FROM THE EDITOR

Theodore Benditt, Editor
University of Alabama at Birmingham

The topic for the current issue of the Newsletter is “Game Theory and Law;” many thanks to Gerald Gaus (Philosophy, Tulane University) for taking on the role of guest editor for this issue. As Gaus points out in his introduction, game theory has, in recent years, appealed to a number of legal theorists as a model that can help explain law; it has been applied to the understanding of specific problems and also to larger questions of the nature of law. Some of the contributions to this issue of the Newsletter continue a debate about the use of game theory and economic reasoning in one particular area of law, whereas other contributions raise questions about the appropriateness of at least certain ways of appealing to game theory in our attempt to explain and understand law.

As always, the Newsletter also contains a number of abstracts of recent articles from law reviews and other journals of interest to philosophers. I want to thank members of the APA Committee on Philosophy and Law who have contributed abstracts for this issue: Julie van Camp (Philosophy, California State-Long Beach), and Brian Bix (Law, University of Minnesota).

New Editor Sought
With the term of the current editor drawing to a close in June 2005, the Committee on Law and Philosophy is now searching for a new editor for the Newsletter on Philosophy and Law. The editor has the responsibility to solicit contributing articles, reviews, and guest editors for each issue of the Newsletter. The term is for five years. The editor is an ex officio member of the Committee on Law and Philosophy. Anyone interested should send a CV and letter of application to Patricia Smith, at patrigsmith@att.net. Application by email is preferred.

Future Issues of the Newsletter
Topics and editors for the next three issues of the Newsletter are:

Fall 2004 – CONSTITUTIONAL INTERPRETATION
Submission Deadline: June 15, 2004
Editor: Theodore M. Benditt
Department of Philosophy
University of Alabama at Birmingham
Birmingham, AL 35294-1260
(205) 934-4083
tbenditt@uab.edu

This issue will be devoted to a variety of issues and topics related to constitutional interpretation.

Spring 2005 – INTELLECTUAL PROPERTY RIGHTS IN CYBERSPACE
Submission Deadline: January 15, 2005
Editor: Richard A. Spinello
Carroll School of Management
Department of Operations and Strategic Management
Boston College
Chestnut Hill, MA 02467 USA
(617)-552-1189
richard.spinello@bc.edu

This issue will consider the scope of intellectual property rights in cyberspace from a moral perspective. Although much has been written about this topic by legal scholars, it needs to be more cogently addressed by philosophers who specialize in ethics or political philosophy. Articles should seek to illuminate various dimensions of the intellectual property debate, which ranges from the theoretical (For example, is Locke’s labor desert theory viable when applied to intellectual property?) to the more practical (Is it morally acceptable to copy software and other digital works?). Do users have an unqualified right to engage in hyperlinking, even if they link to infringing material? Do domain names deserve strong trademark protection? Proposals are welcome.

Fall 2005 – LIFE AND THE LAW
Submission Deadline: June 15, 2005
Editor: Samuel Gorovitz
Hall of Languages 541, Syracuse University
Syracuse, NY 13244 USA
315-443-9331; fax: 315-443-5675
Dearing-Daly Professor of Bioethics and Humanities
Upstate Medical University, Syracuse, NY 13210

The Fall 2005 Issue of the Newsletter will be on “Life and the Law”—more specifically, on issues at the intersection of law, ethics, and health policy. Guest editor Samuel Gorovitz encourages prospective authors to send a brief initial inquiry as soon as possible describing a proposed topic within a few sentences. (Send to vitz1@aol.com ) Selected authors will then be invited to submit preliminary drafts, as work in progress, by September 15, 2004. Those drafts will be rigorously reviewed in a seminar at Yale during the fall 2004 term and will be returned without prejudice to their authors, along with critical commentary, in ample time for revisions to be incorporated prior to the June 15, 2005 deadline for formal submission.
**Game Theory and Law: Introduction**

Gerald F. Gaus  
*Tulane University*

On almost any account, a basic function of law is to facilitate cooperation among strangers, and to provide incentives for people not to cheat on beneficial, cooperative arrangements. Because game theory provides sophisticated formal models of both cooperative behavior (as in coordination games) and cheating behavior (the infamous prisoner’s dilemma), legal theorists have been drawn to game theoretic ideas to explain and/or justify the law. It is certainly an appealing project. If, given some standard assumptions of rational action, we can show that law is a solution to coordination problems—we need to coordinate on some common way of doing things, such as driving all on the right or all on the left—it looks as if law is a device that all rational agents would hit upon to advance their interests. Or, supposing a more Hobbesian view, if individuals are regularly confronted by prisoner’s dilemmas, in which they are rationally compelled to defect even given the knowledge that they will be worse off when all do so, the law offers a way out, punishing free riders so that cooperation can arise.

The first two contributions to this issue evaluate this project: To what extent does game theoretic analysis allow us to understand law? Leslie Green—whose paper on “Law, Coordination, and the Common Good” is a classic—on this issue—returns to the question whether law can be understood in terms of coordination games. As Green points out, on H.L.A. Hart’s theory of law, ultimate legal rules—rules of recognition—are in the end customary: it is the consensus of the actors in the legal system that the Constitution governs law-making that constitutes the rule that it is the source of American law. So in this way law derives from a common practice, thus suggesting the coordination analysis: we coordinate on common rules about legal authority. In his paper, however, Green argues that, despite the obvious attractions, the coordination game analysis of law fails—though, he adds, just why it fails is enlightening. Green rightly points out that we cannot infer the existence of a coordination problem from the observation that we have a shared practice or rule: the advocate of the coordination view must show an independently existing coordination problem that the practice of law solves. Green is especially critical of the failure of coordination game analyses of law to explain the way in which legal norms are authoritative insofar as they provide what Joseph Raz calls “exclusionary reasons”—reasons not to consider some reasons in one’s deliberations. Coordination analyses focus on how the balance of one’s reasons would lead one to follow a rule, and thus fail to explain duty-imposing rules. And it is this duty-imposing character of law that allows it to solve our coordination problems.

Bruce Chapman’s essay also questions whether game theory is the proper way to model legal interaction. In his insightful paper, Chapman distinguishes the conception of rational interaction informing game theory, and contrasts it to rational interaction in legal settings. As Chapman recognizes, game theory is all about the ways in which rational actors interact, with this interaction being understood as requiring that you, if you are to maximize your payoffs, must take into account what the other will do (knowing that the other will be taking account of what you will do, and that what you will do anticipates what he will so, etc.). While in one sense this is a very complex structure of interaction based on rationality, involving layers of mutual anticipation, Chapman argues that it is not really rational interaction at all: the parties’ reasoning does not really adjust to each other. They do not reason together to solve their problems. Chapman argues that embedded in law is a much more public and truly interactive conception of shared reasonableness: law seeks to identify a common conceptual space, not simply a coordination of private aims. Moreover, the reliance of game theory on an overly private (though of course interactive) conception of reason leads it to underestimate the possibilities for reasonable agents to arrive at common, beneficial, outcomes.

The last two essays of this issue comprise a debate about the application of game theoretic and economic reasoning to a specific area of the law: the theory of punishment. Michael Ridge offers a critique of Michael Davis’s well-known auction theory of just punishment. Davis, says Ridge, sees crime as a sort of unfair advantage of free-riding (so we are concerned now with multi-person prisoner’s dilemmas rather than coordination games). Davis envisages an auction in which a limited number of licenses to commit crimes are up for bidding; the price that the licenses would fetch gives one an indicator of how much the unfair advantage is worth, and so a ranking of how serious is the violation. This, in turn, allows us to generate a ranking of severity of crimes that, if punishments are to be proportional to crimes, must correlate with a ranking of severity of punishments. Ridge objects that this cannot be correct, since taking unfair advantage—free-riding—is only one source of moral wrongness. Consequently, a scheme of punishment that perfectly correlated with degree of unfair advantage would fail to make punishments proportional to the total wrongness of the crime. Ridge, however, presses his critique even further: even when the only wrongness involved in a crime is unfair advantage, he insists that the auction theory still fails to generate a ranking of punishments that meets the condition of proportionality. Even in cases where only free-riding is involved, Ridge argues, the wrongness of the free-riding is not simply a function of the unjust gains to the criminal, but the value to others of the public good that the crime has endangered.

In the last essay of this issue, Michael Davis responds to Ridge’s criticisms. Davis objects to Ridge’s method of describing hypothetical cases of wrongdoing and drawing on our “intuitions” about what is appropriate. Davis closely examines the idea of an “intuition.” We cannot have genuine intuitions about cases of which we have no experience, and so many so-called intuitions about cases outside our experience—which he maintains are basic to Ridge’s objections—should not have weight in our deliberations about justice. Moreover, Davis argues, even if people have genuine intuitions about a case, they are often not shared with others, or are not really about the issue at hand. Overall, then, Davis offers a sustained criticism of many appeals to intuitions to invent counterexamples to theories of justice. Davis also insists on the importance of the distinction between moralism and legalism: the latter, which Davis espouses, sees punishment as a purely legal concept, and thus does not seek to track overall moral desert. Ridge, Davis argues, is a moralist; only one who already embraces moralism will be moved by his core examples. In any event, Davis argues, Ridge sometimes misapplies the auction model because the examples are under described.

These sketches, of course, do not do justice to the complexities of these four papers. I hope it is clear, though, that all papers engage the debate about the application of
decision-theoretic rationality and game theory to our understanding of law and justice, a matter of fundamental importance to contemporary legal philosophy.

Endnote

**Articles**

*Strategy and Ultimate Legal Rules*

Leslie Green*

York University, Toronto

Custom requires that one tip for good service. That is no mere habit: it is rule, used as a standard of guidance and appraisal by diners and servers alike. And it is a fairly important rule: refusing to tip is not only rude, it is wrong, for it exhibits a lack of consideration for those who serve and it free-rides on those whose gratuities sustain the service economy in its present form. Yet there is no legal obligation to tip. There is, on the other hand, a legal obligation to pay the bill, even if the food is mediocre and the service poor. Why then is the tipping rule a mere custom but the paying rule is law?

One might say that the paying rule is law because it will ultimately be enforced by people like judges, bailiffs, and the police, whereas the tipping rule will not. That points us in the right direction, but it is inexact. What is relevant is not whether enforcement is likely, but whether it is authorized: some laws are rarely enforced; some predictable deployments of force are unlawful. Validity, not predictability, is the specific mode of existence of a law. So what authorizes enforcement of the paying rule? The law of contract does—a complex amalgam of things law? They are law because they have been made in ways and by people authorized to make law. The enactments, for example, have been created by the exercise of powers granted by other enactments, including a Constitution that gives a legislature the power to regulate business transactions. But why is that Constitution law? Because it was created and amended in a way authorized by an earlier constitution. Yes; but why ...? The lawyer’s old question—*Quo warranto?*—may be answered repeatedly, but not indefinitely. At some point the validity of law rests on something not itself legally authorized.

The most convincing account of these foundations is due to H.L.A. Hart. In *The Concept of Law*, he argued that the ultimate criteria of legal validity are constituted by the customary practices of officials, insofar as their practices amount to a social rule. Of special importance among these practices is a *rule of recognition*, which specifies what features a rule must have before it is to be used in the ways that legal rules are.¹ The rule of recognition is no mere juristic hypothesis as Hans Kelsen thought the “basic norm” to be; its existence is a matter of empirical fact. But neither is it a brute fact in the way John Austin conceived of the habit of obedience to superiors; the rule of recognition is a social rule and has the normative force that such rules have. It is because our rule of recognition certifies certain activities of legislatures, courts and others as sources of law, while denying that status to general social customs, that the payment rule is legally binding and the tipping rule is not. It is worth noticing, though, that the non-binding status of the tipping rule is not, as one might think, a function of its customary nature. It is a function of the fact that this particular custom is not, in our legal system, one that officials are bound to recognize and apply in accordance with their customs. At bottom, law itself is nothing more than a social construction of social customs.

What then is a customary rule? Hart proposed a test. He said that a group of people, G, treat a regularity of behavior, R, as a customary rule when: R is generally practiced in G, deviations from R are widely criticized by members of G, this criticism is widely regarded as legitimate, and some members of G use R to guide and assess behavior—they “accept” R from what Hart called the “internal point of view.”² This test will ensure that R is not a mere coincidental or habitual regularity; the uses to which it is put mark it as lying within the realm of the normative. Unfortunately, the test will not also guarantee that we have found an instance of rule following. People regularly carry umbrellas in the rain; they advise and commend the carrying of umbrellas; they criticize children whose pride allows them to get drenched. But we have no *rule* that umbrellas are to be carried in the rain. There is only a general reason to keep dry, and a set of common normative responses to the modalities and value of doing so. Crucially, we do not cite the practice of umbrella-carrying itself as a reason for carrying them. In contrast, the rule that duly enacted statutes are legally binding is something that courts do treat as a reason for applying them. A customary rule exists only when it can be cited as a reason, when the fact of common practice is a non-redundant part of a practical argument in favor of acting in the way the rule prescribes.

