity that bestows intrinsic value on Being:

Information Ethics holds that being/information has an intrinsic worthiness. It substantiates this position by recognizing that any informational entity has a Spinozian right to persist in its own status, and a Constructionist right to flourish, i.e., to improve and enrich its existence and essence.

Hence, according to Information Ethics, protecting and improving Being constitutes the absolute norm. Now, with Being defined in terms of form, its polar opposite, we have mentioned, consists in lack of form or organization. These notions are intuitive and naturally understandable. Yet Floridi establishes another link, through the notion of information, with entropy. Entropy is a thermodynamical concept that was mathematically related to that of information by Claude Shannon, thanks to their both being defined in terms of order and randomness.

---

**SPRING 2020 | VOLUME 19 | NUMBER 2**
opposed to order. However, the qualitative structure of an ordered state is unspecified by thermodynamics. This can be problematic as it challenges the possibility of a graded normative axiological scale. For instance, two entities may contain the same quantity of information as measured by Shannon’s formula, yet differ qualitatively, as in having different shapes. Do they have identical moral value? Do they deserve equal respect? After all, as Schrödinger said, “any calorie is worth as much as any other calorie.”

Another difficulty for Floridi’s theory of information as constituting the fundamental value comes from the sheer existence of the unilateral arrow of thermodynamic processes. The second law of thermodynamics implies that when there is a potential gradient between two systems, A and B, such that A has a higher level of order, then in time, order will be degraded until A and B are in equilibrium. The typical example is that of heat flowing inevitably from a hotter body (a source) towards a colder body (a sink), thereby dissipating free energy, i.e., reducing the overall amount of order. From the globally encompassing perspective of macroethics, this appears to be problematic since having information on planet Earth comes at the price of degrading the Sun’s own informational state. Moreover, as I will show in the next sections, the increase in Earth’s information entails an ever faster rate of solar informational degradation. The problem for Floridi’s theory of ethics is that this implies that the Earth and all its inhabitants as informational entities are actually doing the work of Evil, defined ontologically as the increase in entropy. The Sun embodies more free energy than the Earth; therefore, it should have more value. Protecting the Sun’s integrity against the entropic action of the Earth should be the norm.

II. FAR-FROM-EQUILIBRIUM SYSTEMS, ORDER, AND ENTROPY

It is surprising that, even though Floridi is well aware of the second law of thermodynamics and the fact that informational entities in one way or another will generate entropy in order to persist in their Being, his theory lacks a conceptual treatment of the crucial case of systems that exist far from thermodynamic equilibrium. Yet these systems present an important obstacle to his view that Being has intrinsic and fundamental value. To see this, consider the following proverbial far-from-equilibrium example: the Rayleigh-Bénard experiment (henceforth R-B).

R-B consists in heating from below a shallow layer of viscous fluid contained in a recipient (think of oil in a circular frying pan.) This creates a uniform potential energy gradient between its bottom temperature and the surface’s temperature at the top of the fluid. Following the second law of thermodynamics, the energy gradient operates as a vector, i.e., a force with a direction, so that the fluid fights the asymmetrical concentration of energy at the bottom (order) by transferring the heat towards the top, thereby restoring thermodynamic equilibrium with the surroundings (entropy). Below a given magnitude for the difference of temperature between the bottom and the top, the transfer occurs by conduction, i.e., stochastic collisions between the moving particles that constitute the fluid. The fluid is thus disordered and disorganized under this regime. Under Floridi’s account, the fluid has little being and therefore value. However, when the magnitude of the potential exceeds a given threshold, a new regular pattern of organization emerges from the interaction between the particles in the fluid. Typically, in a circular recipient, the patterns are constituted by hexagonal convection cells visible to the naked eye. Each cell consists of hundreds of millions of molecules moving in a coordinate fashion. Now the fluid is in a dynamically ordered state and, interestingly, this order or organization constitutes a pattern, i.e., it has a shape, a form. How can this qualitative aspect of organization be understood given the quantitative formalism of thermodynamics?

The answer to this question lies in Ilya Prigogine’s work on far-from-equilibrium systems. The main insight is that the emergence of ordered patterns is due precisely to the requirement to dissipate the free energy pumped into the system from the outside in conformity to the second law of thermodynamics. Concretely, in R-B, the emergence of the very specific pattern of hexagonal convection cells corresponds to an optimal configuration of energy flows within the fluid given the magnitude of the potential and the boundary conditions. The latter include the circular shape of the recipient and constraints such as surface tension. Hexagonal shapes distribute the cells so as to collectively maximize their dissipative surface, which translates into higher entropy generation.

So here we can see a qualitative form of order responding to thermodynamical quantitative principles. The magnitude of the potential field as well as the rate of entropy production vary continuously as a simple scalar; the force simply becomes stronger and entropy increases. However, the state of the system transitions qualitatively from a disordered state to an organized state according to a very specific pattern, which in this case is geometrical. The qualitative aspect serves a quantitative function of maximizing entropy production in response to the asymmetric conditions under which free energy is being pumped into the system. We can see that, in a sense, R-B shows how, in far-from-equilibrium systems, information in Floridi’s sense, or simply qualitative order, is not exactly an intrinsic value, but rather a functional value. Hexagonal cells, as a qualitative ordered Being, have the value of optimizing a natural function, and the function is to conform to the second law of thermodynamics by always creating at least as much entropy as the order is being added.

III. MAXIMUM ENTROPY PRODUCTION RATE

Yet it seems odd to claim that the second law of thermodynamics is responsible for the spontaneous emergence of order. After all, in R-B, order helps maximize a function, yet the second law doesn’t predict any such helping. Why, it may be asked, doesn’t conduction remain the heat transfer regime although simply at a faster rate, proportional to the increase in the potential gradient? The answer is that the second law is only one part of the principle of maximum entropy, the other being, precisely, the maximization function. Indeed, the second law states that, on the long term and on average, entropy tends to increase. In other words, entropy in a system will become
maximal given enough time. But it doesn’t say anything about how entropy is maximized. However, several observations have led many independent researchers to the conclusion that the law of entropy production should state rather that the system will tend to disorder at the fastest rate (given the constraints). With this extension I will refer, following Rod Swenson, to the Law of Maximum Entropy Production (LMEP).

LMEP can be observed even in systems not far from equilibrium. Swenson and Turvey illustrate this by a simple experiment in which an adiabatically sealed chamber is divided by an adiabatic wall into two compartments, each filled with an equal quantity of the same gas although at different temperatures. There is thus a potential field between both compartments with the hottest holding more order or information in Floridi’s metaphysical sense. If a hole is opened on the dividing wall, a channel allows heat to proceed from the hotter to the colder chamber until equilibrium is reached and entropy is maximized, as stated by the second law. If a second hole is opened such that it conducts heat at a different rate from the first one, then depending on the constraints and the configuration of the holes, the system will always distribute the flows along the holes in the optimal way. For instance, if hole 2 can drain some heat before it is all drained through hole 1, then some heat will be drained through hole 2 also. In other words, free energy always seeks to exploit the optimal paths to its own dissipation. The same process can be observed in the mundane setting of a cabin in the woods heated from the inside, where heat will drain to the surroundings through the fastest configuration of windows, doors, and other openings.

Back to the R-B case which is far-from-equilibrium, Swenson and Turvey show that the emergence of the convection cells is inevitable due to the opportunistic exploitation of the configurations that tend to optimize the rate of entropy production. The threshold corresponds to the minimal magnitude of force that will sustain the dynamical ordered state. With enough free energy available within the fluid, a new, non-random configuration becomes possible, and because of LMEP, that configuration will be favored and stabilized “as soon as it gets the chance.”

The point about “getting the chance” deserves a brief pause. The formation of the ordered regime takes time. It is a search the system undergoes, facilitated by the increased amount of kinetic energy produced by the potential field. Akin to a selection process with winner-takes-all rules, the formation of hexagonal cells (1) occurs in time by progressively entraining more and more molecules in the macroscopic motion and (2) is imperfect, perturbed by random fluctuations and many other factors (constraints). These facts allow us to foresee already that what happens relatively quickly and with success in a simple R-B setting, i.e., the establishment of an optimal regime of free energy degradation, will become dramatically more fluctuating, complex, and hence time-consuming in the case of a setting as wide as the Sun-Earth system. This, of course, will have crucial consequences for macroethics.

IV. THE EARTH AS A GLOBAL FAR-FROM-EQUILIBRIUM DISSIPATIVE STRUCTURE

There is life on Earth. Any theory of macroethics worth its salt must have the resources to give a central role to that simple fact. Floridi’s axiological scale leaves room for such a role, thanks to the overdridability of different levels of informational value. However, this is somewhat unsatisfactory. One would have expected from an informational macroethics, which is based on a technical ontological framework, something like an equation, a formula to measure worth, and thus, in some way, to mechanize morals as Alan Turing’s famous formalism mechanized intelligence. The problem is that, as stated above, the quantity of information cannot serve as a gauge of value since two very different entities may contain the same amount of information. For instance, there may be a configuration of some amount of potatoes that is quantitatively equivalent to, say, an innocent child. What is needed is a criterion able to locate qualitative differences on an axiological scale where they can make a difference.

I suggest that the points raised above about far-from-equilibrium systems and LMEP are in a good position to ground such an axiological scale. Consider the question: What difference does life make? To begin with, living organisms are far-from-equilibrium systems. All the metabolic and adaptive processes of living things are sustained by a continuous energy flow, and their ordered patterns are self-organizing, i.e., dynamically maintained, as in R-B. Existing far from thermodynamic equilibrium means existing beyond the thresholds in magnitude mentioned above. This non-linearity implies that the rate of entropy production in living creatures must differ from that of a non-living entity under the same conditions. This implication has been developed by Ulanowicz R. E., Hannon B. M., who hypothesized that it could be proven that forests, for instance, produce more entropy than the desert under the same electromagnetic field. Meyns & Bruers recently tested the hypothesis that “living communities augment the rate of entropy production over what would be found in the absence of biota, all other things being equal.” Using an ecologically inspired model of entropy production in food webs with predators and preys, they showed that the hypothesis holds every time.

The consequence is that far-from-equilibrium systems such as living creatures on Earth operate according to an adaptive principle. In other words, the structures and dynamic patterns that emerge when crossing critical thresholds are such that they tend to optimize entropy production given the constraints. This means that, for a given potential gradient P and a set of constraints C, there is only a restricted set of patterns—perhaps even a singleton—able to optimize the rate of entropy production. In R-B, as we saw, hexagonal cells do the job, but there is a set of geometrical patterns and dynamic organizations different from hexagons that may possess the same amount of information as the fluid yet dissipate the potential at a slower rate under those same conditions. Therefore, one can assume that LMEP working as a thermodynamical selection principle at the planetary level is ensuring that the living forms that emerge in time are coordinated and increasingly evolving towards higher...
rates of global dissipation of the geo-cosmic potential constituted mainly by the electromagnetic radiation from the Sun in which the Earth is immersed. For instance, chlorophyll is particularly efficient in its capacity to absorb blue and red light, thanks to the structural complementarity between the spatial distribution of its p-orbitals and the wavelengths of blue and red light. In this way, we can see value as depending on the fit between the qualitative aspects of living order and the qualitative aspects of the geo-comic potential taken as dynamic patterns. Moreover, visible light corresponds to more than half of the total solar emission, implying a massive free energy influx to be degraded. There is value in the capacity of photosynthetic organisms to contribute drastically to the degradation of this tremendous amount of free energy.

The idea that the Earth is operating holistically as a maximizer of entropy production is increasingly gaining adepts. Special attention is paid to the link between LMEP or similar characterizations of entropy maximization and evolution. Recently, Martyushev and Seleznev have responded to some claims that LMEP doesn’t generalize well and that it shouldn’t be considered a law of thermodynamics. The authors show that such conclusions are based on the wrong application of LMEP’s predictions without properly assessing some key restrictions. One of those restrictions is, I think, of particular importance for the present discussion about macroethics as it concerns the time delays in thermodynamic processes. As I have already mentioned above, the self-organized emergence of new order in a far-from-equilibrium system is a time-dependent process akin to searching. This means that from the onset of a supra-threshold energetical inflow to the actual assembly of an optimal or near optimal dissipative pattern, the system goes through transient heuristic stages analogous to trial and error. During all that time, the system is obviously performing sub-optimally. Yet, one could argue that, even during the searching period, the system is still performing optimally, since the very state of the system during the long process of assembly counts as a constraint and, thus, given that constraint, the system is still “doing its best.” Although such a view sounds Panglossian, it may actually present the advantage of reconciling the apparent unlawful normativity of ethics with the lawful determinism of LMEP. Indeed, at some level of analysis and from a local vantage point, the system is performing sub-optimally and, hence, something like a norm may help seek ways to improve the situation, for instance, by removing or changing the constraints that keep the system from producing entropy at higher rates. From another point of view, however, the system is working optimally given the constraints and in perfect agreement with lawful determinism. A theory of macroethics capable of naturalizing normativity would present a very strong advantage over other alternatives.