Hart later came to acknowledge that his general theory of rules was inadequate; but he wanted nonetheless to insist on one point:

> [T]he rule of recognition is treated in my book as resting on a conventional form of judicial consensus. That it does so rest seems quite clear at least in English and American law for surely an English judge’s reason for treating Parliament’s legislation (or an American judge’s reason for treating the Constitution) as a source of law having supremacy over other sources includes the fact that his judicial colleagues concur in this as their predecessors have done.³

Subordinate legal rules—rules which have been legally authorized—may exist because they are certified by the system’s criteria of validity; but ultimate legal rules must rest on common practice. So the reason for treating the Constitution as law includes the fact that there is a common judicial practice to this effect. Practice may not be the whole story: “includes,” we may assume, signals a necessary condition only. But the fact is that if American judges (and legislators) generally stopped deferring to the Constitution, an individual judge would no longer have a legal duty to do so. At the ultimate level, common practice may even displace formalities. It is sometimes doubted whether the Fourteenth Amendment to the U.S. constitution was validly enacted under the Article V amendment procedures. What is not now open to doubt, and what has therefore displaced any formal basis for the amendment, is the fact of common practice in applying it as supreme law. The Canadian Constitution of 1867 and its amendments of 1982 were valid Acts of the U.K. Parliament. Yet no British law is binding in Canada, and any purported repeal or amendment of those Acts by the U.K. would now be without legal force or effect in Canada, for it would not be recognized by its courts. At the ultimate level, official consensus always counts.
How can the presence or absence of common practice matter so much? After all, as your mother rightly told you, the fact that “everybody else is doing it” is not in general a reason for doing likewise. The answer is that there are special cases in which she was wrong. Common practice can be especially relevant in situations of strategic interaction, where what any one ought to do depends on what the others will do, and where what they will do does depend on what he will do. Of special interest to legal philosophers is the class of strategic interactions that Thomas Schelling and David Lewis called coordination games. Instrumentally rational agents choose strategies to produce the best outcome as they see it, but this choice is conditional on their expectations of how all others will choose, including their expectations about his expectations, and so forth. Coordination problems arise when there are at least two **coordination equilibria**: at least two sets of choices for all such that, having so chosen, at least one of them would be worse off if any one agent had chosen differently. We could all drive on the right, or on the left, so long as we follow a common rule. We would rather meet than miss, but in order to do that we need to settle on one of the equilibria. If a common practice emerged, that would give each a stable reason to comply with it and each could cite the fact of common practice as a reason for acting in the way it prescribes.

Such a practice might emerge in a variety of ways. One equilibrium might attract attention because it is far and away best for everyone, or because they agreed to it. But if there isn’t much to choose among them, or if agreement is difficult because there are too many players or because their preferences conflict, it is unlikely that any one will enjoy a natural salience. In these circumstances, some kind of “artificial” means of marking one as salient could help. There might, for instance, be a directive or order. That is a top-down solution. Or there might be a bottom-up solution: the emergence of a social norm to which most people conform because they expect most others to conform in turn. Pretty much anything that can focus attention might be a source of salience in these circumstances, and the resulting norms, though beneficial, will then have an arbitrary quality. David Hume thought that the rules determining property were like that. He said that there is a common interest in stable, clear rules governing the control of external resources, and that long practice and psychological obviousness are the sorts of things that establish such rules. Schelling thought this idea could be generalized even further: “The coordination game probably lies behind the stability of institutions and traditions and perhaps the phenomenon of leadership itself.”

It is not surprising then that legal philosophers occasionally turn to coordination games to illuminate customary rules, including the rule of recognition. For it is clear that courts do face coordination problems. Some are small-scale: they need to settle on times and dates of hearings. Some are large. The face coordination problems. Some are small-scale: they need...
matter of broad indifference what the sources of law are. Nor did he even think that the reason for official conformity to the rule of recognition is exhausted by common practice: he wanted to leave open the possibility that judges might also apply Acts of Parliament because they think democracy requires it, because the people rely on it, or because they are paid to do so.

When Coleman suggests that Hart conceived of the rule of recognition as a coordination norm, he does not mean that as a point in Hart’s favor. Though once attracted to the idea, Coleman now rejects it as conflating a possible truth with a necessary one. It would be arbitrary to say “it is a conceptual truth about law that it can exist only if officials have ex ante preferences over possible criteria of law that are structured in a certain way—one that is amenable to a certain kind of game-theoretic representation of them.”13 But that seems to leave open the contingent truth. If officials were to have such preferences, could that be the sort of normagenic situation that would give rise to ultimate legal rules? What exactly would the coordination problem be? Hart said we might think of the rule of recognition as if it emerged in response to a certain kind of social situation, one in which increased complexity within a society gives rise to doubts about what rules are binding. The rule of recognition moderates (it can never eliminate) this uncertainty. Could that be the source of a coordination problem? We need to advance cautiously here and remind ourselves exactly what the rule of recognition does: it identifies the sources of law, and it creates a duty to apply them. Let us take these in turn. First, contrary to what sometimes said, the rule of recognition does not “tell us what the law is.” The rule of recognition identifies laws; it does not determine what law requires on a given matter. The rule of recognition specifies certain norms as available for legal use and it directs officials to use them in those ways. But these norms still need to be understood and applied, and that is not a matter for the rule of recognition. Second, the rule of recognition is an ultimate legal rule of a special type. Hart called it a “secondary” rule. Since he sometimes used that term to mean a “power-conferring rule”—a rule providing the capacity to create or vary normative relations—many understandably assumed that he meant that it merely confers powers on the courts. In fact, it is a duty-imposing rule that requires judges to assess conduct by the very rules by which that conduct was to have been guided in the first place. This means neither less nor more than it says. It is not an inert identificatory practice;15 but neither is it a complete guide to means neither less nor more than it says. It is not an inert

— Philosophy and Law —

interpretation. Andrei Marmor proposes we side-step this problem.17 He agrees that ultimate legal rules are conventional, but he says they are constitutive rules only. Just as chess is a game partly constituted by the rule that the bishop moves diagonally, American law is a legal system partly constituted by the rule that duly enacted congressional statues are law. No one would suppose that the constitutive rules of chess give us any reason to play chess, so why worry about how the constitutive rules of law could give us a reason to apply law? The difficulty with this analogy is that the bishop-rule is not a mandatory rule of chess, whereas the recognition rule is a mandatory rule of law. We need to compare, say, the chess rule that it is mandatory to make a move. This does not, to be sure, give us a reason to play chess; but when playing it does indeed give us a reason to make a move. The rule of recognition gives no one a reason to be a judge, but it does purport to give judges a legal reason to apply the law. Moreover, it purports to give them a reason of a particular kind, one that is binding in accordance with their own customary practices.

What is it for a rule of this sort to be mandatory? Coleman does not give an explanation, save to say that duties cannot usually be extinguished by their subject. But this is merely a general feature of all reasons: if x is a reason for A to do it is only in special cases that A is in a position to “extinguish” the force of the reason. According to Hart, “[t]he proof that ‘binding’ rules in any society exist, is simply that they are thought of, spoken of, and function as such.”18 So what is it for rules to function as binding? Hart explains it in terms of one structural and two substantive features. A rule imposes an obligation if (a) it is categorical in form—it applies to the subject independently of his interests or wishes, and (b) it regulates matters thought important to social life and is enforced by serious social pressure.19 This inherits the general difficulty in Hart’s account of social rules, and it compounds it with a problem common to many theories that try to explain obligations in terms of their weight or importance. There are in law many obligations that are of trivial weight, and how and whether they are enforced is a highly variable. A better account, first developed by Joseph Raz, adopts the first feature, but explains the stringency of obligations in a different way.20 Obligations are categorical reasons for action that are also protected by exclusionary reasons not to act on some of the competing reasons to the contrary. Legal obligations exclude some contrary reasons—typically at least reasons of convenience and ordinary preference—but they do not normally exclude all of them: an exclusionary reason is not necessarily a conclusive reason. The stringency of an obligation is thus a consequence not of its weight or practice features, but of the fact that it mandates the required action by insulating it from the general competition of reasons. To believe she has a legal duty to apply a law, a judge need not think that this is a matter of great importance to social life or that it calls for serious social pressure; it is enough to think that it is a categorical reason that excludes acting on certain valid reasons to the contrary, such as the fact that the law in question is stupid, or inefficient, or unpopular.

Interestingly, the realm of strategic interaction does provide a possible model for this sort of reasoning, but it is not to be found in coordination games. In a prisoners’ dilemma, each has a reason to exclude acting on the principle of dominance in favor of acting in pursuit of optimality. If each pursues a dominant strategy, all attain their second worst; but if each acts under an (effective) norm of constraint, they attain their second-best, departure from which would make at least one worse off. Like a mandatory norm, a constraint norm is of a special normative type. A coordination norm, in contrast, is simply any norm that solves a coordination problem. One does
that I did not. After all, it is well-known that games of strategy
importance. I am not sure who has made this mistake; I hope
in a row suggest that Coleman regards this as a point of some
important. The theory is invulnerable to that objection, however,
since (a) does not entail (b). "The rule of recognition does not
solve the prisoners' dilemma. Law solves that problem. Instead,
the rule of recognition solves a coordination problem that must
be solved if law is to be possible." Two italicized sentences in
a row suggest that Coleman regards this as a point of some
importance. I am not sure who has made this mistake; I hope
that I did not. After all, it is well-known that games of strategy
can be nested, so that in order to solve a problem of bargaining
we may have to solve a problem of signaling—how will you
know that by "final offer" I mean "final offer"?—and to solve a prisoners' dilemma we might have to coordinate on
ways to signal trust, or to settle on an enforcer, and so forth. So
we should take care not to make the mistake that Coleman
identifies. Still, there are other mistakes waiting to be made,
the most tempting of which involves the sorts of conflict that
are possible here.

Coordination games suppose that while judges might have
preferences about what shall count as a source of law, they
also have a superordinate preference for consensus on the
issue. Yet we can all think of cases that seem to belie that.
Canada's Justice L'Heureux-Dubé was much more deferential
to administrative decisions than were her colleagues, and she
persisted in this long after powerful lines of authority had built
up to the contrary. And in the U.S. federal courts, some judges
are now on a mission not to coordinate around a settled rule
of recognition, but to try to shift it, beginning with the doctrine
of stare decisis. When we cast about for analogies with other
joint activities, "going on a walk together" or "rowing in time"
here seem less plausible than, say, "tug of war." Admittedly, it
is not necessary in coordination games that agents be strictly
indifferent among equilibria. They may well have preferences—
other things being equal—about meeting places. So many
interesting coordination games involve some conflict, as in
the so-called "battle of the sexes": he prefers boxing to ballet,
she the reverse, but each would prefer a night out together
to one alone (and, presumably, to a night out with the boys
or the girls). Various philosophers have thought that this sort of
conflict explains the need for mandatory norms in coordination
games including, most recently, Jeremy Waldron, who
considers that "for the purposes of normative analysis, partial
conflict coordination games capture the essence of politics." I
am not confident that there is an essence of politics, or if
there is that any single game of strategy provides an adequate
model for its various dimensions. But in any case this idea has
been pretty thoroughly explored by Eerik Lagerspetz, and the
responses to his arguments are still pertinent. Introducing
partial conflict does not help. The worry about coordination
games involves the role they give to overriding consensus, and
this is consistent with there being modest, overridden, conflicts
about the division of the gains of coordination. Overridden
conflicts, however, do not need to be excluded from
consideration. Compare a non-strategic case. I would rather
drink Barsac than Loupiac, other things being equal. But if a
budget-constraint gets in the way, I will drink (and enjoy)
Loupiac. This does not show that I have pre-empted or
excluded my reason to drink the more expensive wine; it does
not show that I have adopted a resolve, a decision, or a personal
rule to do so. I do not use, because I do not need, any of these
more structured forms of practical reasoning to work out what
to do. It is just a matter of rapport qualité prix. Similarly, in
the strategic interaction of a coordination game, non-salient
equilibria do not have to be excluded, for their appeal is
outweighed on the balance of reasons. Coordination norms
change that balance, they do not require or introduce
informational restrictions on practical reasoning. Mandatory
norms function by excluding action taken on otherwise valid
reasons to the contrary; but in the strategic interaction of a
coordination game what reasons are valid depends on the
mutual expectations of the players.

Consider how it is possible for legal requirements to mark
an alternative as salient. Waldron suggests:

For law to be the signaler of last resort (when informal
conventions fail to establish themselves), people
must adopt in advance a certain attitude to it. That
attitude, we may say, is respect for the law as a source
of signals of salience, which general respect is then
related dispositionally to one's respectful response
to a particular signal designating an option as salient
in coordinating a response to a question of common
concern. But if I drive in England, I do not need to adopt in advance any
general normative attitude to the law in order to have a reason
to drive on the left; the fact of common practice is there reason
enough, even though (other things being equal) I would prefer
to drive on the right. What is more, in strategic interaction, it
would be irrational to adopt a certain attitude towards law in
advance, for that depends on what I expect others will do. In
Toronto, as in many large cities, the stoplight convention is in
decay. No one slows at amber, and many take advantage of
the few seconds of delay when all lights are red to play a game
of chicken. Yet the patchiness of deference to law here is
consistent with it being able to provide salience in other areas,
including other conduct of interest to the Highway Traffic Act.
It is possible that the need for deference, and thus for
mandatory norms, comes in at another level. Raz observes
that we may be mistaken as to whether we are in a coordination
problem; perhaps city drivers do not notice the mess they are
in. If the authorities have greater expertise on this matter, and
if they are reliable and trustworthy, then we have a coordination-
related reason to defer. Unless we take their directives as
binding, we would not be able to solve, because we could not
in the first place recognize, a coordination problem in the
offing. Notice merely that this sort of argument is quite
different from one based on the prerequisites of strategic
interaction. In that case, the problem is set by the situation as
the agents see it. An argument for deference based on what is
needed to track right reason from an objective point of view,
whatever its merits, leaves the game-theoretic model far
behind.

The difficulties of identifying an independent coordination
problem and explaining the emergence of duty-imposing rules
within one are specific to the strategic model I have been
exploring here. But there may also be more general difficulties
in any theory that attempts to link customary rules to individual
attitudes of any sort. Coleman, as we saw, said that the game-
theoretic account of conventional norms mistook a possibility
for a necessity. He now suggests we might better understand them by deploying Michael Bratman’s idea of a special congeries of attitudes he calls a “shared intention” from which it supposedly follows as “a conceptual truth that officials are committed to the joint activity and to supporting one another . . .”30 It is not clear that this “conceptual truth” is in any better shape than the one Coleman rejects, and for the very same reason. It is possible that judges are oriented to the rule of recognition as to a joint activity and are trying to support each other in sustaining it; it is also possible that they are struggling to undermine each other within the limits that law and prudence allows and thus to bend the rule of recognition to their own purposes. If a customary norm determining the sources of law emerges from that struggle, its relation to the inputs of individual decision is probably more complex than any of the familiar models suggest.