Another restriction linked to the former that needs to be taken into account concerns local maxima. It would seem that value based on the optimization of entropy production should go against all rationality concerning viability and even common sense. After all, as Floridi points out, entropy is metaphysically tantamount to Evil. Would this imply that forests have to be burned as fast as possible, for instance? Certainly not, because every kind of direct, local application of entropy production might contribute to trapping the whole Earth in a local maximum. Consider petroleum, for instance. In a relatively small and homogeneous system such as R-B, the time delay between energetical transactions is very short. If, say, some small regions of the fluid are hotter than other regions, because of the fast moving particles and transmissive medium, such small local gradients are very short lived, and the global bottom-up force overlies them completely, driving the system very quickly to the formation of hexagonal patterns. In a system as vast and complex as the Earth, on the other hand, the local formation and maintenance of gradients is ubiquitous and inevitable. Petroleum represents one such local gradient, which embodies in its chemical structure a significant amount of free energy. This free energy took thousands of years to undergo a transformation from solar energy to living tissue and then to petroleum, and at every step, entropy was produced. However, until mankind started exploiting petroleum globally and industrially, this source of free energy was sub-optimally unaffected, thus missing a great opportunity for entropy production. Does this mean that we should go ahead and deplete the source at the fastest possible rate as we are currently doing, against all the advice from ecologists? Probably not, but not for the reasons ecologists think. The Earth is still undergoing its transformation towards an optimal regime of geo-cosmic energy degradation. This ongoing transformation has been taking millions of years, and it is not likely to stop anytime soon. However, local potentials such as petroleum and the other so-called fossil fuels may present an opportunity for mankind as part of the Earth system to transition into a higher level of dynamic order, which might improve the rate of solar energy degradation. In other words, consuming the local potentials without taking into account the global field might transiently increase overall entropy production (and therefore terrestrial order), yet as soon as the local source is depleted, the Earth would go back to its earlier regime, having missed an opportunity to move closer to the optimal form. This would be tantamount to destroying your car in order to increase entropy immediately and locally instead of keeping your car and using it to go every day to the supermarket and deplete the higher energy sources present there.

V. CONCLUSION. IS ENTROPY ETHICS AN ETHICS OF EVIL?
I have tried to argue that although Floridi’s looks like a move in the right direction to reconceptualize ethics not only as a holistic foundation of value, but also as encompassing more than just humans or living creatures, it falls short of considering all the aspects related to Being. If I am right, Floridi’s appeal to a notion such as entropy and ontological information is fatally incomplete, since, by deciding to “emphatically” detach those notions from their thermodynamical equivalents, he misses an ontologically crucial link between entropy and order. It is crucial in that it shows that, by relocating intrinsic value not on Being but on entropy production, we can still obtain the astonishingly paradoxical result that Being is protected and promoted as Floridi’s own Information Ethics requires.
In this way I have challenged Floridi’s view, suggesting that contemporary thermodynamical research presents us with ineluctable facts that force us to radically reconsider our axiological principles. Floridi’s information/entropy dichotomy doesn’t seem to make room for far-from-equilibrium phenomena where both are entangled and complementary. It is not possible to identify entropy with Evil when the value of order happens to be contingent on its capacity to optimize entropy production given the constraints. The case of the whole Earth’s evolution as a far-from-equilibrium system makes this point conspicuous.

Considering order as the intrinsic source of value has the disadvantage that we cannot establish a non-arbitrary axiological scale. However, if accelerating (global) entropy production becomes the norm, we can see that those terrestrial forms that our common sense already values most get automatically promoted axiologically because they coincide with the forms that tend to contribute to the production of entropy at optimal rates given the specific context established by the potential field in which the Earth is immersed. If the search for Good is the search for the shortest paths to global enteralical degradation, then life and mankind’s extremely complex cultures and technological achievements get instantly promoted as optimal media for that end. That is because those kinds of entities and processes happen to fit better the structure of the geo-cosmic potential while satisfying the constraints.

In addition, despite being based on a deterministic physical law, the approach presented here leaves plenty of room for human intervention, normativity, and, therefore, responsibility. Indeed, the search for the optimal forms capable of dissipating the geo-cosmic potential at the fastest rate is extremely long and haunted by local maxima where the Earth can get trapped at every moment. Humans are the only entities in the system that have access to distal, higher-order constraints that modulate the overall rate of entropy production at least at the scale of the Sun-Earth system. Yet, because they are also constantly embedded in local gradients, humans also have a tendency to favor the depletion of those more proximal gradients, and the tendency is becoming exponentially stronger with trends such as technological improvement and overcrowding. For this reason, a macroethics theory is needed more than ever, yet it needs to embrace all aspects of reality, including, ironically, entropy itself.

NOTES
13. Ibid., 335.

REFERENCES
Taking the Intentional Stance Toward Robot Ethics

James H. Moor
DARTMOUTH COLLEGE


I wish to defend the thesis that robot ethics is a legitimate, interesting, and important field of philosophical and scientific research. I believe it is a coherent possibility that one day robots will be good ethical decision-makers at least in limited situations and act ethically on the basis of their ethical understanding. Put another way, such envisioned future robots will not only act according to ethical principles but act from them.

This subject goes by various names such as “robot ethics,” “machine ethics,” or “computational ethics.” I am not committed to any particular term, but I will here use “robot ethics” as it suggests artificial agency. I do not exclude the possibility of a computer serving as an ethical advisor as part of robot ethics, and I include both software and hardware agents as candidates for robots.

KINDS OF ETHICAL ROBOTS

Agents, including artificial agents, can be understood as ethical in several ways. I distinguish among at least four kinds of ethical agents (Moor 2006). In the weakest sense ethical impact agents are simply agents whose actions have ethical consequences whether intended or not. Potentially any robot could be an ethical impact agent to the extent that its actions cause harms or benefits to humans. A computerized watch can be considered an ethical impact agent if it has the consequence of encouraging its owner to be on time for appointments. The use of robotic camel jockeys in Qatar has the effect of reducing the need for slave boys to ride the camels.

Implicit ethical agents are agents that have ethical considerations built into their design. Typically, these are safety or security considerations. Planes are constructed with warning devices to alert pilots when they are near the ground or when another plane is approaching on a collision path. Automatic teller machines must give out the right amount of money. Such machines check the availability of funds and often limit the amount that can be withdrawn on a daily basis. These agents have designed reflexes for situations requiring monitoring to ensure safety and security. Implicit ethical agents have a kind of built in virtue—not built from habit but from specific implementations in programming and hardware.

Unethical agents exist as well. Moreover, some agents can be ethical sometimes and unethical at others. One example of such a mixed agent I will call “the Goodman agent.” The Goodman agent is an agent that contains the millennium bug. This bug was generated by programming yearly dates using only the last two digits of the number of the year resulting in dates beyond 2000 being regarded as existing earlier than those in the late 1900s. Such an agent was an ethical impact agent before 2000 and an unethical impact agent thereafter. Implicit unethical agents exist as well. They have built in vice. For instance, a spam zombie is an implicit unethical agent. A personal computer can be transformed into a spam zombie if it is infected by a virus that configures the computer to send spam e-mail to a large number of victims.

Ethical impact agents and implicit ethical agents are ethically important. They are familiar in our daily lives, but there is another kind of agent that I consider more central to robot ethics. Explicit ethical agents are agents that can identify and process ethical information about a variety of situations and make sensitive determinations about what should be done in those situations. When principles conflict, they can work out resolutions that fit the facts. These are the kind of agents that can be thought of as acting from ethics, not merely according to ethics. Whether robot agents can acquire knowledge of ethics is an open empirical question. On one approach ethical knowledge might be generated through good old-fashioned AI in which the computer is programmed with a large script that selects the kinds of information relevant to making ethical decisions and then processes the information appropriately to produce defensible ethical judgments. Or the ethical insights might be acquired through training by a neural net or evolution by a genetic algorithm. Ethical knowledge is not ineffable and that leaves us with the intriguing possibility that one day ethics could be understood and processed by a machine.

In summary, an ethical impact agent will have ethical consequences to its actions. An implicit ethical agent will employ some automatic ethical actions for fixed situations.
An explicit ethical agent will have, or at least act as if it had, more general principles or rules of ethical conduct that are adjusted or interpreted to fit various kinds of situations. A single agent could be more than one type of ethical agent according to this schema. And the difference between an implicit and explicit ethical agent may in some cases be only a matter of degree.

I distinguish explicit ethical agents from full ethical agents. Full ethical agents can make ethical judgments about a wide variety of situations and in many cases can provide some justification for them. Full ethical agents have those metaphysical features that we usually attribute to ethical agents like us, features such as intentionality, consciousness, and free will. Normal adult humans are our prime examples of full ethical agents. Whether robots can become full ethical agents is a wonderfully speculative topic but not one we must settle to advance robot ethics. My recommendation is to treat explicit ethical agents as the paradigm example of robot ethics. These potential robots are sophisticated enough to make them interesting philosophically and important practically. But not so sophisticated that they might never exist.

An explicit ethical robot is futuristic at the moment. Such activity is portrayed in science fiction movies and literature. In 1956, the same year of the Summer Project at Dartmouth that launched artificial intelligence as a research discipline, the movie "Forbidden Planet" was released. A very important character in that movie is Robby, a robot that is powerful and clever. But Robby is merely a robot under the orders of human masters. Humans give commands and he obeys. In the movie we are shown that his actions are performed in light of three ethical laws of robotics. Robby cannot kill a human even if ordered to do so.

Isaac Asimov had introduced these famous three laws of robotics in his own short stories. Asimov’s robots are ethical robots, the kind I would characterize as explicit ethical agents. They come with postmodern brains that are imbued with the laws of robotics. Those who are familiar with Asimov’s stories will recall that the three laws of robotics appear in the Handbook of Robotics, 5th Edition, 2058 A.D. (Asimov 1991):

1. A robot may not injure a human being, or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Asimov’s robots are designed to consult ethical guidelines before acting. They are kind and gentle robots compared to the terrifying sort that often appears in books and movies. Asimov’s ethical laws of robotics seem reasonable at least initially, but, if pursued literally, they are likely to produce unexpected results. For example, a robot, which we want to serve us, might be obligated by the first law to travel into the world at large to prevent harm from befalling other human beings. Or our robot might interfere with many of our own plans because our plans for acting are likely to contain elements of risk of harm that needs to be prevented on the basis of the first law.

Although Asimov’s three laws are not adequate as a system of ethics for robots, the conception that Asimov was advancing seems to be that of a robot as an explicit ethical agent. His robots could reason from ethical principles about what to do and what not to do. His robots are fiction but they provide a glimpse of what it would be like for robotic ethics to succeed.

EVALUATING EXPLICIT ETHICAL ROBOTS
I advocate that we adopt an empirical approach to evaluating ethical decision making by robots (Moor 1979). It is not an all or nothing matter. Robots might do well in making some ethical decisions in some situations and not do very well in others. We could gather evidence about how well they did by comparing their decisions with human judgments about what a robot should do in given situations or by asking the robots to provide justifications for their decisions, justifications that we could assess. Because ethical decision making is judged by somewhat fuzzy standards that allow for disagreements, the assessment of the justification offered by a robot for its decision would likely be the best and most convincing way of analyzing a robot’s ethical decisions competence. If a robot could give persuasive justifications for ethical decisions that were comparable to or better than that of good human ethical decision makers, then the robot’s competence could be inductively established for a given area of ethical decision making. The likelihood of having robots in the near future that are competent ethical decision makers over a wide range of situations is undoubtedly small. But my aim here is to argue that it is a coherent and defensible project to pursue robot ethics. In principle we could gather evidence about their ethical competence.

Judging the competence of a decision maker is only part of the overall assessment. We need also to determine whether it is appropriate to use the decision maker in a given situation. A robot may be competent to make a decision about what some human should have for her next meal. Nevertheless, she would probably justifiably wish to decide for herself. Therefore, a robot could be ethically competent in some situations in which we would not allow the robot to make such decisions because of our own values. With good reason we usually do not allow other adults to make ethical decisions for us, let alone allow robots to do it. However, it seems possible there could be specific situations in which humans were too biased or incompetent to be fair and efficient. Hence, there might be a good ethical argument for using a robotic ethical decision maker in their place. For instance, a robotic decision maker might be more competent and less biased in distributing assistance after a national disaster like the hurricane Katrina that destroyed much of New Orleans. In the Katrina case the human relief effort was incompetent. The coordination of information and distribution of goods was not handled well. In the future ethical robots might do a better job in such a situation. Robots are spectacular at tracking large amounts of information and
could communicate with outlets to send assistance to those who need it immediately. These robots might at some point have to make triage decisions about whom to help first, and they might do this more competently and fairly than humans. Thus, it is conceivable there could be persuasive ethical arguments to employ robot ethical decision makers in place of human ones in selected situations.

THE INTENTIONAL STANCE
I have selected robots that are explicit ethical agents as the interesting class of robots for consideration in robot ethics. Of course, if robots one day become persons and thereby full ethical agents, that would be even more interesting. But that day is not likely to come in the foreseeable future, if at all. Nonetheless, explicit ethical agents, though not full ethical agents, could be quite sophisticated in their operations. We might understand them by regarding them in terms of what Daniel Dennett calls "the intentional stance" (Dennett 1971). In order to predict and explain the behavior of complex computing systems, it is often useful to treat them as intentional systems. To treat them as if they were rational creatures with beliefs and desires pursuing goals. As Dennett suggests, predicting and explaining computer behavior on the basis of the physical stance using the computer's physical makeup and the laws of nature or on the basis of the design stance using the functional specifications of the computer's hardware and programming is useful for some purposes such as repairing defects. But predicting and explaining the overall behavior of computer systems in terms of the physical and the design stances is too complex and cumbersome for many practical purposes. The right level of analysis is in terms of the intentional stance.

Indeed, I believe most computer users often take the intentional stance about a computer's operations. We predict and explain its actions using the vocabulary of beliefs, desires, and goals. A word processing program corrects our misspellings because it believes we should use different spellings and its goal is to correct our spelling errors. Of course, we need not think the computer believes or desires in the way we do. The intentional stance can be taken completely instrumentally. Nevertheless, the intentional stance is useful and often an accurate method of prediction and explanation. That is because it captures in a rough and ready way the flow of the information in the computer. Obviously, there is a more detailed account of what the word processing program is doing in terms of the design stance and then at a lower level in terms of the physical stance. But most of us do not know the details nor do we need to know them in order to reliably predict and explain the word processing program’s behavior. The three stances (intentional, design, and physical) are consistent. They differ in level of abstraction.

We can understand robots that are explicit ethical agents in the same way. Given their beliefs in certain ethical principles, their understanding of the facts of certain situations, and their desire to perform the right action, they will act in such and such ethical manner. We can gather evidence about their competence or lack of it by treating them as intentional systems. Are they making appropriate ethical decisions and offering good justifications for them? This is not to deny that important evidence about competence can be gathered at the design level and the physical level. But an overall examination and appreciation of a robot’s competence is best done at a more global level of understanding.

WHY NOT ETHICAL ROBOTS NOW?
What prevents us from developing ethical robots? Philosophically and scientifically is the biggest stumbling block metaphysical, ethical, or epistemological?

Metaphysically, the lack of consciousness in robots seems like a major hurdle. How could explicit ethical agents really do ethics without consciousness? But why is consciousness necessary for doing ethics? What is crucial is that the robot receives all of the necessary information and processes it in an acceptable manner. A chess playing computer lacks consciousness but plays chess. What matters is that the chess program receives adequate information about the chess game and processes the information well so that by and large it makes reasonable moves.

Metaphysically, the lack of free will would also seem to be a barrier. Don’t all moral agents have free will? For sake of argument let’s assume that full ethical agents have free will and robots do not. Why is free will necessary for acting ethically? The concern about free will is often expressed in terms of a concern about human nature. A common view is that humans have a weak or base nature that must be overcome to allow them to act ethically. Humans need to resist temptations and self-interest at times. But why do robots have to have a weak or base nature? Why can’t robots be built to resist temptations and self-interests when it is inappropriate? Why can’t ethical robots be more like angels than us? We would not claim a chess program could not play championship chess because it lacks free will. What is important is that the computer chess player can make the moves it needs to make in the appropriate situations as causally determined as those moves may be.

Ethically, the absence of an algorithm for making ethical decisions seems like a barrier to ethical robots. Wouldn’t a computer need an algorithm to do ethics (Moor 1995)? Let us assume there is no algorithm for doing ethics, at least no algorithm that can tell us in every situation exactly what we should do. But, if we act ethically and don’t need an algorithm to do it, we do it in some way without an algorithm. Whatever our procedure is to generate a good ethical decision, why couldn’t a robot have a similar procedure? Robots don’t have to be perfect to be competent any more than we do. Computers often have procedures for generating acceptable responses even when there is no algorithm to generate the best possible response.

Ethically, the inability to hold the robot ethically responsible seems like a major difficulty in pursuing robot ethics. How would we praise or punish a robot? One possibility is that robots might learn like us through some praise or punishment techniques. But a more direct response is that ethical robots that are not full ethical agents would not have rights, and could be repaired. We could hold them causally responsible for their actions and then fix them if they were malfunctioning so they act better in the future.
Epistemologically, the lack of ability of robots to have empathy for humans would lead them to overlook or not appreciate human needs. This is an important insight as much of our understanding of other humans depends on our own emotional states. Of course, we might be able to give robots emotions, but short of that we might be able to compensate for their lack of emotions by giving them a theory about human needs including behavioral indicators for which to watch. Robots might come to know about emotions by other means than feeling the emotions. A robot's understanding of humans might be possible through inference if not directly through emotional experience.

Epistemologically, computers today lack much common sense knowledge. Hence, robots could not do ethics, which so often depends upon common sense knowledge. This is probably the most serious objection to robot ethics. Computers work best in well-defined domains and not very well in open environments. But robots are getting better. Autonomous robotic cars are adaptable and can travel on most roads and even across open deserts and through mountain tunnels when given the proper navigational equipment. Robots that are explicit ethical agents lacking common sense knowledge would not do as well as humans in many settings but might do well enough in a limited set of situations. In some cases, such as the example of the disaster relief robot, that may be all that is needed.

CONCLUSION
We are some distance from creating robots that are explicit ethical agents. But this is a good area to investigate scientifically and philosophically. Aiming for robots that are full ethical agents is to aim too high at least for now, and to aim for robots that are implicit ethical agents is to be content with too little. As robots become increasingly autonomous, we will need to build more and more ethical considerations into them. Robot ethics has the potential for a large practical impact. In addition, to consider how to construct an explicit ethical robot is an exercise worth doing for it forces us to become clearer about what ethical theories are best and most useful. The process of programming abstract ideas can do much to refine them.

REFERENCES
(such as artificial limbs, pacemakers, and other devices) are added to a human, the human then becomes a cyborg. The question then becomes, how much of a cyborg is a particular person with particular artificial replacements and enhancements? We can also approach this issue from the other direction. It seems consistent that a robot can be transformed into a cyborg by adding biological parts. Again, how can we quantifiably relate the resulting entity to a human and to a robot?

Being able to somehow measure a distance between a human, a cyborg, and a robot brings up philosophical issues as well as practical issues. A central philosophical issue is how we define personhood; as an entity moves from 100 percent human by replacing or adding mechanical parts, is there a point at which that entity is no longer a person? If there is not such a point, then does that automatically mean that sufficiently sophisticated machines will inevitably be classified as persons? Practical issues include implications in sports, health care, health insurance, life insurance, retirement policies, lawsuits, discrimination, and in software design and implementation for cybernetic devices.

MEASURING THE DISTANCE

If you look at the problem of measuring humans – cyborgs – robots from a purely physical view, it seems logical that there exists a continuum from 100 percent human to 100 percent robot with cyborgs being the transition from one to the other. At the moment, cyborgs typically start out as humans and mechanical parts are added. However, there is nothing in our notion of cyborgs that would theoretically prevent moving in the opposite direction: adding biological parts to robots.

No matter how a cyborg comes into being, we would like to be able to talk about a “distance” from a given cyborg to the ends of the continuum: 100 percent human and 100 percent robot. It’s easy to draw a picture that depicts the basic idea (see Figure 1), but it’s not so easy to define precisely what the distance in this picture means. We contend that we need a metric to mark the distance between a given cyborg to the ends of the continuum: 100 percent human and 100 percent robot. It’s easy to draw a picture that depicts the basic idea (see Figure 1), but it’s not so easy to define precisely what the distance in this picture means. We contend that we need a metric to mark the distance between a given cyborg to the ends of the continuum: 100 percent human and 100 percent robot. If we can define an effective measure for this movement, that measure will have both theoretical and practical significance.

MORE ON CYBORGs

Cyborgs are a popular theme in fiction. The Million Dollar Man and the Bionic Woman, seven of nine from Star Trek, and Detective Spooner in I Robot are well-known fictional characters that are, by our definition, cyborgs.

There are non-fictional cyborgs as well. Scientists Steve Mann and Kevin Warwick, both leading researchers in cybernetics, are also cyborgs because of the mechanical parts they have inside and outside themselves. These parts include RFID chips under the skin and glasses that augment visual reality, an idea now taken up by Google. Another well-known cyborg is Oscar Pistorius, a runner from South Africa. While Steve Mann and Kevin Warwick are cyborgs by choice, Pistorius became a cyborg because he needed a replacement for both his legs below the knees.

Running on his spring steel artificial legs, Pistorius was a successful runner in the Special Olympics. But Oscar wanted to race in the Beijing Olympics. Some of his potential competitors objected, contending that his artificial legs were more efficient than human legs, and that therefore Pistorius would have an unfair advantage. Oscar sued and the courts overruled the Olympic Committee. Pistorius was allowed to compete for a spot on the South Africa Olympic team.

Pistorius, nicknamed “the Blade Runner,” is a classic case of a cyborg: part human, and part mechanical. Pistorius’s legs are not automated but they are artificial. They are also “attached” rather than being internal. However, especially when he is competing, Pistorius is not 100 percent carbon-based human, and he is not 100 percent mechanical.

There are many cyborgs among us today. Artificial parts are becoming increasingly common, and those parts are increasingly sophisticated. They deliver clear advantages to people with impairments and missing limbs. In some cases, these devices are being used to enhance, rather than replace, human functions.

In a recent paper in IEEE Technology and Society, Roger Clarke writes about cyborgs and possible legal issues under the title “Cyborg Rights.” Clarke reviews several aspects of how cyborgs are defined and categorized. The kinds of artifacts used to make a cyborg can be distinguished by their intent: Are they prosthetic, meant to replace missing

![Figure 1. Cyborgs are mixtures of biological human parts and mechanical parts.](image-url)
or diminished functionality; or are they orthotic, meant to enhance normal functioning? Clarke also separates the artifacts by their relationship to the body of the cyborg. Any artifact that is under the skin he calls endo. A cardiac pacemaker and a cochlear implant are endo. An artifact that is attached to the body but not inside the body is labeled exo. Oscar Pistorius’s legs are exo. The third category, external, includes devices that are not inside and not attached to the body, but are still integrated with the human body.9 Eyeglasses, canes, and scuba gear are external.

Clark explores six different kinds of cyborgs using the distinction we’ve already drawn between prosthesis and orthotic, and the distinction among endo, exo, and external. An example of an endo prosthesis is an artificial hip. An example of an orthotic endo device would be a metal plate attached to a bone to make it stronger than a normal human bone. An example of an exo prosthesis is an artificial leg, such as the legs Oscar Pistorius uses. If an artificial leg is somehow mechanically superior to a human leg for some particular function, then it is orthotic instead of a prosthetic, but it is still exo. A pair of eyeglass and a cane are both external prosthesis devices. Devices used to enhance human vision, such as microscopes or telescopes, are external, orthotic devices.

If you agree with Clarke that eyeglasses, contact lenses, and scuba gear all make you a cyborg, then many of us are best classified as cyborgs. This inclusive view of cyborgs is attractive because it emphasizes a broad range of ways in which we can augment ourselves artificially. But since the word “cyborg” came from “cybernetics,” a term popularized by Norbert Weiner with respect to information flows in systems, some scholars may be more comfortable with restricting cyborgs to devices that are electromechanical.

Examples of existing electromechanical devices used with humans are a cardiac pacemaker and embedded RFID chips (common in pets but also used in humans); both of these are endo devices. A robotic hand can be an exo prosthesis or an exo orthotic, depending on how much functionality it delivers. An artificial lung not inside a patient is an external prosthesis device.

In trying to establish a metric for measuring the distance between humans and robots, we found it difficult to include external devices in our initial analyses. External devices can be too temporary, too loosely integrated into the person, to be considered (at least by some) to be well integrated with the original person. When considering the measurement question, external devices led us to some difficulties. For example, including external devices in the definition of “cyborg” made it difficult to precisely distinguish a person using a tool from a cyborg that has integrated a mechanical external part. For now, we will focus on cyborgs made with endo and exo devices; devices attached to the body and embedded under the skin have a clear distinction from tools “at hand.” However, others may wish to be more inclusive than we have been and explore alternative metrics.

**CANDIDATES FOR MEASURING THE DISTANCE**

Now that we have clarified the scope of what we will consider as qualifying devices to move someone along the cyborg continuum, we will focus on our original quest—determining a measure of cyborg-ness. All of the measures will give a reading of 0 percent for 100 percent humans, and give 100 percent for a 100 percent robot.

**CANDIDATE 1: BY WEIGHT**

The first measure is simple, relatively easy to measure, and sadly counter-intuitive. In this strategy, we divide the weight of a cyborg’s mechanical parts by the total weight of the cyborg, and multiply that ratio by 100. If you have a hip replacement, that moves you to the right from 0 percent. A heart valve replacement moves you further. If someday you add an artificial memory implant or artificial lungs, those changes would move you further still on the continuum.

As with all the measures we will consider, there are some good and bad aspects to the weight metric. For this weight-based measure, one advantage is that you can determine this measure precisely in an objective, straightforward way. A significant disadvantage is that the weight of a part is intuitively not indicative of the significance of the replacement. A brain weighs about 1.3 kilograms (about three pounds). A leg weighs about seven kilograms (about fifteen pounds). Realistically most people do not think a leg is five times as important as the brain, but according to the weight measure, replacing a leg with a mechanical device would place you more towards the robot side of the continuum than replacing your brain with a mechanical device.

**CANDIDATE 2: BY INFORMATION FLOW**

The next measure considered is an information-centric measure. For this measure, the information flow into and out of each replacement part in the cyborg is determined. We also measure all the information flow in the whole body. The ratio gives us our position on the continuum. One good aspect of this measure is that, in theory, it could be measured and information flow could potentially relate to significance. (For example, we might be able to approximate the number of bits necessary to contain the same information that travels through the nervous system in and out of the artificial part and/or the biological part replaced.) Also, this measure is consistent with Floridi’s emphasis on information as particularly significant in understanding ethical significance. A problem with the “by weight” measure is that it over-emphasized the importance of mass. The “by information flow” may have a related problem: it perhaps over-
emphasizes the significance of information processing to humans and de-emphasizes other aspects, such as mobility and disease prevention. If the measure was restricted to information flow in and out of the brain, that might make it more practical to measure but less sensitive to important aspects of humaness.