Most coordination models of custom are descriptive and conceptual only. But there is no doubt that other philosophers are attracted to them by what they take to be their moral properties, and I want to conclude with a word about that. In ancient and mediaeval legal theory, custom was sometimes seen as the only true form of positive law, “mere” orders were thought to lack the requisite normative depth to make law.31 That idea still circulates in various forms. One version ties it to the presumed moral obligation to obey the law, and argues that law’s capacity to solve coordination problems, or at any rate a certain class of them, justifies our taking its directives as binding.32 Law is that which is to be obeyed, and only if it can coordinate action could it have that property. These arguments are all vulnerable to the objections sketched above. A different line associates coordination not with duty, but with freedom, an idea seen also in Burke and Hayek’s preference for evolved solutions emerge in games where each player chooses the move that best serves his interest given the anticipated moves of other players.33

Of course, that depends on what one means by “freedom” and this is not the place to explore that large issue. In thinking about choice under constraint, there have always been theories that emphasize the choice, and theories that emphasize the constraints. But in any event, we should hesitate before inferring an “is” from an “ought.” Law, including ultimate legal rules, has the character that it has, not the character that it would be nice for it to have. One of the hallmarks of Hart’s legal thought was his steadfast resistance to all sorts of wishful thinking about law; on that score we do well to follow his lead. Whatever the alleged moral properties of coordination norms, their utility in descriptive jurisprudence turns on their explanatory power alone.

The worries I have sketched here are not meant to suggest that models of strategic interaction are fruitless in legal philosophy. Quite the contrary. Theories of games and decisions have contributed enormously to our understanding of social welfare, bargaining, and coalition formation, to name only a few areas. Models of strategic interaction also train our attention on a very important set of questions about the relationship between reasons and rules. For too long, rationality-based and norm-based explanations were thought to be competing sorts of social theory, with rational economic man seen as the antithesis of an internalizer of norms.34 This artificial opposition could not be sustained,35 for it is obvious that it can be rational to follow norms and coordination games are assuredly one of the cases in which this is so. At the same time they are misleading models for the emergence of duty-imposing norms, and that is what ultimate legal rules are. Perhaps we can console ourselves with the thought that good models are not the only source of philosophical illumination. The truth table for “p→q” is a bad model for the ordinary expression “if p then q”; but thinking about what makes it bad is very fruitful in understanding counterfactual conditionals. The bearing of coordination games on customary rules of obligation is, I think, another area in which the light cast is indirect.

**Endnotes**

2. Ibid., 55-58.
3. Ibid., 266-67.
10. Coleman uses criteria of “legality” to mean what most people mean by “sources of law” rather than what most people mean by “legality,” i.e. the institutionally-specific virtues of the rule of law.
11. Hart, Concept of Law, 134.
14. Positive rules of legal interpretation are, however, part of the rule of recognition.
15. Coleman thinks that some legal positivists take this view. He claims that Joseph Raz treats the rule of recognition as having a “semantic” character only: “It guides no one’s conduct and imposes no duties.” (Practice of Principle, 84-5) In fact, Raz writes, “A rule of recognition is a rule requiring officials to apply rules identified by criteria of validity included in it.” Practical Reason and Norms, rev. ed. (Oxford: Clarendon Press, 1999), 146. And again, “[B]y imposing a duty to apply all the laws satisfying its criteria of validity, the rule of recognition imposes a duty to apply all the laws of the system.” Authority of Law (Oxford: Clarendon Press, 1979), 95.
16. Coleman, The Practice of Principle, 77. See also 85, 87.
18. Hart, Concept of Law, 231
19. Ibid., 85-88.
I. Strategic and Rational Interactions

Game theory offers us an account of how rational agents interact in strategic situations. The situations are strategic in that the determination of rational conduct for any one agent will depend upon what that agent believes another agent in the interaction might do. Of course, this other agent will also typically be working out her own strategy in light of the beliefs she has about the first agent and what the first agent might do. Thus, a rational agent needs to think not only about what to do, but also about what another agent is thinking about what to do and, further, about what that other agent is thinking about what the first agent is thinking about what to do. And so on. In non-cooperative game theory this thinking, while quite sophisticated, and having as its subject matter the thoughts and actions of another, remains very private. Each agent typically works out what to do, or how to interact with another, as a matter of private rationality.

All this is familiar. What is less familiar, perhaps, is that the law and legal theory also contain an account of how rational agents interact. However, it is an account of how rational agents interact, and how they understand their interaction, under the idea of public (or objective) reasonableness. More straightforwardly, we might say that law and legal theory offer an account of how “reasonable persons” interact. However, while this terminology might sound more familiar, it runs the risk of suggesting only that law adds some normative assessment of a rational agent’s conduct into the mix, something in which non-cooperative game theory, as a predictive or descriptive tool, would not claim to have much of an interest. But here I want to emphasize the public or objective nature of reasonableness as something distinct from private rationality, and argue that, in taking this more public orientation, the legal account provides a different interpretation of what an interaction between rational agents is, not simply how it is assessed. Indeed, I shall argue that the law’s account has important advantages over the game theoretic version for explaining the levels of coordination and cooperation that we observe amongst rational agents.¹

Effectively, I will be arguing that while game theory can claim to offer an account of how rational agents interact, unlike law and legal theory, it cannot make the further claim that it provides an account of rational interaction. As I will be suggesting this is a crucial deficiency in game theory, it is worthwhile having some sense early on of what I mean by the distinction. Certainly, a good deal of highly sophisticated reasoning and rationality is exercised by game theoretic agents. The thoughts that occur to rational players in a strategic game (“I think that she thinks that I think, etc.”) are often so complicated that it is sometimes difficult even to articulate them. Moreover, the subject matter of each player’s thoughts is the thoughts of another, and so in this sense the reasoning might even be thought to be “social.” ¹ But how “social,” or public, are they, really? A player’s rational thoughts, once replicated, are “socialized” in a sense, in that they become the thoughts of all the other players at some level. But in another sense what is replicated or socialized is still only the stuff of individual thoughts. When I am thinking about what you are
thinking, there is definitely something intersubjective going on. But, as a different point of emphasis, we have also to concede that our mutual intellectual engagement is still only intersubjective. Our private thoughts are overlapping, perhaps, in that they are "of each other" or "of each other's thoughts," but there is no real rational interaction, at least if we mean by interaction the idea that we meet one another in some public space.

Of course, the game theorist will say that interaction is at the very core of the theory of games. The players interact when each chooses a strategy, say a row or column in one of the familiar matrices depicting a two-person game, and these different choices combine, or interact, to produce a final outcome, that is, some given cell within the overall array of possibilities. What could be more interactive than that? But my claim is that, however interactive this may be, it is not an account of rational interaction. For when there is interaction between players in the theory of games, it is causal rather than conceptual; each player simply determines a part of the world (a row or a column) as a matter of individual choice, and the determination of each part determines the whole as a causal matter. But there is no interaction in a shared conceptual space. So when there is interaction it is not rational interaction.

On the other hand, when there is something rational, the rational is, perhaps, intersubjective, but it is not interactive. The players think, of course, and even think through each other's private thoughts, but they never think together in some more public, or objective, conceptual space. Thus, when there is something rational, it is not interactive, and where there is something interactive, it is not rational. There is never, therefore, any moment where there is a rational interaction.

I am sure that all this will continue to strike some readers as somewhat mysterious, especially the idea that agents can "think together," something that risks conjuring up the (non-sensical) notion of a "group mind." Perhaps it will help to see how the law provides the very sort of account of a rational interaction that I think is lacking in the theory of games. I turn to this in the next section. With this understanding of the law's account in hand, we will be in a better position in section III to see, by way of some simple examples, how such an account can help in the theory of games. The paper concludes in section IV.

II. The Objective Standard of Reasonable Interactions

In his important discussion of common law liability for unintentional harm, Oliver Wendell Holmes reminds us of the special sort of interest that law has in the idea of individual responsibility, an interest that it does not always share with ethical theory. Holmes asks us to consider the defendant in a tort action who has done the best he can to avoid injury to the plaintiff but, because of his particular ineptitude, has not been successful in doing so. On any ethical standard, Holmes suggests, it would be hard to fault the defendant for injuring the plaintiff; how can there be moral fault in doing everything one can to avoid such an injury? Yet, in a passage that is now well known and much quoted for its rejection of the relevance of these subjective abilities for a judgment of legal fault, Holmes remarks:

If . . . a man is born hasty and awkward, is always having accidents and hurting himself or his neighbors, no doubt his congenital defects will be allowed for in the courts of Heaven, but his slips are no less troublesome to his neighbors than if they sprang from guilty neglect. His neighbors accordingly require him, at his proper peril, to come up to their standard, and the courts which they establish declare to take his personal equation into account. The neighbors' standard, of course, is the objective standard of the reasonable person, and, while Holmes thought (in his typical fashion) that such a standard was ultimately justified because it was more conducive to the public welfare, the more general point was that law concerns itself with what is appropriate as a standard of behavior for the man in interaction with his neighbors rather than what is fair as a subjective matter to the man considered on his own.

This difference in what is the proper concern of law as distinct from ethics is also, of course, much emphasized by Kant. In his Doctrine of Right, or that part of The Metaphysics of Morals which deals with his philosophy of law, Kant argues that law concerns itself only with what he calls the "universal principle of Right," or the coexistence of everyone's freedom in accordance with a universal law. "[A]nyone can be free," says Kant, "as long as I do not impair his freedom by my external action, even though I am quite indifferent to his freedom or would like in my heart to infringe upon it." By contrast, "[t]hat I make [the universal principle of Right] my maxim to act rightly is a demand that ethics makes on me."5

Thus, across a broad range of theories, and over a significant span of time, there has been agreement that law (or, at least, private law) attends to what is right between the parties linked by an interaction rather than what is right or ethical as a matter internal to one of the parties. Law is not concerned with a person's thoughts in so far as they do not impact upon another, that is, in so far as they are not acted upon and do not have any potential for interaction. And when they are acted upon, and have some interactive effect, what does matter for law is not what an individual might mean to do as a private matter, but what she does as a public matter, that is, what she (actually) does under a publicly accessible, or (objectively) reasonable, understanding.

This is true for law in general, but it is particularly important for contract and consent, where parties set out to do something together, that is, when they choose to engage in cooperative activities. For example, whether two parties have a contract for the sale of, say, "new oats" or "old oats," will not depend on whether there is a meeting (or overlap) of their (private) minds on this issue. Rather, the court will attend to the most plausible public understanding of the transaction and deem the contract to be for "old oats" if that is the most (objectively) reasonable meaning of its terms in the context in which contracting occurred. Indeed, even where subjectivity does seem to be important, for example, in the criminal law, what the accused will have to attend to as a subjective matter (as a matter of, say, honest belief implicating subjective states of mind like intent, knowledge, recklessness, etc.) will be a public or shared (objective) understanding of what the concept of right conduct requires. Thus, it will not be enough in the case of a sexual assault, for example, for the accused to say (even honestly) "I thought (in my mind) that in her mind she was consenting" if there was no reasonable, or public, manifestation of that consent. The accused's appeal to his thoughts about her thoughts, while exemplifying the same intersubjectivity that is characteristic of strategic thinking in the theory of games, is inadequate because for law the subject matter of his (subjective or honest) belief is insufficiently objective. What the accused needs to be able to establish is that he had an honest belief in a reasonable manifestation of her consent, that is, he needs to be thinking about her consent as a publicly comprehensible matter.

This means that, under law, two parties who are acting together will have the separate individual actions that make up their cooperative activity linked conceptually under some objective or public understanding. Thus, if each party is to understand what her separate obligations are, or what (in law, — 75 —
at least) she should do, she will have first to consult that shared understanding of what it is that they are doing together ("Is this a contract at all?" "Is this a contract for old oats?"); and only then ask what she should do under that shared understanding of the cooperative venture. This may seem obvious enough, but it is important to appreciate that, unlike for game theory, this does mean that the interaction of the parties is a rational interaction. Each party orders her individual action in the cooperative venture as a part of a conceptual whole that likewise also orders the individual action of the other party. In this sense the parties first meet or interact together in some common conceptual space. Of course, it will be the individuals themselves who ultimately act out their respective parts of the cooperative plan as understood in this public way, and at this point the interaction (part with part) might appear to be only an interaction between rational individuals (as in the usual game) and not a rational interaction. But it is precisely because each action (while individual in the causal sense) is ordered by a shared or public conceptual scheme (part with whole) that the interaction of these (rational) individuals is a rational interaction or, as law would articulate this idea, a reasonable one. It now remains to see how this idea of a reasonable interaction might be helpful in the theory of games.

III. Coordination and Cooperation: Rational and Reasonable

Consider the following very simple two person game in which two friends, Row and Column, would like to meet for lunch at one of two restaurants, Andy’s or Bob’s. Unfortunately, they have made no prior arrangement and must choose without the benefit of having already agreed about where to go. In Figure 1, this choice is represented for each of the friends as the choice between A (for Andy’s) and B (for Bob’s), with Row choosing between the two rows and Column between the two columns. The payoffs to each friend are indicated by the numbers in each cell of the matrix, with the first number being the payoff to Row and the second the payoff to Column. Each friend is assumed to know this representation of their situation and that each is rational. Further, all of this knowledge is assumed to be common knowledge (that is, not only does each know the game form representation and that he or she is rational, but also each knows that each knows this, each knows that each knows this, and so on). Notice that there are views that Column could have about Row which would avoid any problem for Column in working out what to do. For example, if it was Column’s view that Row would simply, and somewhat thoughtlessly (compared to Column), go to the restaurant where she most preferred to lunch with Column, namely at A, then Column would have no difficulty conditioning on Row’s choice of A and would choose A himself. Note that Row does not have to actually be this sort of parametric thinker; it is enough that Column thinks she is. And if Row knows that Column thinks this of her, then she too will head for A.

But, under our assumptions, none of this is possible because Column gives Row more credit as a rational actor than that. Column recognizes that Row is every bit as rational as he is and, therefore, in this mirror image situation, he imagines her to be working out a mirror image problem. In this limited respect, therefore, Column thinks of his own reasoning as “social” (or at least “socialized”); it is replicated by other rational agents in the same situation, and that replication is reintroduced into his own thinking as something that he thinks about. Indeed, that is at least partially responsible for what is so paralyzing here. For when he thinks that like-minded Row is thinking about what he is thinking and, further, is thinking about what he is thinking that she is thinking, he realizes that there can, as yet, be no choice by Row (for example, to choose A) upon which he can condition his own choice (to choose A).  

However, consider the following variation on this game, a variation that allows Row to unambiguously choose A as a rational matter, that is, in a manner consistent with the rationality assumptions that are typical of game theory. In Figure 2, the payoffs are changed for Row and show that she would like to eat at A regardless of what Column might choose to do. Of course, just as in Figure 1, in Figure 2 Row would most prefer to eat at A together with Column. Column’s payoffs are unchanged; he would prefer most to eat together with Row at A, but failing that, just as before, he would prefer to eat with her at B rather than eat alone.

![Figure 2](image_url)

In this game Row will head for A regardless of what Column chooses to do. We say that Row has a dominant strategy to choose A. Moreover, Column (knowing the structure of the game form matrix) knows this of Row and, therefore, knows that he should go to A to meet her. (The game is solved by the method of iterated dominance; first, Row chooses A by strict dominance and, second, Column chooses A, since, given that Row chooses A, A now dominates B for Column as well.) Moreover, while Row need not think all that strategically in order to decide what to do, Row likewise knows all this of Column, and so can predict quite easily, and happily, that, as she heads for A, she will meet Column there.

But now consider a variation of the Figure 2 game that has a much less happy result. In Figure 3, the game has begun to look a bit more like a (one-sided) prisoner’s dilemma. The payoffs for each of the two friends at A have been reduced so that they now prefer, when they lunch together, to lunch at B. That is, while Row continues to prefer to lunch at A regardless of what Column chooses, when she lunches together with...
Column, she prefers to lunch at B. (Perhaps she considers the dishes at B are easier to share.) Column continues to prefer to lunch together to lunching alone, and, when lunching together, like Row, he prefers to lunch at B.

**Figure 3**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
<td>1, 1</td>
<td>3, 0</td>
</tr>
<tr>
<td>B</td>
<td>0, 0</td>
<td>2, 2</td>
</tr>
</tbody>
</table>

Can these two friends rationally get to Bob’s together? It seems not. The same reasoning that took each of them to Andy’s in Figure 2 takes each of them to Andy’s in Figure 3. Given a choice of A by Column, Row would choose A; and given a choice of B by Column, Row would choose A as well. So she must, surely, choose A. What (at least in game theoretic terms) could possibly make her choose otherwise? And, given that she chooses A so rationally, how can Column rationally choose other than to go to A as well? Thus, Row will go to A regardless of what Column chooses, and Column, knowing this of Row, will go to A to meet her.

While the result is an unhappy one for the two friends (that is, it is Pareto-inferior to the outcome where they lunch together at Bob’s), it is hard to deny that it is the rational result for them given the structural nature of strategic reasoning. Strategic reasoning, it will be recalled, conditions the rationality of a choice of action by one agent on a choice of action by the other. Put differently, but equivalently for the game form matrix in Figure 3, the rationality of an action, say, the choice of a row by Row (or a column by Column), is the rationality of that choice of row (or column) given a certain column choice by Column (or row choice by Row). And, surely, it is tempting to ask, yet again: How could it be otherwise? Rationality for each of the two friends goes to the rationality of his or her individual actions, where Row chooses row by row, and Column chooses column by column. There can be no question of what it is to choose rationally in anything other than a strictly vertical (row by row within a given column) or horizontal (column by column within a given row) direction. Choices in a diagonal direction (or choices for Row outside a given column or for Column outside a given row) are simply not available as choices for any one individual. Thus, if rationality or reasoning is to govern choices, and choices are ultimately made by individuals, then it is natural to think that the rationality of an action for one individual is the strategic rationality of that action given the choice of action by another.

However, while natural enough, strategic rationality is not the only way to think about how rationality and reasoning might govern individual choices in cases of social interaction. An alternative form of reasoning, one that is exemplified, I think, in the law’s idea of a reasonable interaction, while conceding that choices are ultimately made by individuals, might not assess the rationality of these individual actions directly, but could derive their rationality from a prior assessment of the rationality of a pattern of actions for the group of individuals taken as a whole. In other words, the rationality of a part—some individual action—is derived from the rationality of the larger whole of which it is a part. I will (hoping that I can avoid some of the baggage that comes along with this term) refer to this alternative way of thinking or reasoning about one’s individual action as holistic. But, as already suggested from our discussion of the law, one might also want to label this sort of thinking as reasonable.

This alternative form of holistic reasoning about individual choices, even when replicated across all the agents in identical situations within some social interaction, has important implications for the problems our two friends have been confronting in choosing to meet at a restaurant. Consider first the pure coordination problem that we observed earlier of our two friends, Row and Column, in Figure 1. The problem of coordination arises under strategic reasoning because what is rational for each agent to do depends so crucially on what the other agent chooses to do. Under strategic reasoning, each agent asks: what should I do (given my thoughts about what the other agent might do)? However, under the alternative more holistic approach, Row and Column each ask themselves, first: what is it that we should do here? (Each answers easily: We should meet. Where? At Andy’s.) And then they ask themselves, second: and what is it that I do when we do that? (Just as easily each answers: I should choose A.)

Of course, the first question does require each of the two friends to make a diagonal comparison in Figure 1, a comparison that neither friend can immediately follow through on as a matter of individual choice. But the second question returns each to the issue of what action he or she should choose as an individual. In particular, each friend should choose that row or that column which allows each to do his or her part in the achievement of the outcome deemed most rational (here, meaning best in terms of preferences) for the group as a whole. However, while individual actions must ultimately follow the strict vertical and horizontal contours of the game in this way, a prior assessment of what is rational in individual action need not. It is a peculiar feature of the sort of reasoning that game theory contemplates of interacting agents that it simply assumes that the empirical requirements of individual action must somehow be reproduced in an individual’s prior assessment of that action’s rationality. But there is no obvious reason why our conceptual world should track what is possible in the causal world in just this way. The causal world or, more particularly, how individuals act upon or interact with each other in some shared physical environment, could be informed by, or track, some independent judgment that we make of our actions (together) as a purely conceptual matter.

Consider now the problem in Figure 3, which is not, strictly, a pure coordination problem, although it shares the feature with Figure 1 that both friends would very much like to find a way to lunch together at one of the two restaurants rather than the other. Again, strategic reasoning identifies a dominant strategy for Row to choose A, and then, given (the predictability of) Row’s choice of A, Column chooses A as well, this despite the fact that both friends can see this coming and would be better off each choosing B. However, under the alternative more holistic approach to reasoning through their predicament, each friend again asks himself or herself: What is that we should do here? The answer for Column would appear to be as easy as it was before, only this time it identifies a different restaurant: We should meet, says Column. Where? At Andy’s. (Whether the questions and answers are in fact as easy as this for Column, who must also, even under holistic reasoning, take into account how the situation has changed for Row, is a question I will return to momentarily.) But what about Row? Given her preferences over the four possible outcomes in the game, Row has some difficulty even articulating a common conception of what the two friends might do. She seems not to be able to say easily that “We should meet for lunch. Where?” since her most preferred outcome is, first, to eat alone and, second, to do so at Andy’s. This, after all, is one of the crucial differences between Figures 1 and 3. However, she also cannot easily say that “We should eat alone,” as this contemplates the possibility that she might eat alone at Bob’s,
her least preferred outcome. What she has to say in answer to the question about what “we” should do is something less categorical and more particular, something like “We should eat alone, but only if I eat alone at Andy’s and you at Bob’s; otherwise we should eat together at Bob’s.”16

This obviously leaves something to be desired as a categorical call to rational action for the two friends.17 Moreover, it fails in this respect for two reasons. First, the call to action is so infused with particularity it hardly seems categorical at all; there really is no common conception here that informs action by each of the two friends. In other words, there is no common conception of their interaction (as reasonable) that informs their individual actions as rational (that is, that renders them “sensible” under some common conception), and it is this that we were seeking under the alternative more holistic approach. Second, given his preferences, such a highly particularized call to action by Row is hardly likely to be acceptable to Column; after all, it calls for an interaction which results in one of his least preferred outcomes. Moreover, under our common knowledge assumptions, Row can anticipate this reaction from Column. Therefore, Row can anticipate that Column will fail to find this call to action rational, at least in the holistic sense, not only because it seems so unorganized by general categories of thought that they can share, but also because, even as a particularized call to action, it is so contrary to the preferences that Column has.

However, it is only the first deficiency that I really want to emphasize here. In part this is because one can imagine Row also saying to Column that his categorical call to action (to lunch together at Bob’s) is very contrary to her preferences. Indeed, we could intensify her preference for lunch alone at Andy’s by giving her a payoff of five rather than three for that outcome, all without changing the decisive structure of the game in Figure 3 (i.e., from the point of view of strategic reasoning the game would still have the same “solution” under iterated dominance), and then Row could say of Column that his categorical call to action is more costly to her than her particularistic call to action is costly to him (supposing that these payoffs were cardinally comparable utilities and that costs are measured as departures from each friend’s most preferred outcome). So this contrary-to-preference notion of a non-rational call to action cannot do much work to distinguish Row from Column. But the non-categorical or highly particularistic nature of Row’s call to action does seem to distinguish her from Column. Column’s call to action can organize the respective contributions of each friend to the overall social interaction under some general category of thought that each can share ("What am I doing in choosing B? I'm doing my part in us getting together for lunch at Bob’s."). But in Row’s call to action (“We should eat alone, but only if I eat alone at Andy’s and you at Bob’s; otherwise we should eat together at Bob’s”), there seems to be no such general category or concept that can pose as the “whole” of which action by each friend is some part.

Notice too, if we return to the situation where Row’s preference for lunching alone at Andy’s is intensified to the point where she has a payoff of five in that outcome, that while the decisive structure of the game again remains the same from the point of view of strategic reasoning (i.e., it is solved by iterated dominance), now there is a holistic call to action that could be quite categorical for Row, and it is a call to action that would avoid the Pareto-inferior outcome where both friends lunch together at Andy’s. Now she could say, “We should each choose our restaurant such that total overall utility is maximized,” and then, by choosing A, go on to do her part in that shared conception of the friends’ interaction (again supposing that these payoffs are interpersonally commensurable cardinal utilities). Moreover, if Column shared that conception of their interaction, then he too would know to do his part by choosing B. (This illustrates how Column’s call to action under a shared conception of the interaction must take into account how the situation might change for Row; even though his own payoffs are unchanged after the payoff to Row is changed, the change in payoff to Row makes available to Column, as much as to Row, a shared categorical conception of the interaction, where Row eats alone at Andy’s, that was not available before the change.) Again, this indicates that what does the real work here in a categorical or holistic call to action, and what avoids the Pareto-inferior outcome where each friend chooses A, is that is holistic or categorical, not that it is consistent (somehow) with the preferences of all the interactors.18 After all, under an interaction where the shared conception is that each friend does his or her part in the maximization of total overall utility, a shared conception that also helps these friends to avoid choosing the Pareto-inferior outcome, the result can be, as it is here, significantly contrary to Column’s preferences. Note too that the result under this shared conception of their interaction is not merely a coordinated choice between Nash equilibria, but, more strongly, a coordinated choice of a non-Nash equilibrium.

These variations in the examples suggest, not surprisingly, that different shared conceptions of an interaction are available, and that these different shared conceptions will differently inform agents about what each should rationally (and individually) do (as (his or her part) under that shared conception. The possibility of quite different shared conceptions will also suggest that there is a good deal of indeterminacy for an individual agent in thinking about what he or she should rationally do, something that might have one wondering whether our two friends are much helped, in their attempts to get to lunch together, by these more holistic ideas. The skeptic might conclude that the problem now is only that our two friends need to coordinate their sense of “shared conceptions,” a problem that does not seem, perhaps, all that much more easy to solve than the problem of coordinating their individual actions more directly.

I do not intend to argue in any length against that criticism here, although even this conclusion does suggest that our social institutions, including our legal institutions, will have a quite different problem of coordination to address from what is more conventionally supposed. However, it does strike me that this skeptical conclusion is also very likely wrong. There will be fewer degrees of freedom for individual action under the thought that one’s (preferred) action must be disciplined, first, by a conception of what it is that one is doing and, second, by a conception of what it is that we are doing, that is, by a conception of our interaction that is capable of being sensibly shared. The different categories and concepts that organize the actions of different individuals will have to fit together in a “sensible” way, and if these categories and concepts are shared, then it seems reasonable to think that the actions of the different individual actors will have to fit together in a sensible way as an interaction. At a minimum, if there are fewer degrees of freedom for rational individual action under shared conceptual schemes than there are under individual preferences, even the common knowledge of such individual preferences, then one might expect fewer problems of coordination and, further, as our examples have suggested, fewer problems of cooperation than the standard game theoretic literature suggests.