CANDIDATE 3: BY FUNCTIONALITY
The third measurement option adds up the functionality contributed by the artificial parts and divides that by the total functionality of the cyborg. This avoids the counter-intuition of the weight measure and allows us to conclude more considerations than information flow. But, as stated, this is a vague measurement and measuring precisely the “amount of functionality” is difficult. We have by no means solved these problems, but looking at attempts to measure medical outcomes of injuries and therapies may offer us directions for future research in measuring “cyborg-ness.”

For many reasons, including insurance payments and scholarly studies about effective treatments, medical professionals seek to quantify patients’ quality of life. This has led to numerous attempts to assign numerical values to a patient’s physical and mental well-being.

One way to attempt a functionality measurement is to adopt or adapt one of the existing systems that helps assign a number to the impairments a patient exhibits, or to the improved condition of a patient. The adaptation would have to isolate and then measure the effect of the cyborg’s improvement due to the endo or exo artifact that was added. These measures attempt to measure both time (increased life span) and quality of life.

A positive aspect of using (or adapting) an existing quality of life measure is there is an extensive body of literature and years of active practice in using these measures. There has been considerable effort to make these evaluations repeatable, consistent, and objective. However, there is not as yet a consensus on any one particular measure, and the objectivity of these measures is an ongoing object of research. A significant problem for our purposes is that the measures now in use necessarily use statistical measures of large groups of patients, and the effect of an integrated artifact may vary significantly between individual cyborgs.

The use of these medical function-based measures for quality of life seems more clearly appropriate for prosthetic artifacts than for orthotic enhancements. For prosthetic devices, a certain function that was human becomes mechanical, and that gives rise to a movement to the right in our diagram, a distance proportional to the percentage of functionality. However, an orthotic enhancement introduces new functionality and perhaps longer life. Seemingly, this is significant for quality of life type measures. However, how do we reflect these improvements in a measure for cyborg-ness? Reflecting these improvements could, for example, push us beyond 100 percent robot, which intuitively makes no sense. Perhaps a way out of this problem is to restrict our measurement to the percentage of functionality of the replaced or enhanced part based on a 100 percent human. But that restriction clearly loses some of the explanatory power of functionality-based measuring.

The intuitive appeal of measuring functionality has convinced us that this area requires more study. However, the complications involved will require the involvement of medical professionals in order to make a more intelligent suggestion about how these measures might be adapted to measuring the distance between humans, cyborgs, and robots.

CONSIDERING CHALLENGES TO THE PROPOSED MEASURES
We are not overjoyed with any of the alternatives listed above; no measure seems ideal. Perhaps some measures can be combined, but that adds unwanted complexity. If the measure is too complex, people will not readily understand it, and fewer people and institutions are likely to use it. We hope, therefore, to be able to use a single measure if at all possible.

A challenge for any cyborg measure is trying to include sensitivity towards how an artifact is used by a human, not just what the artifact and its capabilities are. As an example of this challenge, consider what we call the “surrogate problem.” The idea of artificial entities under the control of human operators was explored by a recent movie called Surrogates, and in a different way by the more commercially successful (but less well explained scientifically) movie called Avatar. The surrogate problem for our cyborg metric is that the degree to which external devices are used by the humans as a substitute for life without the surrogate should be a significant factor in our measure. A person who spends almost every waking hour living “through the surrogate” seems clearly more of a cyborg than a person who uses a very similar device, but uses the device sparingly, a few minutes a day. The purpose and duration of these surrogate sessions should help determine our measure of cyborg-ness, not just the artifacts themselves. Our measures above, which exclude external devices, do not wrestle with this issue, but it clearly is an issue worthy of further study.

The surrogate problem is related to another problem with external devices, which we call the “puppet problem.” We referred previously to the problem of distinguishing between a tool and an external integrated artifact. One prominent existing example of what might be classified as an orthotic external device is a Predator military drone. These drones are already ethically and legally controversial and thinking of the operators and drones collectively as a cyborg adds another layer of complexity to the ethical questions. Human/machine collaborations (like a physical puppet) become more complex and more significant when the “puppet” has onboard intelligence. Predator drones started out as electronic puppets, controlled in a way that is similar to video games. But as the drones become increasingly sophisticated, they are gaining more and more internal control, and plans have been developed to make them independent from direct human control for longer and longer times. A challenge for adapting a metric to external puppet devices is how to measure the sophistication of an artifact, and the degree of control the human has over the mechanical artifact. The degree of sophistication and independence of an artifact in a “puppet cyborg” seems like
a significant factor in understanding the cyborg as a whole, but artifact sophistication and independence are difficult to quantify.

These questions of measuring puppets also can be applied to endo and exo artifacts. Should not our metric take into account the intelligence and perhaps the “autonomy” of the artifacts that are part of the cyborg? If so, how exactly should this be measured? If not, then aren’t we missing a potentially significant aspect of the cyborg? For example, if a cyborg includes an endo computer that can override the brain’s signals in case of an emergency, or in the case of a defected malfunction in the brain, then that automatic control appears to be an important movement towards the robot side of the continuum that should be included in our measure. None of the measures proposed above is sensitive to this kind of distinction.

A separate challenge has to do with the history of how a cyborg is constructed. We explicitly draw our arrow between humans and robots as a double-headed arrow, but typically the literature discusses a cyborg as a human plus mechanical parts. That is certainly an interesting direction, and it is the direction we are going with many experiments in cyborgs today. But it isn’t the only direction possible. We can also make cyborgs by going in the other direction, placing human parts into robots and making a cyborg that way (see Figure 2). This has interesting ramifications. Our measures above could work for these “left-handed cyborgs,” but people may feel quite differently about a robot with added biological parts than a human with added mechanical parts, even if the measurements of the two resulting entities are identical. On the basis of fairness, we suspect that the measure should not be directly sensitive to the history of the cyborg’s formation, although a measure could be sensitive to different functionalities that resulted from different formations of a cyborg.

**100% HUMAN**  
**100% ROBOT**

*Figure 2. A cyborg could be built by adding biological parts to a robot.*

**FURTHER IMPLICATIONS**

We have only started to touch on several important questions about measuring cyborgs. Once we establish a measure (or measures), that is only the beginning of the philosophical work. With a measure established, you then have to face difficult questions. For example, is there a particular place on your scale where you start to restrict how a cyborg is treated by people or the law? Is there some point at which this someone should not be called “human” anymore? If so, what is that number, exactly? If not, then when all the biological parts are replaced, is the resulting entity (which on our account is now a robot) still a person?

Should voting rights change if you become too close to a robot? If you live hundreds of years as a cyborg, and you vote every year, then you will get many more votes than a human limited to about a century. Is that acceptable, especially for 100 percent humans? How will health care financing change when cyborg replacement parts become a huge part of health spending? Will people who are NOT cyborgs become a distinct minority? If so, humans without artificial enhancements may find themselves living shorter lives than cyborgs, and cyborgs (with increased life span and enhanced capabilities) may routinely surpass non-cyborg humans in business and government; how will society adapt to these kinds of changes? Will non-cyborg humans insist on a legal status separate from cyborgs? What are the ethical and political arguments for and against making this distinction in laws and regulations?

Many scholars are becoming interested in how we are being transformed from a society of humans to a society that combines humans and other intelligent entities. We think that we can make progress in exploring these issues by concentrating on cyborgs. First, unlike other entities being discussed (like robots who could readily pass for humans in a physical encounter), cyborgs are already among us in large numbers. Second, in the foreseeable future it is likely that many of us will be moving towards robots in the human-robot cyborg continuum. Increasing numbers of people will have an immediate, personal stake in these questions about cyborgs.

**CONCLUSIONS**

In this short article we have asked more questions than we’ve answered. But there are several ideas that seem clear:

5. Cyborgs are among us, and the number of cyborgs is likely to increase. The sophistication of the mechanical parts will continue to increase.

6. Software used to control the mechanical parts will become more sophisticated and complex.

7. The idea of a continuum from 100 percent human to 100 percent robots can be a useful notion in our philosophical and ethical analyses, even without settling on a particular metric.

8. Practical problems in making policies, laws, and regulations will not be well served by a theoretical, under-specified continuum. Policy must be spelled out in a way that is unambiguous and precise, so for any given law or regulation that establishes rules based on cyborg-ness, a particular measure (or perhaps measures) will have to be chosen.

9. Despite the many challenges to determining a fair and accurate measure, we are convinced that work should continue on establishing a metric to measure cyborg-ness.

We plan to ask exactly these kinds of questions in our future work.

**APPENDIX A: HUMAN CLONING**

If someday we have human clones, they will seriously complicate our view of a cyborg continuum between
Consider a new continuum where all the entities on the continuum are biological (not mechanical), and where the two extremes are no cloning on the left and 100 percent cloned on the right (see Figure 3). As far as we know there is not a consensus term for humans who have replaced some original parts with cloned parts, so we have marked the middle of this continuum with the phrase “somewhat cloned.”

**Figure 3.** Cloning requires a new scheme for measuring the distance between biological humans and cloned humans.

**BIBLIOGRAPHY**


**Remediation Revisited: Replies to Gaut, Matravers, and Tavinor**

Dominic McIver Lopes

**UNIVERSITY OF BRITISH COLUMBIA**


A Philosophy of Computer Art was conceived of a hunch that thinking about computer art might allow us to come at large and familiar problems in aesthetics and art theory from a new angle. Berys Gaut, Derek Matravers, and Grant Tavinor touch upon some of these large and familiar problems in earlier issues of this Newsletter. One of these is Richard Wollheim’s “bricoleur problem.”

Wollheim asked what makes some stuffs or processes—or “media”—suitable vehicles of art, and he proposed that a solution to this “bricoleur problem” will be largely determined by “analogies and disanalogies that we can construct between the existing arts and the art in question” (1980, 43). In seeking these analogies and disanalogies, we may draw from the “comparatively rich context” of critical and historical discussions, as we did when photography and the movies were new arts (1980, 152).
Critical, historical, and theoretical discussions of digital art typically do root it in precursor art practices and do draw analogies to traditional art while identifying disanalogies that represent reactions against tradition. Going a step further, some theorists hold that this process is itself and by necessity part of digital art practice. A work of digital art is nothing but digitally rendered literature, depiction, film, performance, or music, and so its significance must lie in how it “remediate” these traditional art media by rendering them digitally (Bolter and Grusin 2000). Through remediation, digital art is the art of bricolage.

Running against this grain, *A Philosophy of Computer Art* distinguishes digital art from computer art, whose medium is computer-based interactivity. As Matravers points out, this means that computer art faces a seriously exacerbated bricoleur problem. If interactivity is not a medium in traditional art, then what is the basis for an analogy to computer art?

Every work of computer art has an interface or display made up of text, images, or sound; and perhaps these provide a basis for constructing the comparisons needed to solve the bricoleur problem? Remediation to the rescue after all? Not so fast. The argument in *A Philosophy of Computer Art* assumes that to appreciate a work of computer art for what it is, one must appreciate it, at least in part, for its computer-based interactivity. So we cannot understand why computer-based interactivity is a suitable vehicle for appreciation by seeking analogies between the computer-based interactivity of computer art and the computer-based interactivity of traditional art. There is no computer-based interactivity in traditional art.

Some readers will have noticed a sneaky reformulation of the bricoleur problem as concerning what is a suitability medium for appreciation instead of art. This reformulation is harmless as long as what makes something art is at least in part features of its medium that make it apt for appreciation. Institutional theories of art deny that what makes something art has anything to do with features of its medium for appreciation, but institutional theories of art are inconsistent with the bricoleur problem. Gaut’s argument proceeds first by objecting to the bricoleur problem. If interactivity is not a medium in art. Institutional theories of art are inconsistent with the bricoleur problem because they realize? Perhaps the analogies we need to solve computer art’s acute case of the bricoleur problem are not to be found by comparing interactivity to media like acting, narrative, depiction, tone-meter-timbre structures, and the like. Why should a solution to the bricoleur problem send us in search of analogies at the level of realizers and not at the level of the formal, expressive, and cognitive properties that they realize? Perhaps the analogies we need to solve computer art’s acute case of the bricoleur problem are not to be found by comparing interactivity to media like acting, narrative, depiction, tone-meter-timbre structures, and the like. Simply put, interactivity is a suitable medium for appreciation if interactive works can realize features worth appreciating. This suggestion must fall flat if a solution to the bricoleur problem must tell us how a medium can be a suitable vehicle for art when it is not a suitable vehicle for appreciation. However, as already noted, only institutional theories try to understand what makes for art without appeal to appreciation, and the bricoleur problem does not arise for these theories.

If this thinking is sound, it is possible to solve the bricoleur problem without appeal to remediation. To the extent that the problem pushes theorists to emphasize that digital art is an art of remediation, we now have room to downplay digital art as a theoretical concept and to counterbalance it with the concept of computer art. However, all of this is wheel spinning if computer art is not an art form in the first place.

Gaut argues that there is an art form—call it “computer-based art”—which conjoin computer art and digital art. Gaut’s argument proceeds first by objecting to the argument in *A Philosophy of Computer Art* that digital art is not an art form and then by proposing a feature, automated algorithmic processing, which is medium for all computer-based art—for computer art and digital art alike.

Tavinor suggests another solution to the bricoleur problem, in discussing why the artstic aspects of video games involve interactivity. Some games (e.g., checkers) have no representational elements, some games (e.g., chess) have interactive and representational elements that are completely independent of each other, but in most video games (e.g., The Sims) representation and interactive gameplay are inseparable. One appreciates The Sims for how its little dramas are realized through interaction: the interaction is what it is only given the representational elements and the representation is what it is only given the interaction. So, in trying to understand why video games are suitable vehicles for appreciation, why not draw analogies between drama-realized-interactively and drama-realized-by-actors-following-a-script? And if video games are the popular end of computer art, then this proposal solves the bricoleur problem for computer art.

The proposal can be generalized in a way that makes it clear that remediation has not snuck in the back door. We appreciate works for such formal, expressive, and cognitive properties as having balance, being sad, and bringing out how none of us are free of gender bias. In different arts, these are realized in different ways—by acting, narrative, depiction, tone-meter-timbre structures, and the like. Why should a solution to the bricoleur problem send us in search of analogies at the level of realizers and not at the level of the formal, expressive, and cognitive properties that they realize? Perhaps the analogies we need to solve computer art’s acute case of the bricoleur problem are not to be found by comparing interactivity to media like acting, narrative, depiction, and tone-meter-timbre structures, but rather by comparing the formal, expressive, and cognitive achievements of interactivity alongside those of acting, narrative, depiction, and tone-meter-timbre structures. Simply put, interactivity is a suitable medium for appreciation if interactive works can realize features worth appreciating.

This suggestion must fall flat if a solution to the bricoleur problem must tell us how a medium can be a suitable vehicle for art when it is not a suitable vehicle for appreciation. However, as already noted, only institutional theories try to understand what makes for art without appeal to appreciation, and the bricoleur problem does not arise for these theories.

If this thinking is sound, it is possible to solve the bricoleur problem without appeal to remediation. To the extent that the problem pushes theorists to emphasize that digital art is an art of remediation, we now have room to downplay digital art as a theoretical concept and to counterbalance it with the concept of computer art. However, all of this is wheel spinning if computer art is not an art form in the first place.

Gaut argues that there is an art form—call it “computer-based art”—which conjoins computer art and digital art. Gaut’s argument proceeds first by objecting to the argument in *A Philosophy of Computer Art* that digital art is not an art form and then by proposing a feature, automated algorithmic processing, which is medium for all computer-based art—for computer art and digital art alike.

Here is the argument in the book for the claim that digital art is not an art form. An art kind is an art form only if we normally appreciate any work in the kind by comparison with arbitrarily any other work in the kind. Digital art is an
art kind but we do not appreciate any given work of digital art with arbitrarily any other work of digital art. Therefore, digital art is not an art form.

Gaut doubts whether the major premise of this argument can individuate art forms. Rembrandt self-portraits and paleolithic cave paintings are pictures but we do not normally compare Rembrandt self-portraits with cave paintings. Fair enough, there is some truth in that. Bear in mind two points, however.

First, the claim is not that we consciously or actively compare any given work in an art form with each and every other work in the art form. Rather, the claim is that the KS are the works whose comparison class does not exclude any K. Appreciating a Rembrandt self-portrait as a Rembrandt self-portrait does exclude cave paintings, but appreciating a Rembrandt self-portrait as a picture does not exclude cave paintings. An appreciation of the self-portrait that would have gone differently were cave paintings included cannot be an appreciation of the self-portrait as a picture—it must be an appreciation of it as something narrower—a seventeenth-century painting, perhaps. In the book, the same idea is expressed by saying that an art form is “a group of works that share a distinctive feature in common and that are normally appreciated partly for having that feature” (Lopes 2009a, 18).

Second, the “normally” requires a word of explanation. It is possible to appreciate a K as a K* (Lopes 2008). For example, it is possible to appreciate a building as a sculpture, though buildings are not sculptures, and it is also possible to appreciate a building as an antelope, though it would probably not come off very well (it depends on the building!). However, what makes architecture an art form is that buildings are works of art and there is a norm to appreciate them as buildings. Works made on Tuesdays are an art kind and it is possible to appreciate a work as a Tuesday work, but there is no norm to appreciate anything as a work made on a Tuesday, and that is why Tuesday works are not an art form.

There is no norm to appreciate digital art works as digital art because we do not in fact appreciate them as digital art, though we do appreciate them as digital songs, photographs, and the like. Have you ever appreciated a digital song in comparison to arbitrarily any digital photograph? If you do appreciate 1234 as a work of digital art, then you would have no more reason to exclude Jeff Wall’s A Sudden Gust of Wind from its comparison class as you would have to exclude any other digital song.

The examples of David Cope’s EMI and Harold Cohen’s AARON do double duty in Gaut’s argument. EMI and AARON output (non-interactive) works that we appreciate as products of automated algorithmic processing. As a result, they seem to counter the claim that we never appreciate any given work of digital art in a comparison class with arbitrarily any other work of digital art.

Gaut is right to say that we can and do appreciate the works output by EMI and AARON in comparison with one another. AARON’s drawings are more original but less impressive formally than EMI’s compositions because of the algorithms and databases that each employs. This admission is consistent with the argument against the proposition that digital art is an art form so long as the case of AARON and EMI is not like that of 1234 and A Sudden Gust of Wind. The question is whether there is room to allow for the appreciation of AARON’s drawings alongside EMI’s songs without having to squeeze A Sudden Gust of Wind into the same art form as 1234. There is if computer art has a sister art form, “generative art,” wherein algorithms are run on electronic computers to output new works of art (see also Andrews 2009). On this proposal, AARON and EMI output generative art. But A Sudden Gust of Wind and 1234 are not works of generative art; they belong instead to digital images and digital music, which are genres of the traditional arts of depiction and music.

This discussion is not, as it might seem, empty taxonomizing, for it brings us face to face with the bricoler problem, whence it leads us into fundamental questions of value in the arts and the role of media in realizing that value. Gaut, Matravers, and Tavinor raise plenty of other issues besides these that merit study and dialogue. A Philosophy of Computer Art was never to be the last word on its topic, but rather the first.

REFERENCES

FROM THE EDITOR: NEWSLETTER HIGHLIGHTS

Peter Boltuc
UNIVERSITY OF ILLINOIS–SPRINGFIELD
THE WARSAW SCHOOL OF ECONOMICS

This note is to be read in addition to the table of contents (or, even better, the actual contents) of the main part of the current issue. Here I mention several articles that were pre-selected but did not make it due to the lack of space. Most of them were just too long to be re-published. For instance,
we selected the shorter article by Hintikka (2011) and not one of his last written papers (Hintikka 2013), and we had to skip the posthumously published paper by Pollock, which is over fifty years old. In the case of publications by Franklin et al., we republished their philosophically important response to his critics, while not being able to publish their original article (Franklin et al. 2008), or the actual commentaries. Also, in the case of Lopes we were able to publish his 2010 commentary, while not publishing the 2009 Précis or any of the commentaries. Where we have published commentaries, we had to balance between the best and the shortest. While the additional list does not do justice to all of our important papers, especially those that seem to be aging well.

On a lighter note, I end this issue with the editor’s choice of the most important paper, and most important issue—I have surprised myself on both counts.

I would like to turn our readers’ attention to the following:

**TOP PAPERS FOR GENERAL PHILOSOPHERS**

*not anthologized in this issue*

- **Chrisley, Ron, and Aaron Sloman.** "Functionalism, Revisionism, and Qualia," 16, no. 1 (2016): 2–12.

**TOP PAPERS IN THEORETICAL COMPUTER SCIENCE**

- **Tani, Jun, White, Jeff.** "From Biological to Synthetic Neuromorphic Approaches to Understanding the Structure Essential to Consciousness" (Part 3), 17, no. 1 (2017): 11–22 (Three Parts, Part 3 crucial).
- **Thaler, Stephen L.** "The Creativity Machine Paradigm: Withstanding the Argument from Consciousness," 11, no. 2 (2012): 19–30 (the more fundamental presentation of Thaler’s cognitive architecture than his 2019 novelty article we were able to publish).

**TOP PAPERS IN COMPUTER ETHICS**


**TOP PAPERS IN COMPUTER ART**


**TOP PAPERS ON E-LEARNING IN PHILOSOPHY**

- **Barnette, Ron.** "Reflecting Back Twenty Years," 10, no. 2 (2011): 20–22 (among several pioneering papers in various issues).

**TOP PHILOSOPHICAL CARTOONS**

Half a dozen of Ricard Manzotti’s philosophical cartoons in various recent issues.

**EDITOR’S CHOICE**

Let me end with a few subjective reflections—I let myself pick the editor’s choice of the article and of the issue after a thorough re-reading of the whole newsletter, from 2001 to 2019. I did surprise myself by the choice, in both instances.

**EDITOR’S CHOICE: PAPER**

Having re-read the articles we have published during those years, I think that the most relevant paper for philosophy of AI published in the newsletter may have been a discussion-piece by Stan Franklin, with Bernard Baars and
Uma Ramamurthy, “Robots Need Conscious Perception: A Reply to Aleksander and Haikonen,” 9, no. 1 (2009): 13–15 (reprinted in the current issue). This short piece seems philosophically even more pointedly relevant than their original article (“A Phenomenally Conscious Robot,” 8, no. 2 [2008]). The paper is a response to two criticisms launched from the opposite standpoints: Igor Alexander’s non-reductive phenomenal experience approach and Pentti Haikonen’s view that machines cannot really have phenomenal consciousness.

This also relates to Haikonen’s more robust critique of the Global Workspace Theory, arguing that there is no unified workspace (or blackboard) where all subsystems in the brain communicate. Haikonen makes this argument in his critique of Shannahan’s position. Haikonen also argues that such full inter-communicability may be a serious flaw in advanced AI. Franklin’s main paper provoked a separate discussion with Gilbert Harman. Recently, Franklin and his research team shared with us details on recent developments in LiDA cognitive architecture.

EDITOR’S CHOICE: ISSUE

Another surprise. Although I am really proud of the original papers by Harman, Rapaport, Moor, Dodig-Crnkovich, and Miller in the first issue I (guest) edited (Vol. 6, no. 2, spring 2007), it turns out that the best issue is Vol. 9, no. 1, fall 2009, with two articles that we republished:


The editor’s choice paper by Franklin, Baars, and Ramamurthy also belongs to this issue (see above), as does the response by Haikonen.

Of at least equal value is the longer, featured article of that issue by Raymond Turner, “The Meaning of Programming Languages” (pp. 2–7).

“A Semantics for Virtual Environments and the Ontological Status of Virtual Objects” (pp. 15–19), by David Leech Anderson, one of the most active former members of this committee, could have been the centerpiece of most other issues.

I should also mention Robert Arp’s paper on “Realism and Antirealism in Informatics Ontologies” (pp. 19–22).

Last but not least, the issue contains a block of pioneering articles with feminist perspectives on online education with papers: “Women Don’t Blog” by H. R. Baber and “Gender and Online Education” by Margaret A. Crouch (both former members of this committee).

MORE INFORMATION ABOUT THIS NEWSLETTER

(AN OUTSIDE CHAT):

More about this newsletter can be found at:


My note in the previous issue (fall 2019) also contains a thorough note about the newsletter’s older history.

TOWARDS THE FUTURE

The committee hopes that this newsletter will be available free of charge on the APA website indefinitely. Preferably, it would be good if this (and perhaps the other newsletters) would be available on the basis of creative commons—I understand from the APA that the main problem with this is limited bandwidth. If not, perhaps the newsletter could be mirrored on other sites for the purposes of research, teaching, and intellectual pleasure.

Free access forever was the hope of many of our authors, including John Pollock, who has asked Terry Horgan to publish his deathbed paper wherever it would be most reliably stored and freely available to all. The same goes for Jaakko Hintikka who entrusted two of his last articles to Dan Kolak (one of the former chairs of this committee) with a very similar request.

In terms of editorial plans, the journal Philosophies is interested in hosting a yearly block of articles on Philosophy and Computers, so this may be one avenue for our proven and new authors. Feel free to email me at pboltu@sgh.waw.pl in this matter.

REPORT FROM THE CHAIR

Peter Boluc
UNIVERSITY OF ILLINOIS SPRINGFIELD/WARSAW SCHOOL OF ECONOMICS
PBOLTU@SGH.WAW.PL

On July 1, 2019, I began my one-year term as the final chair of the APA Committee on Philosophy and Computers; by decision of the APA, the committee is one of the three ‘philosophy and...’ committees to be sunsetted mid-2020.

During its last year the committee decided to take up the following tasks:

1. To make deep changes to the committee’s charge so as to reflect the committee’s actual function and the approval of the new charge by the Board of the APA. This was for two reasons: to make sure that the committee’s role was well understood, and to lay a clear path for the new affiliate organization.
The committee approved the change of the charges of the committee, which has led to the mission statement for the new association. To our pleasant surprise, with support from Amy Ferrer, the APA board approved the new charge in November 2019, despite the fact that the committee was to function only for about seven months after that.

The new affiliate organization (see point 5 below) adopted as its own the committee’s new mission statement, with minor changes.

2. To award the Barwise Prize for 2019 and 2020. Also, to make sure that people familiar with the area of computing and philosophy would be involved in awarding the future Barwise Prizes.

The committee awarded the 2019 Barwise Prize to Professor Margaret Boden (University of Sussex) and the 2020 Barwise Prize to Professor Aaron Sloman (University of Birmingham). A note from Aaron Sloman appears below this report.

As a result of discussions with the committee, the APA agrees that at least one former Barwise Prize winner and at least one representative of the new affiliate organization, being a follow-up on this committee (provided such exists and functions well), would be joining the APA’s selection committee in their role as full members of that committee insofar as the selection of future Barwise Prize winners.