However, even if the skeptic were to concede this point as a theoretical matter, he might still want to argue that there is little evidence that rational agents actually interact in this
way rather than the more strategic way that is contemplated by the game theorist. However, here I can be more definitive in my reply. There is good experimental evidence that agents in game theoretic situations actually do reason this way. Consider, for example, what Eldar Shafir and Amos Tversky discovered about how subjects play the familiar two-person prisoner’s dilemma game, a game very similar to what appears in Figure 3.\(^{20}\) In their experiment the rate of cooperation in the prisoner’s dilemma was 3\% when the subjects knew that the other player had defected, and 16\% when they knew that the other player had cooperated. One might well have expected some rate of cooperation between 3\% and 16\% when the subjects were uncertain whether the other player had cooperated or not. However, when the subjects were confronted with this uncertain situation, the rate of cooperation rose significantly to 37\%, a number that cannot even be rationalized as some weighted average between the strategically formulated actions “cooperate given that the other player cooperated” and “do not cooperate given that the other does not.”

Shafir and Tversky attribute this pattern of responses to the different conceptions or understandings that a subject will have of her choice situation depending on whether she knows if the other player has already made his choice of strategy. When she knows that the other player has already chosen his strategy, whether it is to cooperate or not to cooperate, the subject thinks of herself as acting “on her own.” Given the choice of the other player, she alone will determine the final outcome of the game. This encourages her to bring a highly individualistic perspective to bear on her choice of strategy, a perspective that leads her more naturally to choose against cooperation. However, in the uncertain situation, where all four possible cells of the prisoner’s dilemma game are still very much in play, the outcome of the game is to be determined by a combination of the strategy choices of both players taken together. Shafir and Tversky argue that this provides for a more collective understanding of the situation, and from this more collective point of view the optimal strategy for both parties is to cooperate. Thus, say Shafir and Tversky, it is less surprising that cooperation is chosen more frequently in this situation.

These results support the argument that individual agents in a game theoretic situation behave differently depending on how they conceive of the choice they are being asked to make. So there seems to be more than preference and information variables involved in these choices. Moreover, these differences are such that when the choices are presented to them in the causally interactive way that game theory most often contemplates (that is, when the choices are presented to them under the conception “this is still to be determined by our combined actions,” or “it’s up to us”), rather than “this is to be individually determined” or “it’s up to me”), the individual actors think about what to do as something that they should do together. That is, each thinks first under the diagonal comparison “what should we do?”, and then lets that judgment inform how he or she acts individually. In this section I have tried to suggest, by way of some simple examples, that this can help these individuals both to coordinate their actions and to avoid certain non-cooperative outcomes that might make them all worse off.

**IV. Conclusions**

The analysis that I offer here is highly preliminary. The details are certainly far from worked out, and the approach may fail fundamentally, either as a way to understanding better why people coordinate and cooperate in game theoretic situations, or as an account of what is a reasonable cooperation in the law. I hope only to be suggestive about what might be possible under a quite different approach, one that law already makes available. But if I am right in making the suggestion, the returns are large. Not only are we closer to understanding why people coordinate across multiple Nash equilibria, that is, in the cases of pure coordination games or coordination games where there might be some conflict of interest (e.g., the game of chicken or the battle of the sexes), but also we might be able to understand better why people choose to cooperate and coordinate around non-Nash equilibria, something that challenges game theory as an account of social interactions at a very fundamental level indeed.

---

**Endnotes**

1. Some of the empirical (experimental) literature that shows people have a tendency to coordinate and cooperate at higher levels than game theory predicts is nicely surveyed in John H. Kagel and Alvin E. Roth, eds., The Handbook of Experimental Economics (Princeton: Princeton University Press, 1995).
3. Philosophers are typically very careful to avoid any suggestion of a shared mental state, even when their arguments appear to bring them close to the possibility. For example, in a paper discussing “shared intention,” Michael Bratman cautions that “a shared intention is not an attitude in the mind of some super agent.” See Bratman, “Shared Intention” Ethics, 104 (1993): 97, 99. John Searle makes a similar point in his essay, “Collective Intentions and Actions,” in Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, eds., Intentions in Communications (Cambridge: MIT Press, 1990), 404: “[T]alk of group minds . . .[is] at best mysterious and at worst incoherent. Most empirically minded philosophers think that such phenomena must reduce to individual intentionality.”
7. See Smith v. Hughes (1871), L.R. 6 Q.B. 597 (a buyer refused oats that the seller had delivered on the grounds that he had meant to buy old oats, whereas the oats delivered were new). For excellent discussion of the case under the idea of a public rather than merely intersubjective understanding (“the meeting of minds” idea), see Brian Langille and Arthur Ripstein, “Strictly Speaking – It Went Without Saying,” Legal Theory 2 (1996): 63, 76.
8. The facts of Director of Prosecutions v. Morgan [1975] 2 All ER 347 (HL) come to mind. After an evening out, Morgan, the accused, invited some of his friends to come back to his house and have intercourse with his wife. Somehow, contrary to the truth, he managed to convince them that she would likely feign her resistance, but that she would be willing nevertheless. One issue was whether the friends’ honest beliefs in her consent, no matter how unreasonable in the face of her manifest unwillingness, were enough to undermine the mens rea requirement for their assaults. The House of Lords, reversing the Court of Appeal on this point (although upholding the conviction of the accused), held that they were. Needless to say, the House of Lords decision on this point of law has been controversial. For good discussion, see E. M. Curley, “Excusing Rape” Philosophy and Public Affairs, 5 (1976): 325-360.
9. Sometimes this is expressed as a requirement that his belief in her consent be honest and reasonable. But that is problematic because it obscures too much the distinction that properly exists between the fault standards that are appropriate for criminal law and those that are appropriate for private law. It is more accurate to say that in criminal law the accused must have an honest belief in what is an objectively reasonable (public) manifestation of consent. Without that he is at least guilty of a reckless indifference to what right conduct requires as a matter of law. For discussion, see Bruce Chapman “Responsibility and Fault as Legal Concepts,” King’s College Law Journal, 12 (2001): 212. Also, see Ripstein, Equality, Responsibility, and the Law, 200-14.

10. It has become (itself) conventional to credit David Lewis, Convention (Cambridge: Harvard University Press, 1969) with the first formulation of the concept of common knowledge, and Robert J. Aumann, “Agreeing to Disagree,” Annals of Statistics, 4 (1976): 1236-9, with the first rigorous formulation, demonstrating its fundamental importance for game theory. However, for a recent argument that suggests there are important differences between Lewis’s account and that which has been imported into game theory by way of Aumann, see Robin B. Cubitt and Robert Sugden, “Common Knowledge, Salience, and Convention: A Reconstruction of David Lewis’ Game Theory,” Economics and Philosophy, 19 (2003): 175-210. 11. Formally, the problem here is that there are two (pure) Nash equilibria in this game. “A Nash equilibrium is an array of strategies, one for each player, such that no player has an incentive (in terms of improving his own payoff) to deviate from his part of the strategy array”; see David M. Kreps, Game Theory and Economic Modelling (New York: Oxford University Press, 1990), 28. In a single play game, the equilibrium notion shows up in the idea that only Nash equilibria can survive, under common knowledge of rationality, the recursive thought processes of individuals thinking through the rational thoughts of another. Thus, in Figure 1, there are the two possible Nash equilibrium outcomes (hence the indeterminacy for the friends’ choice of actions), and only two possible Nash equilibrium outcomes. The “only” will be important in Figure 3.


13. Ibid., 219.

14. The holistic reasoning being contemplated here is akin to the sort of reasoning that Bacharach has characterized as “we” thinking, something that he contrasts with the more usual “I/he” thinking that we see in game theory. See Michael Bacharach, “‘We’ Equilibria: A Variable Frame Theory of Cooperation” (unpublished, June 24, 1997), 5. Sugden’s “team reasoning” is also closely related; see Robert Sugden, “Thinking as a Team: Towards an Explanation of Nonsellish Behavior,” Social Philosophy and Public Policy, 10 (1985): 69, and Robert Sugden, “Team Preferences” Economics and Philosophy, 16 (2000): 175.

15. I develop this argument more generally (that is, in a way that looks for its implications in the theories of individual choice and social choice as well as the theory of games) under the idea of “categorical reason” in Bruce Chapman, “Rational Choice and Categorical Reason,” University of Pennsylvania Law Review, 151 (2003): 1169-1210. All the discussion of the “causal” as distinct from the “conceptual” in the text at this point should not suggest that what is in play here is the same distinction that separates “causal decision theory” from “evidential decision theory.” On the latter distinction, see Robert Nozick, The Nature of Rationality (Princeton: Princeton university Press, 1993), 41-63. The evidential decision theorist takes how another might behave (e.g., a Newcomb predictor, another similarly situated prisoner in the prisoner’s dilemma, etc.) as evidence (sometimes) of how you should behave even though there is no causal connection between the other’s behavior and the payoffs you face. This sort of thinking (while allegedly more “informed” by what others are doing) is still highly individualized or private, and differs, therefore, from what is being envisaged when your conduct as an agent is informed by a shared conception or understanding of what it is that you are doing as one part of a larger whole.

16. I have tried to suggest elsewhere that the difficulties that some agents might face in articulating for others a shared conception of what they should do as a group might be useful if certain difficulties in social choice theory are to be avoided; see Chapman, “Rational Choice and Categorical Reason,” 1195-1203, and Chapman, “More Easily Done Than Said: Rules, Reasons, and Rational Social Choice,” Oxford Journal of Legal Studies, 18 (1998): 295-322. The problem, as here, is that the agent cannot organize the particular action proposed under categories of thought that the other agents might share. “Eating alone,” for example, is not unambiguously (or categorically) better, or worse, than “eating together.” In these earlier papers, I relate this problem to some of the domain restrictions in social choice theory (in particular, Sen’s “not between” value restriction) that are sometimes required if stable social choices (e.g., where there are no majority voting paradoxes) are to be ensured.

17. Of course, no actual “call to action” is being contemplated here, as there is (by assumption) no actual conversation between the friends. Rather, what is being contemplated by each friend is only a thought experiment, albeit one that is a little more “public” than what goes on in strategic reasoning. The friend asks herself (silently) whether her conception of what they are (reasonably) doing (or what they ought to do) is something that the other friend can (reasonably, sensibly) be expected to share. Only then does the conception order what each will do together. Row should see that her call to action is not one that is likely even to occur to (or be conceived by) Column.

18. This also shows that what is being proposed in this paper to solve coordination games where one of the Nash equilibria is Pareto superior to another is to be distinguished from the mere addition of something like a joint dominance principle. For examples of the latter approach, see David Gauthier, “Coordination,” Dialogue, 14 (1975): 195, and (for a cautious endorsement of this approach in limited circumstances) Kreps, Game Theory and Economic Modelling, 31. A shared conception of oneself as a total utility maximizer will avoid Pareto-inferior outcomes without necessarily endorsing the relevant Pareto-superior outcome. In other words, under such a shared conception, some (but not all) agents might be made worse off.

19. Two points need to be emphasized. First, when the law addresses coordination in the way suggested here, it does not address a coordination game, that is, the problem of how best to coordinate actions by individuals (e.g., say, by making some actions more salient than others). On this it contrasts with the interesting argument in Richard H. McAdams, “A Focal Point Theory of Expressive Law,” Virginia Law Review, 86 (2000): 1649-1729. Rather, the law should address more general understandings or conceptions of action and interaction, developing and/or reinforcing these general understandings so that coordination games are solved as a matter of individual reasoning, through a particular shared conception, to the choice of an action that is not otherwise the best. On this it helps individuals to avoid more than just the problems modeled by coordination games, even coordination games where there is some conflict of interest (as in the game of chicken or the battle of the sexes game, both discussed by McAdams, supra). If the examples in this section are any indication, a shared conception of their interaction may well help the interacting agents to solve problems modeled by the prisoner’s dilemma as well. That is, such an approach might encourage the agents to choose not only
If the Price is Right: Unfair Advantage, Auctions, and Proportionality

Michael Ridge
University of Edinburgh

At one point in England, it was a capital offense to “appear on a high road with a sooty face.” I do not know whether anyone was executed for this offense, but many people were sent to Australian penal colonies for such petty crimes as stealing a handkerchief. More recently, Kenneth Payne was sentenced to 16 years in prison for stealing a Snickers Bar in Texas. When the Assistant District Attorney in this case was asked how she could justify putting a man in prison for 16 years for stealing a candy bar she replied, “It was a king size,” and indeed it was—value, $1.00. These punishments are vulnerable to many criticisms but the most obvious is that they do not fit the crime. Unfortunately, the idea that the punishment should fit the crime is as difficult to characterize precisely as it is plausible. The doctrine of lex talonis embodied in the slogan “an eye for an eye” provides the most obvious answer to this question but it has the unfortunate virtue of being false. In some cases the problems with lex talonis are moral ones—we would debase ourselves if we were literally to rape rapists. In other cases the problems are conceptual—taken literally, the theory seems to have no intelligible application for a crime like treason, for example. These problems are familiar and defenders of lex talonis typically qualify the doctrine in ways that are supposed to avoid them. However these qualifications typically are either ad hoc or too metaphorical to be helpful. Intriguingly, Michael Davis argues that we can formulate a more precise and non-metaphorical account of proportionality by applying some ideas from game theory. In a series of articles culminating in his 1992 book To Make the Punishment Fit the Crime, Davis defends the ingenious idea that proportionality should be understood in terms of an auction for licenses to commit crimes. The guiding idea behind Davis’s account is that a criminal takes unfair advantage of others’ obedience; a punishment fits a crime insofar as it annuls this unfair advantage. Davis is not alone in his appeal to the idea of unfair advantage to explicate proportionality. What is unique and interesting (some might say eccentric and bizarre) about Davis’s account is its game-theoretic model of unfair advantage.