3. To organize one or two sessions of the Committee on Philosophy and Computers at all three APA divisional meetings during the 2019-2020 academic year.

The committee proposed one session at the 2020 Eastern Division meeting (organized by Daniel Susser) and two sessions at the 2020 Central Division meeting (organized by Susan Sterrett and Gary Mar, respectively). There were also two sessions proposed and accepted by the Pacific Division for its 2020 meeting—the Barwise Prize lecture by Gualtiero Piccinini (the 2019 winner), and a session organized by Josha Bach and Peter Boltuc devoted to machine consciousness. Those sessions did not take place due to cancellation of the Pacific Division meeting; we hope that those sessions will be re-scheduled for the Pacific Division meeting in 2021.

4. To secure the legacy, and some level of continuity, to the APA Newsletter on Philosophy and Computers.

As the table of contents of the current issue indicates, the APA Newsletter on Philosophy and Computers has published original articles by luminaries of our discipline (e.g., two original articles by Hintikka). Such papers were made available to us under presumption that they are, and would remain, available free of charge to the world philosophical community. It would be highly unfortunate if those were to be lost, or made password-protected and limited to the small group of paying customers.

We have an understanding with the current APA leadership that those issues would remain accessible on the APA website indefinitely; however, the current leadership cannot prejudge decisions of the follow-up boards. Thus, we proposed that this, and other, APA Newsletters be made open access-based on Creative Commons. The APA executive director, Amy Ferrer, indicated that the issue is worth taking up with the new board of officers later in 2020. In particular, the new organization (APC, see point 5 below) would like to mirror those issues on our website (not just link to them).

In terms of continuity, we have an offer, addressed to Peter Boltuc, to host a permanent section devoted to philosophy and computing within the online journal, Philosophies (MDPI). This is a good offer, although some of the articles would have to incur processing fees. The APC decided to avoid engaging in recurring costs of publication, which resulted in putting aside some offers that required such payments of the new association.

5. Create a follow-up organization, based on the APA’s initiative to award it a status of APA affiliate organization.

This last task has taken the longest. The new organization took shape gradually and as of May and early June became The Association for Philosophy and Computing (APC).

A note from APC Secretary Robin Hill, discussing the main aims and structure of the APC, follows this report. I just want to mention that, despite some level of similarity among the names, APC is not in the same niche as the IACAP (International Association for Computing and Philosophy). First, in the latter’s name “Computing” comes first, while APC primarily focused with philosophy. More importantly, we do not aim at being “international” (although we have two or three international members in the initiating committee); we are primarily an APA affiliate with a programmatically limited scope of operations.

A more thorough version of my report should be available, and posted on the APA website, by September 2020.

NOTE FROM THE 2020 BARWISE PRIZE WINNER

My Philosophy in AI: A Very Short Set of Notes Towards my Barwise Prize Acceptance Talk

Aaron Sloman
UNIVERSITY OF BIRMINGHAM

In a sense, everything I’ve done since about 1959 has been part of, or a continuation of, my attempt to defend Kant’s philosophy of mathematics in my 1962 thesis (after converting from maths to philosophy because I disagreed with what I heard philosophers saying about mathematics—while thinking Kant had got it right, but with a lot of gaps in the details). I don’t know how I first encountered Kant while still a maths student. I suspect it was a result of a philosophy graduate student lending me a copy of Russell’s History of Western Philosophy because he thought...
I needed help clarifying my arguments with philosophers and theologians!

Russell summarized and then criticized Kant, but I felt Kant was basically correct, so after reading more of his work I switched to philosophy to defend his view of mathematical discovery, finishing my DPhil in 1962; I am still trying to develop some of the key ideas!

I started learning about AI and programming only around 1969 (from Max Clowes) and then resolved to demonstrate how Kant was right by showing how to program a baby robot that could “grow up” to be a mathematician making discoveries like Archimedes, Zeno, and others. The Computer Revolution in Philosophy (1978) was an attempt to do some of the groundwork for that project, but not using logic-based AI as McCarthy and Hayes had proposed in 1969. The project keeps growing as I learn more, especially as a result of talking to biologist Jackie Chappell since about 2004, which led to the Meta-Configured genome idea, still developing...

I know that my ability to write is going downhill as a result of age (83): I have to spend too much time looking things up that I previously had at my fingertips. Half asleep now. I’ll get back to you later—and once again, many thanks!

Aaron

P.S. About receiving the Barwise Prize: I am particularly amazed because I have had the impression that professional philosophers, with very few exceptions, seem to regard my work as insignificant. So I am delighted to hear that enough of the committee members thought otherwise.

P.P.S. I recently stumbled across the work of another Barwise Prize recipient, William Rapaport, and felt that because we both make so much of our work freely available online, it could be usefully advertised as a combined resource for philosophy students and others. So with his permission and help, I produced an announcement (http://www.cs.bham.ac.uk/research/projects/cogaff/misc/two-books.html) that mentions our two books and websites, then branches out into separate descriptions:

Rapaport:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/two-books-rap.html

Sloman:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/two-books-slo.html

---

ANNOUNCEMENT

The Association for Philosophy and Computing

Robin Hill
UNIVERSITY OF WYOMING

The Association for Philosophy and Computing (APC) was formed on May 24, 2020. The new association works to provide forums for the critical and creative examination of information, computation, computers, and other computationally enabled technologies (such as robots), and endeavors not only to enrich philosophical research and pedagogy, but to reach beyond philosophy to enrich other discourses, both academic and non-academic. This mission statement reflects the APC’s roots in the long-standing APA Committee on Philosophy and Computers. The Association for Philosophy and Computing will be an affiliate of the American Philosophical Association.

The Association for Philosophy and Computing also publishes related blocks of articles. Together with the APA, it looks after the legacy of the Newsletter on Philosophy and Computers (2001–2020) of the APA Committee on Philosophy and Computers. In addition, the APC is represented in the APA group that awards the Barwise Prize, which is given on a yearly basis for significant and sustained contributions to areas relevant to philosophy and computing.

Initiating members include Peter Boltuc (University of Illinois at Springfield), the long-time editor of the APA Newsletter on Philosophy and Computers, and other community researchers including Jack Copeland (University of Canterbury, NZ), Robin Hill (University of Wyoming), Riccardo Manzotti (IULM University, Milan), Gary Mar (Stony Brook University), Robin Smith (Texas A&M University), Susan G. Sterrett (Wichita State University), Daniel Susser (Penn State University), and Kristen Zbikowski (Hibbing Community College). APC coordinators are Boltuc, Copeland, Sterrett and Susser; APC secretaries are Hill and Sterrett, with subcommittees handling publication, conferences, Barwise Prize advising, and other duties and activities. The group welcomes participation across related disciplines and perspectives. See the website at http://www.philosophyandcomputing.org/, and email the association at philosophyandcomputing@gmail.com.
LETTER FROM THE EDITORS

Tziporah Kasachkoff
THE GRADUATE CENTER, CITY UNIVERSITY OF NEW YORK

Eugene Kelly
NEW YORK INSTITUTE OF TECHNOLOGY

We are happy to welcome readers to the spring 2020 issue of the APA Newsletter on Teaching Philosophy. This issue features one article on teaching, a review essay on C.D.C. Reeve's new translation of Aristotle's *De Anima*, poems by two philosophers, and a list of recently published books that will be of interest to philosophy instructors.

The article on teaching, by Yakir Levin of Ben Gurion University of the Negev (Israel), describes a series of courses that he teaches to advanced undergraduate and graduate students. These courses, which Levin himself describes as “idiosyncratic,” are, as readers will readily see, indeed unusual. Each has as its focus a book specially chosen by Levin for its philosophical interest, both to Levin himself and to his students. The books are read by both Levin and his students as the course progresses so that the philosophical exploration and discussion of the content of the book represents a philosophical journey that both instructor and students take together as, week-to-week, they examine and deliberate about the book’s thesis and its supporting arguments. Levin describes in some detail how he conducts the courses, tells us of some of the books he has used as the foci for such courses, and includes a sample syllabus. He also indicates what he requires of the students in the courses and, importantly, the benefit—for both his students and for himself—of teaching such courses as he describes.

(As Levin indicates in a footnote, he believed that the fact that the courses are so atypical would disqualify an account of them from being suitable material for our publication. One of the editors thought otherwise on the grounds that—withstanding that most of us are required to teach courses that introduce the various fields of philosophy to students—many philosophy instructors do offer advanced elective courses in which students have the opportunity to bring their philosophical perspective to matters presented in books that are not part of the “philosophy canon.” But, as Levin notes in his footnote, the readers should be judge. We would appreciate any feedback readers wish to share with us.)

We are happy to publish a review essay in this issue, written by Rosemary Twomey, of the new translation by C.D.C. Reeve of Aristotle’s *De Anima*. Twomey’s essay comprehensively covers what Reeve’s translation achieves, and indicates how the Reeve translation of *De Anima* compares with other, earlier, editions. Twomey makes some interesting observations about Reeve’s introduction to the volume as well as about the many notes that Reeve appends to the translated text. Twomey also gives readers a good sense of some of the complexities of translating Aristotle, and of how other translators’ different choices might affect the readers’ understanding of Aristotle’s thought. Of especial usefulness are Twomey’s citations of, and references to, these other translations for this allows interested readers to make comparisons between Reeve’s choices and those of others. Twomey also indicates for readers the varying usefulness of the Reeve translation for, on the one hand, Greek scholars and, on the other hand, for initiates.

In this issue, we also offer a number of poems, submitted by (long-term and much appreciated contributor) Felicia Nimue Ackerman of Brown University and also by (first-time contributor) Alexandru Manafu of York University. We welcome the playful insights of both these writers.

As always, we include a list of books that we have received from publishers which might be of interest to our readers. And, again, as always, we encourage our readers to suggest themselves as reviewers of books and other materials that they think may fruitfully be used either in the classroom or in their own preparation. It is especially useful to receive reviews of materials from those philosophy instructors who have used those materials in their own classrooms and so can comment from experience on the merits and/or disadvantages of their use.

SUBMISSION GUIDELINES

When writing for our publication, bear in mind that we are devoted to accounts and discussions of successful pedagogy in the philosophy classroom and not to theoretical discussions of philosophical issues. This should be borne in mind not only by those who write articles but also for those who review material for our publication.

As always, we encourage our readers to suggest themselves as reviewers of books and other materials. It is especially useful to hear from philosophy instructors who have used those materials in their own classrooms and so can comment.
from experience on the merits and/or disadvantages of their use.

We warmly encourage our readers to write for our publication. We welcome papers that respond to, comment on, or take issue with any of the material that appears within our pages.

Guidelines for submitting papers to be considered for publication in the APA Newsletter on Teaching Philosophy

The author’s name, the title of the paper, and full mailing address should appear on a separate sheet of paper or, if the paper is sent to the editors electronically, on a note that will not print out within the text of the paper itself.

Nothing that identifies the author or his or her institution should appear within the body of the paper or within its footnotes/endnotes. The title of the paper should appear on the top of the paper itself.

Please submit the paper in electronic form. If this is not possible, four complete copies of the paper should be sent to one of the co-editors listed below. Authors should adhere to the production guidelines that are available from the APA.

In writing your paper in electronic form, please do not use your word processor’s footnote/endnote function; all notes should be added manually at the end of the paper.

All articles submitted to the newsletter undergo anonymous review by the members of the editorial committee:

Tziporah Kasachkoff, co-editor
Emerita, PhD Program in Philosophy
The Graduate Center, CUNY
(tkasachkoff@yahoo.com)

Eugene Kelly, co-editor
New York Institute of Technology (ekelly@nyit.edu)

Robert Talisse
Vanderbilt University (robert.talisse@vanderbilt.edu)

Andrew Wengraf (andrew.wengraf@gmail.com)

Contributions should be sent (if not electronically) to:

Tziporah Kasachkoff, PhD Program in Philosophy, The Graduate School, The City University of New York, 365 Fifth Avenue, New York, NY 10016, at tkasachkoff@yahoo.com

and/or

Eugene Kelly, New York Institute of Technology, Department of Social Science, Old Westbury, NY 11568, at ekelly@nyit.edu

---

ARTICLE

Devoting a Course to the Exploration of a Book: Journeying Intellectually with the Students

Yakir Levin
BEN GURION UNIVERSITY OF THE NEGEV

For quite some time by now, I have been teaching at the advanced undergraduate/graduate level only, where all the students attending my classes have had a solid background in philosophy, and many of them also in psychology cum cognitive neuroscience. Each of the courses/seminars I have given since my teaching took that route (and to some extent even before that insofar as my advanced courses were concerned) has been based on a book that I think will be of interest to both the students and myself. I usually choose books that deal with central philosophical issues in the areas in which I myself am interested—philosophy of mind, metaphysics, and history of philosophy—or in areas with which I would like to be more acquainted. From time to time I also choose books that have relevance to issues that are on the current public and cultural agenda—e.g., fake-facts; the importance of the value of truth; identity, diversity, pluralism, and authenticity; and liberal democracy. Every now and then I also choose a classical text—e.g., Augustine’s Confessions; and Plato’s Phaedo—into which I myself would like to delve. Apart from the latter texts, the books on which my courses are based are books that I myself haven’t yet read beforehand and are therefore books which I will cover with the students rather than in preparation for the course/seminar that the students take with me. That is, I advance with the reading along with the students rather than read the whole text before they do.