After characterizing Davis’s version of the auction theory in more detail (section I), I argue that what is often taken to be one of its most serious vices can easily be avoided, that it faces other more insoluble problems as a theory of justice but might nonetheless be useful as part of a theory of deterrence. As it stands, Davis’s version of the theory is implausibly committed to supposing that a society that chose to punish unlicensed murderers and unlicensed shoplifters in the same way could be reasonably just. However, those favoring the auction account should simply drop the assumption that a society can provide an adequate model of unfair advantage only if it is reasonably just (section II). The auction theory nonetheless fails as a general theory of proportionality (section III) because of its exclusive focus on unfair advantage. Intuitively, criminal punishment should be proportional to wrongness more generally. This suggests that the auction theory might nonetheless provide useful guidance and illumination if its scope were dramatically narrowed to include only crimes that are morally wrong only because they involve an unfair advantage. Perhaps in those cases, at least, punishment should be proportional to unfair advantage since unfair advantage exhaustively characterizes its wrongness. However, the auction theory fails even if we restrict its scope in this way; proportionality in these cases is a function of the importance of the public good at stake rather than the unfair advantage taken (section IV). I conclude with some brief and speculative thoughts as to why the auction theory seems to fit well with actual criminal practice (Conclusion).

1 The auction theory is a theory of proportionality in criminal punishment and is compatible with a variety of views about the justification of punishment. Moreover, the auction theory is compatible with a wide range of views about what the criminal law should prohibit. Finally, the theory concerns criminal punishment; it is compatible with a range of views about appropriate civil penalties and non-legal punishments. Davis’s version of the auction theory aims only to explain how to determine the maximum punishment allowed by considerations of proportionality. The theory is silent on the issue of whether proportionality also requires minimum punishments. Finally, whether a particular crime deserves less than the maximum set by the theory should be determined by judges in virtue of other considerations.

The theory aims to explain proportionality in terms of unfair advantage understood in terms of the auction model. The theory presupposes that crime constitutes a kind of free-riding. Insofar as it is reasonably just, the criminal law is a clear case of a mutually advantageous cooperative practice and failing to do your share in a mutually advantageous cooperative practice is free-riding. Hobbes’s dystopian vision of the state of nature may be overblown but we would be much better off with a reasonably just system of criminal law than we would be without one. An egregiously unjust system of criminal law may make some worse off than they would be in a state of nature—legally enforced slavery is a good example. However, crimes committed by those disadvantaged by the law in this way do not deserve punishment simply in virtue of the fact that they are crimes. Davis argues that such crimes deserve punishment, but only because they are morally wrong independently of the positive law. Davis suggests that these cases are ones of purely moral desert rather than criminal desert, and hence fall outside the scope of his theory. The distinction between moral desert and criminal desert invoked here is problematic, and I return to it below (section IV).

In determining the unfair advantage associated with a given crime we must exclude the incidental advantage associated with it. Rather, the issue is what sort of advantage the criminal gains simply in virtue of violating the law. The auction theory maintains that the best measure of the unfair advantage associated with a given crime is how much a license to commit that crime would fetch in an open market in a reasonably just society with a fixed number of licenses is available. The suggestion is that the unfair advantage associated with a crime varies directly with the price a license to commit the crime would fetch in such a market. Of course, price is a function of both supply and demand, so the theory needs a criterion for the number of licenses to be offered. Davis suggests that the number of licenses for a given crime should correspond to the prevalence of that crime the society is willing to tolerate given the estimated costs of prevention. Politicians score points with rhetoric of “zero-tolerance” but any serious attempt at the complete elimination of crime in a modern nation-state

---

would be futile, unimaginably financially costly, and involve
the widespread violation of civil liberties. Ideally, the auction
model provides an ordinal ranking of crimes in terms of unfair
advantage. If a license to commit first degree murder would
cost more than a license to commit assault then the maximum
punishment for murder should be greater than the maximum
punishment for assault. Davis argues that this account does
better at legitimizing actual legal practice than its rivals.

One problem with Davis's formulation of the auction
theory is the obscurity of the idea that a criminal necessarily
reaps an advantage simply by committing a crime regardless
of any so-called "incidental" benefits. The thought is that
simply by committing a crime a person gets a certain advantage
regardless of whether the crime succeeds. As Davis's critics
have argued, this is a strained use of "advantage." If my
attempted robbery fails then intuitively I have gained nothing
by it. The objection reveals that the main idea behind the
auction theory is, contra Davis, not actual unfair advantage.
We should instead understand the auction theory as measuring
the expected advantage associated with a given sort of crime.
My particular attempted robbery may provide no advantage at
all. Nonetheless, my crime falls under a sortal recognized in
the law—robbery—and crimes falling under that sortal can be
assigned an expected value as a class. The expected value of
a crime is general in two respects. First, its content is general—
the expected value is of robbery in general. Second, it is the
expectation not of this or that particular criminal but of society—the expected value we would assign to robbery. It
remains a good question why punishment should be proportional to unfair expected advantage so understood; I
return to this question below (section IV).

II

Unless the auction model includes a penalty associated with
committing an "unlicensed" crime, there is no obvious reason
to buy criminal licenses. The auction theory faces a trilemma.
Either the theory imposes no punishments for unlicensed
crimes or it imposes different punishments for different
unlicensed crimes or it imposes the same punishment for all
unlicensed crimes. The first option is a non-starter since it
provides no incentive to buy licenses in the first place. The
second option raises the question—what is the standard by
which these different punishments are arranged? If the
standard for those punishments is fixed by another lottery
then a vicious infinite regress threatens. If, on the other hand,
those punishments are fixed by some auction-independent
standard then it is unclear why we could not simply use that
standard to fix punishments in the actual world. After all, on
Davis's account the society in which the hypothetical auction
takes place is supposed to be reasonably just, so the
punishments assigned to crimes there must also be reasonably
just. The third option, however, is incompatible with Davis's
assumption that the society in which the auction takes place
is reasonably just. For any society that punishes unlicensed
jaywalking and unlicensed murder identically is not reasonably
just. Such a society will either assign penalties that are far too
weak for some crimes or far too severe for others. A society
that punishes unlicensed murder with no more than a small
fine is not reasonably just. Perhaps even more clearly, a society
that punishes unlicensed jaywalking with the death penalty or
life imprisonment is not reasonably just. Indeed, such a regime
would be draconian in the most literal sense. For "draconian"
comes from an Athenian lawgiver Draco who punished all
crimes with death: "[W]hen Draco was asked why all of his
decreed punishments were the same, he apparently replied
that he thought small offences obviously deserved death, but
that he couldn't think of a worse penalty for more serious
crimes."

Critics of the auction theory have not been silent on this
point. Davis's considered response is to embrace the third
horn of the trilemma, arguing that such punishments are not
so unjust as to render the entire scheme unjust because
everyone had the option of buying a license to commit the
crime in question. This is unconvincing. If Draco had offered
a Davis-style auction then perhaps his regime would have been
less unjust but monstrously unjust nonetheless.

Davis is fighting an uphill battle on this front for no good
reason. He is right to embrace the third horn of the trilemma
but wrong to hold onto the idea that the auction must be
understood against the background assumption of a reasonably
just society. For nothing in the theory requires that the society
be reasonably just. After all, the auction itself is not a model of
justice in any sense. Quite the opposite, it is a model of the
unfair advantage (better: unfair expected advantage)
associated with a particular sort of injustice —free-riding. Nor,
in any event, would a model of justice necessarily need to be
just. As Davis himself has taken pains to emphasize, models
are not just like the subjects they model. A model of
microscopic particles does not need to be microscopic, so
why must a model of justice be just?

So far as I know, Davis has never given an argument that
the society in which the auction takes place must be reasonably
just. The only reason I can imagine for such an assumption
would be if we were supposed to carry out such an auction to
determine appropriate punishments. However, Davis explicitly
rejects this interpretation, claiming that the auction was not
intended as a real possibility but as a model. Moreover, if
the theory did require us to carry out such a bizarre experiment
then it would clearly be hopeless. For such an experiment is
politically impossible precisely because it would be so unjust.
It is unjust not only because it would require that we punish all
unlicensed crimes in the same way, though that is already a
serious injustice. It is also unjust because it would allow those
convicted of very serious criminal wrongdoing to avoid criminal
punishment altogether. Imagine telling a woman whose son
was raped, tortured and brutally murdered that although we
know who committed the crime and can prove it and have
him in custody, we unfortunately will have to allow him to go
free because he had licenses for all of his crimes! Such a
woman would quite rightly be morally outraged, nor would
the availability of civil litigation adequately address her outrage.
Clearly, this is not justice but a mockery of it. Indeed, I cannot
resist pointing out that something close to this appears for
comic relief in Terry Pratchet's famous and popular "Discworld"
series of amusing fantasy books. Here is a representative
passage from Pratchet's Guards! Guards!:

One of the Patrician's greatest contributions to the
reliable operation of Ankh-Morpork had been, very
early in his administration, the legalizing of the ancient
Guild of Thieves...In exchange for the winding down
of the Watch, they agreed...to keep crime levels to a
level to be determined annually. That way, everyone
could plan ahead, said Lord Vetinari, and part of the
uncertainty had been removed from the chaos of
life...a complicated arrangement of receipts and
vouchers saw to it that, while everyone was eligible
for the attentions of the Guild, no-one had too
much. It is with good reason that Pratchet presents us with this scenario
in a farcical story; it is funny in part because it is so absurdly
unjust. Davis is right not to suggest that we should try to adopt
his model in the real world precisely because it would be
unjust. Fortunately, we can drop the assumption that the
society in which the auction takes place must be just.

— 82 —
III

The auction theory may well provide a good model of the expected unfair advantage associated different crimes, but does expected unfair advantage provide a good account of proportionality? As Davis puts it, what we want is a theory of “how much to punish criminals” under the laws of a relatively just society.14 The auction theory’s answer is, “Punish criminals who violate just laws in proportion to the unfair advantage associated with the crime.” The most obvious problem with this answer is that intuitively criminal punishment should be proportional to the overall wrongness and not merely unfair advantage. To take an unfair advantage of others is a kind of free-riding. However, being a free-rider is only one sort of moral wrongdoing, and often not the most serious sort at that. Admittedly murderers, rapists and child abusers are free-riders and that this is part of why their crimes are wrong. Still, this is only a small part of the wrongness of those crimes. That they take the life of an innocent person, that they violate another person’s bodily integrity and treat them as a mere object, for example, seem like much more serious wrong-making features. Intuitively, the primary reason a murder deserves greater punishment than a shoplifter is that taking an innocent person’s life is a much more serious sort of wrongdoing than shoplifting. It might turn out that murder also happens to be a form of free-riding associated with a greater (expected) unfair advantage, but this is not why murder deserves a greater punishment than shoplifting. Even if the auction theory gets the right results, it gets them for the wrong reasons.

Nor is it entirely clear that the theory gets the right results. The main theoretical problem is that criminal punishment should be proportional to wrongdoing tout court and not merely wrongdoing qua free-riding. The limiting case of this problem is a crime that does not involve free-riding at all. Consider the following disturbingly timely example. A terrorist masterminds an attack against the United States from overseas which kills several people. This hypothetical terrorist (unlike Osama bin Laden) has never reaped any benefits from the United States nor has he reaped any benefits from the existence of international law. The United States manages to capture this terrorist and prosecute him for murder. The case is an unusual one but very plausibly the terrorist deserves criminal punishment. However, if we suppose that maximum criminal punishment must be proportional only to unfair advantage then he deserves no punishment whatsoever. For given that he never reaped any benefits from the United States system of laws he is not guilty of free-riding and has not taken unfair advantage of Americans’ compliance with criminal law. Here we have a case in which severe criminal punishment is warranted but where the auction theory implies that any criminal punishment is too much. Whereas if we held that punishment should be proportional to wrongness tout court we could easily maintain that the terrorist deserves severe criminal punishment because of his wanton destruction of innocent human life regardless whether he is guilty of free-riding. There are much worse things than being a cheater. I emphasize that this example is a limiting case and that the main and more general point is that even if the auction theory does get the right results in other cases then it does not get them for the right reasons.

Davis is aware that free-riding is not a measure of overall moral gravity.15 He draws a distinction between morally deserved punishment and deserved criminal punishment and argues that our intuitions in these sorts of cases are intuitions only about morally deserved punishment and not intuitions about deserved criminal punishment as such. Since his theory is a theory of deserved criminal punishment as such these intuitions are supposed to be irrelevant. Here is Davis: “Those who do not benefit from the criminal law can deserve punishment only insofar as their offences are immoral as well as illegal. They could, for example, deserve punishment for murder but not for sleeping in the park.”16 It is true that those who do not benefit from the criminal law can deserve punishment only insofar as their offences are immoral as well as illegal, but this is equally true of those who benefit from the criminal law. It is just that those of us who benefit from the law act immorally whenever we break a just law from which we benefit—that is free-riding. Perhaps Davis’s point is that while those who do not benefit from the criminal law do deserve punishment when they act wrongly, they deserve this punishment regardless of the criminal law. The examples seem to support this interpretation; murder is wrong regardless of the law whereas sleeping in the park is wrong (when it is) only because of the law. However, it does not follow from the fact that someone does not benefit from the law that he cannot do something that is wrong only because of the law (mala prohibita). The suppressed premise of the argument seems to be that all actions that are wrong only because of the law must be wrong in virtue of being free-riding, but this is not so. The rules of the road are a good case in point. Driving on the incorrect side of the road is for most people a form of free-riding but it is also knowingly risking grave harm to others, and it is morally wrong for both of these reasons. It risks grave harm to others independently of the law, though. So its falling into this moral category is dependent on the law. If someone who did not benefit from the criminal law at all (a runaway slave, say) were to drive on the wrong side of the road then his action would still be wrong only because of the law (mala prohibita) but it would not involve free-riding.