In selecting a book for a course, I usually read a few reviews of the book (if there are any), and then quickly browse through the entire book. I try to choose books that are central, wide-ranging, and as self-contained as possible. During the period in which I have been teaching in this manner, only once a choice that I made turned out to be a failure. The students and I struggled with the book for two or three classes, and then, together deciding that it wasn’t worth the effort, we switched to a different book on a related topic—one which all of us greatly enjoyed reading and learning from.

In a typical course of this type, the students and I read the book to which the course is dedicated cover-to-cover. If the book is comprised of between ten and fourteen chapters, each three-hour class is dedicated to covering one of the book’s chapters. (There is one class held per week and thirteen to fourteen weeks in the semester.) The students and I read one of the book’s chapters during the week before the class that is to be devoted to that chapter, and then we discuss that chapter in class. Sometimes it happens that we do not finish discussing a particular chapter in one class session and in that case—and since I don’t take it as necessary to discuss each chapter in full—we just move on to the next chapter unless we find the topics we haven’t managed to discuss particularly interesting or important.
In the latter case we spend more than one class on a given chapter, and that may come on the expense of later chapters—i.e., sometimes we don’t manage to cover a book in full. The price of this might be that we miss interesting material that is to be found further on in the book. But in the spirit of Wittgenstein’s saying that in philosophy the race goes to the one who can run slowest, we don’t rush anywhere. We go with the discussion wherever it takes us, and that’s a great part of the fun.

While I try to choose books that aren’t too lengthy, it has also happened that I chose fairly lengthy books. It has even happened once that a book that I chose—Robert Pasnau’s Metaphysical Themes 1274–1671—was so lengthy (it is comprised of thirty chapters and is almost eight hundred pages long) that I devoted a year-long double seminar to it. The students in this particular seminar were a very dedicated group who really “got into it” and we all greatly enjoyed the book and learned a lot from it. In books where the chapters are fewer in number than the number of class meetings in the semester, more than one class meeting is devoted to the discussion of each chapter. This is what often happens in the case of books which have fewer but rather lengthy chapters. And if the book is a bit short for a whole semester’s treatment of it, we read parts of another book that deals with related topics. This is what happened in a class that I gave this year: we started by reading, cover-to-cover, A. A. Long’s rather short monograph Greek Models of Mind and Self, a gem of a book that contains five not very long chapters. We then moved on to read quite a few chapters from Stephen Everson’s Psychology, a fine collection of essays by different authors on topics related to Long’s book. But this is relatively rare occurrence as usually the books I choose are sufficiently lengthy to serve for a whole semester’s study.

It has also happened twice or thrice that I have dedicated a seminar’s weekly four-hour-class to a book. One such seminar was devoted to Tyler Burge’s Origins of Objectivity, which although comprising only eleven chapters, is almost six hundred pages long. Another such seminar was devoted to Evan Thompson’s Waking, Dreaming, Being, which, again, although comprising only ten chapters, is almost four hundred fifty pages long. Both these seminars were at the graduate level.

Below you will find a brief description of courses/seminars that I gave in the last academic year, as well as one that I gave this year, each conducted in the manner described above. Due to the form of these courses (described above), the syllabus for each is not very detailed. I have therefore appended a syllabus of only one of these courses, as that syllabus is pretty much exemplary of the others.

As can be seen from the examples given below, the topics of my courses/seminars are rather varied. Each of these courses forms a sort of journey with the students, one in which neither I nor the students know in advance how it will develop, or where it will lead—a true intellectual adventure. I myself usually learn a lot from these classes—indeed, often enough, points raised and discussed in them find their way into my own work—and when they go well, I enjoy them tremendously. Judging from the student papers at the end of the course, it appears that the students get a lot from them as well. It is also the case that, since I read the books along with the students, and so do not have to prepare much in advance apart from doing the week-to-week reading of the chapters to be discussed in that week’s class, I am able to offer completely new courses/seminars each and every year. Indeed, each year I now offer four such new courses—two per semester. Having been teaching for quite a long time, I have reached a point where I find it difficult to repeat a course while still maintaining the joy of teaching the material of that course. I need fresh intellectual experiences and excitement both to do the teaching properly and also to enjoy the teaching. It is also the case that, not only do my many years of teaching motivate my manner of teaching, but together with the skills and background knowledge that I have acquired along the way, they enable me to teach in that manner.

Besides the personal motivation behind my manner of teaching, there is, of course, also a pedagogical motivation. As Kant is reported to have said, his intentions in teaching weren’t to teach philosophy but rather to teach philosophizing “not thoughts merely for repetition but thinking.” This is, in a sense, what any philosophy course worthy of the name is supposed to do (or so it seems to me). I think that collectively thinking with students through a major philosophical work—indeed, doing philosophy together with them—is a particularly good way of achieving such a goal, i.e., of improving students’ thinking-cum-analytical-cum-philosophical skills. My teaching manner, then, is a way of helping students achieve this goal. But, of course, it requires some intellectual maturity on the part of the students, and some prior—indeed, considerable—initiation into the intricate practice of philosophical thinking. That’s why my particular manner of teaching suits upper-level students and not beginners. My manner of teaching also does not suit students who would like to be taught in a very orderly and systematic way. In my experience, however, advanced students who are willing to participate in an intellectual adventure—though at times haphazard and often leading to unexpected destinations—are students who greatly benefit from the manner of teaching described here. But not only do they profit from this manner of teaching, they delight in it as well.

A few technical points about my courses:

1. Attendance in the classes comprising a course is compulsory—i.e., students must attend at least ten classes out of thirteen in the semester.

2. Apart from the required weekly reading, students are required to write one relatively short mid-term paper (usually on one of the chapters in the book that we are reading). I do not grade these papers, but I do sit and discuss each student’s paper with him or her tête-à-tête.

3. I assign a second, longer, paper whose topic and relevant bibliography I announce to the students—i.e., all students in the course write on the same topic. I do grade this longer, final paper and this grade becomes the student’s final grade in the course.
Syllabus for the Seminar:

Ben-Gurion University of the Negev
Department of Philosophy

Madness and Modernism
1st Semester, 2019

Lecturer: Professor Yakir Levin, yakirl@bgu.ac.il.

Office Hours: Tuesdays, 14:00-15:00, or by appointment, Diller Building (74), room 336.

Time and Location: Wednesdays, 17:00-20:00, Diller Building (74), room 343.

Seminar Description:
The similarities between madness and modernism in both literature and the visual arts are striking: defiance of convention, nihilism, extreme relativism, distortions of time, strange transformations of self, and much more. In this seminar we shall examine a severe psychopathology such as schizophrenia in light of these similarities. On the psychopathological side we shall rely on clinical work and reports by both patients and therapists. On the modernist side we shall refer to the works of such artists and writers as Kafka, Beckett, and Duchamp, and to the ideas of philosophers such as Nietzsche, Heidegger, Foucault, and Derrida.

Assignments and Grading Policy:

• Reading from week to week of the reading assignments from the book whose details are given below.

• Each week you will receive the reading assignment for the following week.

• All reading assignments will be put in advance on the Moodle site of the seminar.2

• In the middle of the semester there will be a short mid-term paper assignment. This paper will not be graded. But I will meet tête-à-tête with each and every one of you in order to discuss it. Submission date for this paper will be announced when you get this assignment.

• At the end of the seminar you will receive a topic, bibliography, and instructions for a longer seminar paper. Submission date for this paper is according to the Faculty regulations. This paper will be graded, and the grade you will get on it will be your final grade in the seminar.

• Attendance Policy:
You are required to attend at least ten out of the thirteen class sessions of the seminar.

SAMPLE COURSES/SEMINARS

A Seminar on Louis Sass’s Madness and Modernism: Insanity in the Light of Modern Art, Literature, and Thought (revised version)

In this book, Sass (a clinical psychologist) outlines his phenomenalistic approach to schizophrenia, and suggests that we might best understand this mental disorder by means of suggested parallels with modernism in both literature and the visual arts. Some such parallels are, for example, (a) defiance of convention, (b) distortions of time, and (c) strange transformations of the self. In the course of drawing these parallels between schizophrenia and major modernist literary and artistic trends, Sass makes reference to the works of such artists and writers as Kafka, Beckett, and Duchamp, and to the ideas of philosophers such as Nietzsche, Heidegger, Foucault, and Derrida.

Sass’s book (in its revised version) has eleven chapters. We read the book cover-to-cover, devoting one class session to each chapter. Some of the discussions of some of the chapters spilled over to further class sessions (that were supposed to be devoted to subsequent chapters), which is why it took us thirteen weeks to read the entire book.

I should mention that among the fourteen students who took this seminar—many of whom were majoring in both philosophy and psychology—five students were actually working with schizophrenic patients, and one was a PhD psychology student whose dissertation was on schizophrenia. Not surprisingly, the personal experience and background knowledge of these students greatly enriched the discussion in class.
• Bibliography:

A Course on Quassim Cassam’s *Self-Knowledge for Humans*

Cassam’s book develops an account of self-knowledge that tries to do justice to the respects in which humans are not model epistemic subjects. The book rejects rationalist and other mainstream philosophical accounts of self-knowledge on the grounds that these accounts do not accurately account for how humans actually come to have knowledge about themselves. Instead, Cassam defends the view that inferences that we draw from both behavioral and psychological evidence constitute the basic source of our knowledge about ourselves. In addition, Cassam provides an account of how we come to have (what he calls) “substantial” self-knowledge, including the knowledge we have of the values we hold and of our character.

Cassam’s book is divided into fifteen chapters, of which each of the first twelve were covered/discussed in class. The discussion of the first twelve chapters was sufficient to prepare students to read the last three chapters of the book on their own, and then to write and submit a final paper on some of the topics (such as self-ignorance and the value of self-knowledge) that are covered in these chapters.

A Course on Bernard Williams’s *Truth and Truthfulness: An Essay in Genealogy*.

In this book Williams explores the values of truth and truthfulness focusing on questions such as: What does it mean to be truthful? What role does truth play in our lives? Williams proceeds by deploying a Nietzschean genealogical method (appealing to both fictional and real genealogies) and in doing so he explores related values like sincerity and authenticity.

We read the book throughout the semester, devoting one class to discussion of each of its ten chapters. As in the seminar on Sass’s book, and for similar reasons, it took us thirteen weeks to read the entire book.

A Course on Alvin I. Goldman’s *Simulating Mind, The Philosophy, Psychology, and Neuroscience of Mindreading*.

In this book Goldman seeks to make a case for the importance of simulation in both low-level and high-level tasks of third-person mind reading. Interdisciplinary to the core, the book bridges philosophy of mind, psychology, and neuroscience, thereby providing perhaps the definitive account of simulation theory. We read the book throughout the semester, devoting one class to discussion of each of its eleven chapters. As in the case of the former two courses, and for similar reasons, it took us thirteen weeks to read the whole book.

A Seminar on Anthony Appiah’s *The Ethics of Identity*.

In this book Appiah seeks to develop a Millean (i.e., liberal) account of identity that takes seriously both the claims of individuality—the task of self-creation or of making one’s own life—and the larger claims of social categories such as gender, nationality, race, and religion through which we come to define ourselves as the persons we are.

Although 272 pages long, Appiah’s book is divided into only six chapters, each of which is rather long, very rich in content, and raises many questions. Because of this, we read the text in fairly small portions, devoting, for example, five classes to discussion of the first two chapters of the book.

Before a semester starts, my university’s computing department opens a Moodle website for each of my courses for that semester. It’s a superb means for communicating with the students, e.g., for sending them collective emails (which I do at least once a week); for listing reading materials for the course; for giving instructions for the midterm and final papers; for informing students of the details of occasional lectures and workshops that are relevant to the course and to which I recommend attendance.

Among the reading materials I put on the Moodle website of my courses, I usually include additional materials relevant to the book we are reading in those courses. For example, for the seminar on Appiah’s book, *The Ethics of Identity*, I put on the site a few relevant articles from the press about voting patterns in Israel amongst different identity groups. Articles which had been part of an ongoing heated public debate about these issues. In addition, I put a reference to the website of a thought-provoking and very touching documentary which was shown on the Israeli public TV channel about the Queer community in Israel. (Needless to say, the students in class were much more familiar with this identity group than I, and I learned a lot from them during our discussion of this particular example in the context of our examination of Appiah’s account of identity.)

Another example: for the course on Williams’s *Truth and Truthfulness*, among other things that I placed on the Moodle site was an article about Bernard Williams and his intellectual legacy that appeared in the *Times Literary Supplement* as part of their marvelous series “Footnotes to Plato.”

**NOTES**

1. The co-editor of this newsletter, Tziporah Kasachkoff, asked me to share with readers of this publication my rather idiosyncratic manner of teaching. Though it occurred to me that my method of teaching might be too idiosyncratic to be useful to others, Professor Kasachkoff, a good friend of mine to whom I cannot really say no, thought otherwise. So I agreed to her request. You, the reader, will be the judge.

   Many thanks to Tziporah for her wise advice, most helpful comments, and excellent suggestions.