The more basic point, though, is that Davis’s theory addresses the question, “how much to punish criminals” under a just system of laws and it is clear that this is a moral question. As Davis’s example of punishing someone who does not benefit from the criminal law for murder but not for sleeping on a park bench strongly suggests, we do have a reasonable albeit vague idea of moral desert of punishment that goes beyond moral desert for free-riding. Insofar as that more general idea is sound, though, there is no reason it should not be relevant to determining the maximum criminal punishment someone morally deserves for a given crime. It is not as if our more general moral notion of deserved punishment is not general enough to catch criminal punishment in its net. Intuitively if one person’s crime is immoral only because it is a case of free-riding then intuitively her maximum criminal punishment should be much lower than someone whose crime is also immoral because it is murder. Davis’s position seems unstable. His discussion of those who do not benefit from the criminal law strongly suggests that we can make reliable moral judgments about what the maximum penalties appropriate for forms of wrongdoing other than free-riding17 but he also denies that those intuitions are relevant to determining the maximum criminal penalty. If the intuitions are generally sound and take criminal punishment within their scope, it is arbitrary simply to exclude them from a theory of proportionality. The result may be a theory that is messier than the auction theory, but we should not sacrifice truth for precision.

Nor is it useful to make the terminological appeal to “criminal desert” as opposed to moral desert more generally. For our question, as Davis puts it, is “how much to punish criminals” and this question should take into account moral desert more generally and not mere criminal desert.18 “Criminal desert” is stipulated to include only moral desert associated with free-riding. A theory of mere “criminal desert” in this sense simply would not answer the more general question of

--- 83 ---
“how much to punish criminals” in precisely the sense relevant to judges and legislators. A judge might reasonably respond, “OK, this is how much criminal punishment he deserves because what he did was a crime associated with such-and-such unfair advantage, but how much more criminal punishment does he deserve because he murdered someone?” Substantive concerns cannot be avoided by terminological manoeuvres.

In a discussion of methodology in punishment theory, Davis suggests that the justification of a theory should be comparative, otherwise we may overlook the possibility that even the best theory has fairly serious vices. In this spirit, the critique offered here is comparative. My suggestion is that a theory of proportionality that holds that maximum criminal punishment should depend on overall wrongness is more plausible than one that makes it depend only on wrongness qua free-riding. I hasten to add that it is actual degree of overall wrongness that matters on this theory, and not merely what sociological surveys suggest most people take to be the degree of overall wrongness; Davis argues that the latter approach has unacceptable consequences.

How we determine actual overall wrongness is a difficult question and requires serious moral philosophy as well as various forms of empirical inquiry. Undoubtedly there will be reasonable disagreement about such matters, but it is equally clear that there will be reasonable disagreement about the outcome of an auction for licenses to commit crimes. Davis at one point suggests that such a theory would force judges to pry unacceptably into people’s personal lives to determine how wrong they acted, but this conflates the distinction between the wrongness of an action and the wickedness of the agent. We need not assess someone’s character to determine the wrongness of her action though we do perhaps need to determine her beliefs and intentions. This move also ignores that the theory might well focus primarily on types of crimes rather than particular tokens (as Davis’s own theory does). A focus on the wrongness of types rather than tokens would also assuage some of Davis’s worries about predictability. Such a theory is admittedly less precise than Davis’s but this seems like a less serious vice than handling cases in a highly unintuitive way.

### IV

Perhaps the argument of the last section illustrates only that Davis’s version of the auction theory is too ambitious in its scope. If the problem is that the theory only gauges free-riding perhaps the solution is to narrow the scope of the theory to crimes that are wrong only because they involve free-riding. The resulting theory is admittedly less sexy than the original but sexiness is a less important theoretical virtue than truth.

Unfortunately, this move does not work. To see why it fails, we must return to the question, why should penalties for free-riding be proportionate to the extent of unfair (expected) advantage associated with it? That is the guiding idea behind the auction theory and it does sound plausible at first blush. However, on deeper reflection it looks deeply problematic. For intuitively, the maximum penalty associated with a form of free-riding should correspond to the importance of the public good at stake rather than in terms of unfair advantage. Still, the auction theory does seem to do an impressive job of predicting actual criminal penalties, as Davis argues in many places, and this does not seem to be entirely accidental. I therefore tentatively conclude that the auction theory fails as a theory of proportionality even if its scope is restricted to the kinds of crimes for which it is tailor-made.

### Conclusion

In spite of its apparent attractions the auction theory fails as a theory of proportionality. So far as justice goes, punishment should be proportional to overall wrongness and not merely wrongness qua free-riding. Even in cases of pure free-riding, justice requires that punishment be proportional to the seriousness of the crime understood in terms of the value of the public good at stake rather than in terms of unfair advantage. Still, the auction theory does seem to do an impressive job of predicting actual criminal penalties, as Davis argues in many places, and this does not seem to be entirely accidental. I conclude with some speculative thoughts on this point, which suggest that the theory may be a useful one after all.

The ability of a moral theory to legitimize existing practice is a virtue only insofar as existing practice actually is legitimate; existing practices can be unjust even when widespread. In some of the cases Davis discusses, this is relevant. For example,
the practice of punishing unsuccessful criminal attempts less seriously than successful ones is problematic precisely because attempted murder seems just as wrong as murder. As is often noted in the literature on moral luck, either it is just an accident that one person’s attempt is successful or it is merely evidence of some other feature of the case (wholeheartedness, perhaps) that is the genuinely morally relevant one. The latter possibility would not warrant different punishments in general for unsuccessful attempts but rather indicates that juries should consider the failure of an attempt as evidence of some other feature—half-heartedness in the attempt, say. After all, in many cases this merely defeasible evidence of the presence of some other feature does not warrant the inference that the other feature really is present. In some cases, it really is just a matter of luck that one person’s attempt failed where another person’s was successful. In those cases, at least, it is intuitively implausible to think that justice would allow a higher maximum penalty for success as opposed to failure. We might be willing to live with such a scheme for utilitarian reasons but this just highlights that justice is but one (admittedly important) value amongst many. The issue of moral luck is a famously difficult and vexed one and I cannot do justice to it here. It is, however, worth registering that Davis’s argument that his theory would legitimize existing legal practice in this area may be more of a vice than a virtue.

Still, the question remains, “Why does the auction theory seem to do so well at legitimizing actual ordinal ranking of the severity of punishments insofar as they seem reasonable?” as Davis argues it does. Here it is worth reminding ourselves that the unfair expected advantage associated with a given crime very clearly is an indication of the strength of the incentives to commit such crimes. Of course, the actual incentive of a particular criminal to perform a particular crime will depend on an enormous number of idiosyncratic features, but it would be impossible for the criminal law to try to deal with such highly variable features. Perhaps the auction theory provides a very rough and useful guide to the strength of the incentives to perform various crimes. In that case, there is another explanation as to why Davis’s theory gets the right results when it does. All else being equal, severity of criminal punishment should vary directly with the incentive to commit the crime. Otherwise the law is unlikely to deter crime in a way that is efficient, humane and consistent with genuine constraints of proportionality (not understood in terms of unfair advantage). If we aim to avoid unnecessarily draconian laws and an overtaxed criminal justice system then we should try to find the minimum punishment needed for deterrence. Perhaps this explains why the auction theory legitimizes existing legal practice as well as it does, particularly since rough considerations of deterrence historically have played a large role in setting criminal punishments. For the greater the incentive associated with a given crime, the more severe the punishment needs to be, all else being equal. So if the auction theory succeeds in giving an ordinal ranking of the expected advantage associated with a crime then it should be no great surprise if actual schemes of criminal punishment were roughly isomorphic with this ranking.

If this is correct then one of the main apparent advantages of the auction account—a its ability to fit with existing legal practice—provides no support for the theory as a theory of proportionality. For on the proposed explanation, the theory tracks existing practice in virtue of rough considerations of deterrence rather than considerations of proportionality. This also suggests that we should not be too quick to abandon the auction model. For it might well be useful as part of a larger theory of deterrence. This would be a very different function from the one Davis proposes. On Davis’s account, the auction model helps determine maximum penalties but on this account it would help determine the minimum penalties needed for deterrence. Moreover, instead of determining proportionality and justice, the theory would now determine punishments in accordance with the utilitarian need for deterrence. Since deterrence often seems to conflict with justice, this is an ironic reinterpretation of the auction model but a potentially fruitful one nonetheless. We remain indebted to Davis for articulating the model even if we “kick away the ladder” of proportionality and replace it with a less grandiose scaffolding of deterrence.

Endnotes

6. See Davis, To Make the Punishment Fit the Crime, 246-7.
8. Actually, there is the fourth option of simply stipulating that people in the auction would not poach and know this about themselves, but this seems rather artificial and raises awkward questions about how their psychology differs from the psychology of actual people such that they would never poach. Davis rejects this account as unrealistic and I shall follow him in putting it to one side.
11. See Davis, To Make the Punishment Fit the Crime, 255-256.
14. Davis, To Make the Punishment Fit the Crime, 244.
15. See, e.g., ibid., 249.
16. Ibid., 248.
17. This also seems implicit in his discussion of non-criminal punishment in extra-legal settings, such as punishment of someone in a concentration camp by his fellow prisoners. See ibid., 237-8.
19. See, e.g., Davis, To Make the Punishment Fit the Crime, ch. 2.
20. Ibid., 31.
21. Many thanks to Sean McKeever for emphasizing that the objection I am pressing here actually is independent of the dispute about criminal versus moral desert raised earlier.
Michael Ridge’s "Unfair Advantage, Auctions, and Proportionality" is an extended critique of my theory of setting justified statutory penalties ("the fairness theory"). His focus is the auction I use to show that the theory sets penalties according to criminal desert (a certain kind of moral desert), a showing necessary to preserve my credentials as a retributivist. Ridge has nothing to say about the seven-step method by which I claim legislators should, and generally do, set statutory penalties, or about my derivation of this method from what I identify as the presuppositions of the criminal law, or about my explanation of the relation of these presuppositions to any justification of the criminal law, or about other important elements of the theory. This silence on so many important topics would be disappointing in a book, but they are inevitable in an article of reasonable length. I mention his silence only to place what he does say in context.

What is disappointing (to me, at least) is that Ridge rejects almost every element of the theory he does discuss. In compensation, he not only gives a fair statement of part of the theory, even suggesting its appeal, but also an opportunity to dispose of objections that may have occurred to others. I propose to use that opportunity, but not to refute his objections one by one. While his errors are few, their kin are not. Disposing of objections one by one merely opens the way for others. What I propose to do instead is something likely to be more effective and interesting. I propose to identify errors of method or understanding upon which Ridge’s objections depend and thereby both dispose of his objections wholesale and prevent many others now quietly gestating somewhere.

1. Intuition

Like many philosophers, Ridge appeals to “intuition” at crucial points in his argument. Intuition, as he uses that term, seems to be a mere feeling or impression that such and such is so, a spontaneous judgment having epistemic authority independent of argument. Since Ridge does not seem to believe in a priori knowledge of such matters as whether a criminal deserves less punishment for an attempted robbery than for a successful one, any epistemic authority intuition has in his argument must rest on experience. To support epistemic authority, the experience in question must have a reliable connection with whatever the intuition supposedly reveals. Our intuitions about familiar situations, especially if shared by others, do (generally) have such a connection. Experience disciplines intuition. Experience that disciplines everyone’s intuition more or less the same way seems to belong to the same epistemic category as perceptions of color, shape, texture, and so on. They are good (but not decisive) evidence for what they seem to tell us.

We may then distinguish at least four kinds of intuition that lack the appropriate epistemic authority: (a) intuitions not resting on the right experience; (b) pseudo intuitions (that is, mere verbal formulas); (c) intuitions not shared (or, at least, not shared widely enough); and (d) intuitions open to more than one interpretation. Let us take these in order.

(a.) An appeal to intuition must rely on the right experience. For example, if asked to consider what we would say if we were in a certain situation, that situation must be enough like what we have experienced for our response to be likely to reflect what we would actually say in the situation were we put into it. Where we lack similar experience, our response lacks the discipline experience provides and, lacking that, lacks the corresponding epistemic authority. An appeal to intuition is (more or less) an implicit argument from analogy.

At several crucial points in his argument, Ridge appeals to an intuition that does not meet this standard. In each case, his purpose is to show that our intuitions are at odds with the results of the auction I use to model fairness. Consider, for example, Ridge’s request that we “[imagine] telling a woman whose son was raped, tortured, and brutally murdered that although we know who committed the crime and can prove it and have him in custody, we unfortunately will have to allow him to go free because he had licenses for all of his crimes” (§II).

Imagine such a woman in our society, I have the same intuition Ridge does: “this is not justice but a mockery of it” (§II). But, of course, this society (the only one of which we have direct experience) does not have the system of licenses I imagined to make an auction possible. Though perhaps a reliable indicator of what we would think if aliens imposed such a system on us tomorrow, our intuition of extreme injustice is unlikely to be a reliable indicator of what we would think had we grown up under the system imagined. We have no direct experience of the society in question or of one much like it.

Why do I think intuitions of injustice are likely to differ when experience of justice differs? Anthropology and history. Consider the very spot where Ridge had his intuition (Edinburgh): Only a thousand years ago, a woman living there, in what was then Saxon Northumbria, would—or so it seems from what we know of that time and place—have accepted as full justice for the brutal murder of her son a “wergild” or “man price” (that is, a payment negotiated between the victim’s family and perpetrator’s family). Those who live in a society where murderers must pay blood money rather than suffer the death or imprisonment we consider just do not necessarily share our intuitions of justice. Our intuitions about their practices should yield to those of people more familiar with them. Only our arguments need not yield to their intuitions.