2. Moodle is an online learning management system which enables one to create a private website for one’s courses. (All universities in Israel—as well as, I think, many universities elsewhere in the world—use it.)
REVIEW ESSAY

Aristotle, De Anima


Reviewed by Rosemary Twomey
QUEENS COLLEGE, CITY UNIVERSITY OF NEW YORK

Over the past several years, Hackett has published a host of C.D.C Reeve translations of Aristotle under the series title “The New Hackett Aristotle.” De Anima [DA] joins Metaphysics, Nicomachean Ethics, Politics, Physics, Rhetoric—and Topics and Sophistical Refutations, which is forthcoming. While the present edition is not explicitly oriented towards teaching and speaks only generically of readers’ interests, the Introduction starts with basic biographical detail of Aristotle’s life, suggesting students are among the intended audience. This is an ambitious project, and one for which there is some need: since W.D. Ross in the early twentieth century there has not been a systematic effort to render all or most of Aristotle’s surviving works in one translator’s voice. Still, Aristotle’s language is fairly technical, so translations do not show their age as quickly as they might otherwise. (Among decades-old editions of DA, in addition to Ross’s later [1961] edition, Hicks’s excellent 1907 translation is still a viable option for classroom use.) Nonetheless, those who teach Aristotle in the twenty-first century often use more modern translations such as that of Hamlyn [1968, repr. 2002] or Shields [2016], or anthologies such as the excellent one of Irwin and Fine.1 Reeve’s translation is a good addition to this list, especially if one is also teaching another text translated by Reeve so that there is consistency in terminology. It is important to keep in mind that translators render critical terms differently: nous can be rendered as “understanding,” “mind,” “intellect,” “thought,” or even “reason”; aisthêsis as (either) “perception” or “sensation”; and so on with important terms including phronêsis, kath’ hauto, kata sumbebêkos, and epístêmê. Because of this variation, reading multiple texts with different translators may have the effect of obscuring clear connections. Of course, Aristotle is not himself always consistent at keeping his technical terminology uniform, a fact of which Reeve is mindful. For example, in DA I.2 Reeve translates phronêsis as “wisdom” rather than as “practical wisdom” (as he does in his edition of NE) because Reeve (rightly) thinks that Aristotle does not have his doctrine of practical reasoning in mind here (84n38). Still, Reeve rarely varies his terminology, which allows readers who do not know Greek to draw parallels among Aristotle’s works more easily.

The Reeve translation is readable and clear. It is worth noting that Reeve’s commentary takes up twice as much space as the text itself. This is to be expected as Aristotle’s writing is frequently elliptical and assumes knowledge of technical terms that Aristotle defines elsewhere. But whereas other popular editions divide the text into digestible chunks with long expository commentaries, Reeve chooses to use frequent endnotes—448 in all and often one per sentence. Indeed, if one attends to all of the notes, reading the text may prove to be a disorienting experience. Furthermore, not all of the endnotes will be of relevance to all students as many of the endnotes, especially in the later parts of Book I, provide reference information. Reeve includes a note every time Aristotle refers back to something he has previously said in another place. While the notes can sometimes be helpful in pointing us to other texts, oftentimes the note refers us back to an obvious place just a few lines earlier (e.g., notes 245 & 280). Other sorts of notes that Reeve provides will be more helpful for students: whenever technical terms are introduced Reeve takes care to provide in his endnotes some context for those terms—though sometimes the supplied context comes in the form of an unelaborated-upon quotation from another work. (For example, in II.4 Aristotle says that the soul is the cause of the body “in three of the ways distinguished.” Aristotle is here referring to his four causes, most fully explicated in Physics. In his note, Reeve quotes without comment a summary of the four causes from Metaphysics. Students unfamiliar with the general doctrine of the four causes will not get much elucidation from this rather dense quotation.) In general, I had the impression that the notes would be more helpful to Aristotle scholars than to beginning or intermediate students.

Reeve has many interesting interpretations, especially in the later parts of Book III. But because he does not survey the literature—an editorial decision he mentions in the preface (ix)—students will not get a sense of interpretive controversies by attending to the notes.

Reeve provides a wide-ranging and thought-provoking introduction, but, like the notes, it will be of variable interest to students. The discussion of Aristotelian science and methodology will help students understand what Aristotle is trying to accomplish in DA. I agree with Reeve that DA should be understood as an attempt to provide the indemonstrable first principles upon which the rest of psychology would (must) rely. In the introduction and notes, Reeve does an excellent job of presenting Aristotle’s view of the structure of science and the role both of first principles and of the demonstrations that rely on them (though he could also have appealed to some evidence from within the text itself rather than referring exclusively to other Aristotelian treatises (e.g., II.1 402a7-10 & II.2 413a11-16)). But the introduction misses chances to outline basic concepts that students will need to have while making their way through the text. For example, Reeve does not explain the range of uses of the term psuchê (soul), and his commitment throughout the introduction to the separability of intellect from body (a controversial interpretive claim) could, understandably, lead students to interpret Aristotle’s interest in soul as part and parcel of a more general dualism. It would have been helpful had Reeve indicated to readers that while there is no perfect English translation of the term, psuchê refers to whatever it is that distinguishes living things from non-living ones. Reference to “souls” is compatible with a wide range of views about their underlying nature and the connection that obtains between them and corporeal bodies. There is also no explanation of basic Aristotelian concepts such as form and matter and potentiality and actuality. While these
concepts are touched upon in the notes when they occur in the text, it would have been helpful if Reeve had explained these concepts before the student encounters them in the text itself.

Other parts of the Introduction bear a less obvious relation to the text. For instance, Reeve discusses pneuma at some length even though (arguably) it does not feature in DA. Though elsewhere in his oeuvre Aristotle employs a technical notion of pneuma (especially in his biology and his account of the living body), in DA the word and its cognates only occur three times (420b20 & 24; 421b15), and each time it refers to breath. Though Reeve disagrees with Abraham Bos’s specific claim that pneuma (rather than the visible human body) is, according to Aristotle, the body to the soul’s form (xxiv), he apparently embraces Bos’s general position on the importance of pneuma in Aristotle’s psychology: Bos’s The Soul and Its Instrumental Body is among the few secondary sources mentioned in the “Further Reading” section at the end of the book. But Reeve doesn’t do enough to defend his very controversial reading of pneuma in his Introduction, nor does he even indicate that it is controversial. Reeve does claim that pneuma is implicitly appealed to in explaining transparency in media such as air and water as well as in the transmission of intellect to the fetus, but Reeve’s argument here is hasty and has little textual support, and in any event it will not be clear to the novice that this reading is idiosyncratic. But more fundamentally, while pneuma as a general concept will be foreign to many students, Reeve assumes students’ familiarity with its meaning. Students would have benefited had Reeve proceeded more slowly and included some of the history of the term and its connections with the concept of breath.

Like his other translations, Reeve’s rendering of De Anima is accurate and (for Aristotle) quite readable. Students should find the wording familiar. For the most part his translation does not make choices that render his favored interpretations all but inevitable. This is as it should be since a good teacher should be able to use the text as written in outlining controversies and presenting alternative interpretations of that text. One exception to this salutary practice occurs in Reeve’s discussion of Aristotle’s attitude towards the idea of the “parts of soul.” Aristotle is reluctant to acknowledge “parts of the soul” as having the same sense that Plato gives it in the Republic—that is, as distinct, separable, agents within the soul. Aristotle expresses this reluctance by twice explicitly doubting whether it is apropos to use the word for “part” (morion) in discussing the capacities of soul (413b13-32 & 432a22-b8). However, when trying to refer neutrally to that which performs the soul’s various functions, Aristotle does freely use a construction combining the activity in question with the neuter article and the ending -ikon. Translators have traditionally rendered these as references to faculties: to aisthêtikon—usually translated as “the perceptual faculty.” Such a reading is in some ways quite loaded since it suggests a faculty psychology (a very plausible suggestion, but a charged one nonetheless). David Hamlyn instead translates aisthêtikon neutrally but awkwardly as “that which can perceive.” Reeve, on the other hand, translates the word as “perceptual part,” which makes Aristotle’s reluctance to use the word “part” elsewhere look rather puzzling, most strikingly in in II.2, where Aristotle asks whether we should think of the primary functions of the soul as being “parts” of souls (413b13-15). He does not appear to settle this issue, yet just a few lines later, Reeve has him speaking freely of the perceptual “part” and the believing “part” (413b29-32). Aristotle returns to the issue in III.9, where he again struggles to decide whether or not to refer to “parts” of soul (432a22-23) even though, in Reeve’s translation, Aristotle appears to have been referring to “parts” of the soul throughout the text.

The book includes a brief collection of Further Readings consisting of: a few biographies of Aristotle, the ancient commentaries which have been translated into English, and a highly selective list of recent sources.

The book has a very thorough index covering not just the text but also the notes, and this will prove useful to students and scholars alike.

In my judgment Reeve’s translation is a good and a well-organized resource, though I prefer the notes in other recent editions—especially Shields’—as they cover more controversy and so give students a better sense of the diversity of positions. The Reeve-translated text itself is to be recommended, but students who want exposure to the range of interpretations will need to supplement Reeve’s edition with one or more of the others.

ACKNOWLEDGMENTS

Thanks to Iakovos Vasiliou for comments on an earlier draft.

NOTES

1. Irwin and Fine, Aristotle: Selections. For other recent editions, see Miller, Aristotle: On the Soul and Other Psychological Works; and Polansky, Aristotle: De Anima.
3. Such as Hamlyn, Aristotle: De Anima; Hicks, Aristotle, De Anima; Polansky, Aristotle: De Anima; Ross, Aristotle: De Anima; and Shields, Aristotle: De Anima.

BIBLIOGRAPHY

Hicks, R. D. Aristotle: De Anima. 1907.
POEMS

To Teachers Who Hope to Inspire Their Students
Felicia Nimue Ackerman
BROWN UNIVERSITY
This poem first appeared in Daily Nous.

I never had a teacher more inspiring than Ms. Burr.
She led me to resolve that I would never be like her.

To Those Who Think the Unexamined Life Is Not Worth Living
Felicia Nimue Ackerman
BROWN UNIVERSITY
This poem first appeared in Daily Nous.

Lloyd always acts without thinking.
Reflection is hardly for him.
Lillian’s mind has been shrinking.
Dementia is making her dim.

Both find enjoyment in living.
So don’t be so ready to scoff.
Why are you so unforgiving?
How harsh to be writing them off.

To Cynthia Ozick*
Felicia Nimue Ackerman
BROWN UNIVERSITY
This poem first appeared in Daily Nous.

Aesthetics and logic,
Injustice and war:
Philosophers ponder
These topics and more.

We needn’t relinquish
This varying focus.
Our field would be meager
With only one locus.

*The novelist and essayist Cynthia Ozick says, “Novelists, poets, philosophers and theologians agree: Mortality, that relentless law of universal carnage, is the sole worthy human preoccupation.”

In Praise of Campus Culture Wars
Felicia Nimue Ackerman
BROWN UNIVERSITY

A campus that is truly free
Has denizens who disagree.
There isn’t any culture war
In Orwell’s Nineteen Eighty-Four.
Philosophy Rap

Alexandru Manafu
YORK UNIVERSITY

This poem was read on the first day of an Introduction to Philosophy class. For most students in attendance this was their first day of university.

Great minds like Stephen Hawking said:
"Philosophy is dead!"
But I’m here to prove him wrong
Philosophy’s still strong!

For as long as there will be people
There will be big questions too
Like: Who am I?, What’s right or wrong?
What do I have to do?

Is there a God? What’s truth?
What can I truly know?
What’s love and what is beauty?
What’s the transcendent, yo?

But this course is no joke
Philosophy is hard
You’ll have to read some Plato,
And Nietzsche, and Descartes

You’ll have to write some essays
And argue, go to bed at 3
Philosophy’s not easy
And I’m philosophy, you see?

But stick around, because
In this course you will find
Philosophy is good for you
It feeds your soul, your mind

And with these words I close my sketch
And throw this in the bin
This poem is now ending
Let our first lecture begin!

BOOKS RECEIVED

*Books marked with an asterisk will be reviewed in one of
our forthcoming issues.

(Readers are welcome to suggest themselves as reviewers
of one of the books listed.)

BROADVIEW PRESS
Cahn, Steven M., Philosophical Adventures*

HACKETT PUBLISHING COMPANY
Aristotle. Rhetoric, Translated with Introduction and Notes,
by C. D. C. Reeve.

Forcehimes, Andrew T. and Semrau, Luke. Thinking Through

Perry, John. Dialogue on Consciousness: Minds, Brains, and
Zombies.

Plato. Charmides, Translated with Introduction, Notes and
Analysis by Christopher Moore and Christopher Raymond.

Ariew, Roger & Watkins, Eric., eds., Modern Philosophy: An
Anthology of Primary sources, 3rd edition.

OXFORD UNIVERSITY PRESS
Cahn, Steven M., Exploring Ethics: An Introductory
Anthology, 5th edition.

PRINCETON UNIVERSITY PRESS
Singer, Peter. 82 Brief Essays on Things that Matter.* (with
new afterward by the author)

ADDRESSES OF CONTRIBUTORS

Dr. Felicia Nimue Ackerman
Brown University
Department of Philosophy
Box 1918
Providence, RI 02912
Email address: felicia_ackerman@brown.edu

Dr. Yakir Levin
Ben Gurion University of the Negev
Department of Philosophy
P.O. Box 653
Beer Sheva 84105
ISRAEL
Email address: yakirl@bgu.ac.il

Dr. Alexandru Manafu
York University
Department of Philosophy
S414A Ross Building
4700 Keele Street
Toronto, Ontario
Canada M3J 1P3
Email address: alexman@yorku.ca

Dr. Rosemary Twomey
Queens College, City University of New York
Department of Philosophy
65-30 Kissena Blvd
Flushing, New York 11367
Email address: rosemary.twomey@gmail.com