Among the arguments to which Ridge might appeal to dismiss what the anthropologists and historians report about other people living under other systems of justice is that the satisfaction reported arises from “false consciousness.” They are satisfied because, and only because, they lack experience of better institutions. Had they experienced our institutions as well as their own, they would prefer ours precisely for the reasons Ridge (and I) would; they would see ours as more just.

I agree that false consciousness is a possible explanation of Saxon satisfaction with wergild in Northumbria. But I do not agree that is the best explanation. I have two reasons. First, Ridge’s claim that false consciousness is the best explanation seems to rely on an experiment we cannot perform. Absent an empirical test, dismissing Saxon intuitions of justice in favor of our own seems to be no more than a polite way of stating a prejudice. We have not put their intuitions, or ours, to the test that a claim of false consciousness deserves. Second, and more important, we have good reason to doubt that the test would
come out as Ridge predicts. Not all modern people, not even all the educated and thoughtful, prefer our way of doing justice to the Saxon. At least some recent calls to privatize justice suggest the Saxon system as a model we might want to copy.\(^2\) So, for all we know, were we to bring a sample of Saxons from their time to ours, many, perhaps most, might prefer their system of justice to ours even after experiencing both. Ridge might, I suppose, still reject their intuitions in favor of ours, but he would then have to claim that some false consciousness is incorrigible. Perhaps some is, but any argument that tries to distinguish between such false conscious and mere reasonable difference of opinion is likely to have enough problems to make it useless in criticism of anything I have said.

Of course, the intuition that Ridge appeals to is not of any actual society but of one we merely imagine. We are free to imagine the society anyway we like. But we are not free to imagine the intuitions of its inhabitants anyway we like if we want those intuitions to carry weight in our argument. Any appeal to intuition must (at least implicitly) rely on actual experience, not on ours necessarily but on that of those as close as possible to what we want the intuitions to be about. Where those with experience much closer to the imagined society’s experience than ours seem to have an intuition that differs from ours, we are not entitled to rely on our intuition. Or, at least, we are not entitled to rely on it without further argument. Ridge gives none. Hence, we must dismiss the intuition he appeals to, even if we all share it. We must also dismiss any argument that relies on such an intuition.

(b.) An appeal to intuition is not a mere verbal formula; it must appeal to an actual intuition. An intuition, we agreed, is a judgment or feeling disciplined by a certain range of experiences we actually have had. So, where there is no possible experience corresponding the intuition appealed to, there can be no intuition (or, at least, no intuition of the right sort). We must, then, always take care to describe the situation about which we are supposed to have a certain intuition in enough detail to know that the situation is possible (as well as a situation at least somewhat like those of which we have experience). Ridge does not always take care to do that. As a result, he sometimes mistakes a verbal formula for an imaginable situation. Consider, for example, Ridge’s terrorist who “masterminds an attack against the United States from overseas which kills several people [but who] (unlike Osama bin Laden) has never reaped any benefits from the United States and its system of laws nor...from international law” (§III). Ridge thinks it “very [plausible] the terrorist deserves criminal punishment” (§III). But never having had the benefits of those legal systems, the terrorist cannot have taken unfair advantage of them. So, Ridge concludes, deserved criminal punishment cannot be (as I claim) a function of mere unfair advantage the criminal takes by committing the crime.

The “plausibility” of the terrorist deserving punishment, that is, the intuitive appeal of the bald judgment that he deserves punishment, presupposes that we can imagine a terrorist who, though not benefiting from international law, can kill people in the U.S. without coming here. I do not think we can imagine such a terrorist—or, at least, imagine him in a way vouching for the reliability of any intuitions we have about him. The very isolation that would eliminate all benefits of international law—including the money, technology, and movement of conspirators necessary for the terrorist act—would, it seems, eliminate the terrorist’s ability to strike the U.S. using familiar means. Of course, we can imagine a terrorist killing by voodoo or some other unfamiliar means not relying on Western technology and institutions. But, since we have had no experience of terrorists striking us (or anyone else) in that way, we (again) can have no (reliable) intuitions about what such terrorists deserve.

I do not claim that Ridge cannot fix this example so that we have intuitions about it. Though I doubt that he can fix it, all I am now claiming is that, as he has stated the situation, even a reasonably charitable reader cannot have the intuition in question. Unless Ridge can re-imagine the terrorist in a way generating a real intuition (and serving his argument), he has only a verbal formula, not the terrorist he needs.

If Ridge cannot fill out his description in a way making it more than a verbal formula, how is it possible for him to think he has an intuition regarding it? Obviously, the intuition must be of something else, something easily mistaken for the desired intuition. My guess is that Ridge’s intuition is actually of a terrorist we can all imagine, a terrorist who, like bin Laden, did benefit from international law. Any intuition about this easily imagined terrorist would, of course, not serve Ridge. He needs to have us imagine the terrorist he described, the one that cannot be understood as what Ridge prefers to call “a free rider” (that is, someone taking unfair advantage of the criminal law). Ridge therefore needs to state his example in enough detail so that we can be sure his intuition—and ours—is real, not something mistaken for a real one.

(c.) Having achieved that, Ridge would have another problem. We might not share his intuition. Often, intuitions are not as widely shared as the author appealing to them supposes. Ridge thinks it “plausible” (that is, intuitive enough to need no argument) that his imaginary terrorist deserves criminal punishment. I don’t think it plausible. My reason is that I do not have that intuition even about bin Laden. If the U.S. government were to ask my advice about what to do with bin Laden should he be captured, I would, I admit, advise trying him in open court under ordinary civilian law. But I would advise that on principle, not because I feel that he deserves such treatment. Criminal justice—as I understand it—is a system of law in which deserved punishment is the ordained outcome of criminal wrongdoing, a system that treats its subjects as rational agents, moral persons not only capable of guiding their conduct by the criminal law but inclined to be so guided. What I have heard of bin Laden makes me feel that he is more like one of Locke’s polecats, a danger rather than a criminal. Hence, my intuition is that punishment is too good for him. We should just shoot him where we find him or hold him in some secure place until he dies or his fanaticism dissipates. I doubt I am alone in that feeling. The Bush administration’s incarceration of hundreds of “suspected terrorists” without trial seems the natural expression of that feeling—as is the public’s acceptance of what might otherwise seem an outrage against justice, an outrage without obvious benefit in domestic safety or international esteem.

(d.) Even if an intuition is clear and shared, it may not be clear about the relevant question. Consider the passage from Pratchet’s *Guards! Guards!* that Ridge quotes. It describes an arrangement between the Guild of Thieves and the government in “Discworld.” The arrangement is a complicated system of vouchers and receipts designed both to keep crime within limits and to distribute crime more or less evenly across the population. The arrangement sounds a lot like the background for the auction I use to model fairness. Ridge concludes—without argument—that the passage is funny “precisely because [the arrangement] is so absurdly unjust” (§III). I share Ridge’s intuition that the passage is funny, but I do not agree that the absurdity of the injustice explains why the passage is funny. Good satire is seldom so flat-footed. Indeed, I think the joke is precisely the opposite of what Ridge supposes. As far as I can tell, I laughed at Discworld’s system of
crime control for the same reason I used to laugh at the
crude machinery by which one of Rube Goldberg’s cartoons
would propose to do something ordinarily done much more
directly. Yes, the machinery would do what it is supposed to
do—but in what a fantastic way! Unless Ridge has an
independent argument showing Discworld’s system of crime
control to be unjust, he has no right to move from our intuition
that the arrangement is funny to the conclusion that it must be
unjust. Our intuition, however clear, is not clearly about
injustice.²

If we strike from Ridge’s paper all passages relying on one
of these erroneous uses of intuition, not much is left. Cleaning
up his erroneous uses of intuition would, in effect, wipe out
his argument against the fairness theory. I might then stop
here, but I shall go on. Ridge has committed at least four other
serious errors other critics might make as well: (2) a failure
to be clear about criteria of adequacy for a theory of punishment;
(3) a failure to treat my model as a model; (4) a failure to
understand how the model is constructed; and (5) a failure to
distinguish between two fundamentally different senses of
“deterrence.” I shall conclude this response by explaining these
four additional errors.

2. What Should a Theory of Punishment Do?

There seem to be two (primary) views about what punishment
theory should do. One view, the moralist, assumes that
punishment is a widespread social phenomenon of which
criminal justice is a special case. We must shape our theory to
fit such diverse practices as parents punishing their children,
gods punishing humans, humans punishing each other in “the
state of nature,” and even children punishing their pets.

Punishment is as central to morality as moral praise and blame.
A theory of punishment that fit only criminal justice would be
radically incomplete. That is the moralist view, punishment as
part of ordinary morality.

On the other view, the legalist, punishment is primarily a
legal concept. Its home is that relatively complex system of
standing procedures, rules, and penalties we call “the
criminal law.” What a parent does with a child—say, deprive of
some privilege or misbehaving— is a mere analogue, lacking some
or most of the features of punishment strictly so called (rational
agent, rules announced in advance, set penalties for violation,
and so on). One cannot argue against a legalist theory of
punishment (a theory of criminal justice) by appealing to such
marginal cases. Their marginality qualifies them from serving
as counter-examples. For legalists, a theory of punishment
that fit only criminal justice would be quite good. Any failure
to deal with marginal cases, even as analogues or metaphors,
would be no worse than an absence of frosting on an already
rich cake.

These two views might live side by side in relative peace,
much as moral theory and legal theory do, did the moralists
not regularly criticize the legalists for (in effect) not being
moralists. The moralists have some justification for their
criticism. Early work in criminal justice, that is, its foundational
literature, tended to take moralism for granted. Why that was
so is a question for historians. Perhaps Christianity had a part in
this. (Understanding God as maker of a law the violation of
which is to be punished in the torture houses of Purgatory and
Hell makes sin a kind of crime and the divine response a kind
of criminal justice.) Perhaps the analogy between law and
morality, the extreme form of which is natural law, made it
easy to think of morality as a natural legal system (with “reason”
promulgating both statutes and penalties). Perhaps the practice
of most of Europe’s legal systems contributed, since they then
generally accepted natural law as (in some sense) part of
positive law.

Legalism is a recent break with the philosophical literature.
But legalists have at least three justifications for that break.
The first is etymological. All the terms central to punishment
theory (“penalty,” “crime,” “condemnation,” and so on, as well
as “punishment” itself) have their origins in systems of public
justice. Etymology warns us that we are not entitled to assume
that children punishing their pets, parents punishing their
children, or even God punishing his people, is the older or
more central form of punishment. Second, there is context.
Very little writing on punishment (except by theologians) is
about justifying private or divine “punishment.” At least since
Beccaria (1764), the focus of punishment theory has been the
practice of relatively complex legal systems. Third, while
legalism shares with legal positivism an awareness of the
difference between law (in which there are authoritative
decision-makers) and morality (in which there are none),
legalism does not depend on legal positivism for its defense.
All it depends on are the obvious differences between law
and morality.

One obvious difference between law and morality
concerns punishment. Legal systems have relatively complex
procedures for setting penalties in advance of wrongdoing.
Morality does not. Indeed, as far as morality is concerned, ten
days in jail is, all else equal, as appropriate a penalty for theft as
is ten months or ten years.

While I believe that Ridge would, or at least should,
agree with this abstract statement, he often argues as if it were false.
Consider, for example, Ridge’s claim that I “[allow] that we
can make reliable moral judgments about what the maximum
penalties appropriate for forms of wrongdoing other than free-
riding but [deny] that those intuitions are relevant to
determining the maximum criminal penalty” (§III). His only
evidence for this claim is in a footnote 17. “This seems implicit
in his discussion of non-criminal punishment in extra-legal
settings, such as punishment of someone in a concentration
camp by his fellow prisoners.” Ridge cites two pages in my To
Make the Punishment Fit the Crime as the authority for his
claim—without quoting either. The relevant passage on the
first of those pages seems to be this:

We could, I agree, talk of deterrence, revenge,
training, condemnation, and perhaps even education.
Since each of these categories includes a principle of
proportion, we could find a use for “deserved
punishment” even in a concentration camp—just as
we can find a use for “deserved punishment” when
explaining why we kicked a dog. But such uses are
easily distinguished from the use of “deserved
punishment” characteristic of criminal law.¹

As I read this passage, it allows for moral uses of “deserved
punishment” that are analogical or metaphorical rather than
literal. The passage does not imply that my use of “deserved
punishment” is literal (or, at least, any more literal than is
speaking of kicking the dog as punishing it). On the second
page, I add (in part):

even if I agreed that we could find in a concentration
camp a use of “deserved punishment” much like that
characteristic of the criminal law, that would show
only that a concentration camp is still enough like
ordinary life for ordinary moral categories to apply.
Our judgments of desert there would be judgments
of simple moral desert, not of criminal desert.³

Note that I did not actually agree that we could find such a use
of “deserved punishment” in a concentration camp. My
reasoning is counterfactual: Even if we could find such a use,
the use would be the analogical or metaphorical use of
ordinary morality, not the literal use characteristic of the criminal law. True, I do not explicitly say there that ordinary morality lacks a literal use of “deserved punishment.” But the preceding chapter criticizes von Hirsch’s moralism; and other chapters include such statements as the following: “Punishment [as I define it] cannot be conceived apart from the criminal law (or some close analogue).”

To me, my position is clear. “Deserved punishment” in its literal sense has no use in a concentration camp or in any other radically unjust institution. Whatever one prisoner might say to justify harming another in response to a moral wrong is better put in terms other than “deserved punishment” (deterrence, incapacitation, and so on). Morality may limit our response to wrongdoing, even in a concentration camp, but those limits are not enough to justify the claim that the resulting response is literally “deserved punishment.” The problem is as much “punishment” as “desert.”

Alas, whenever a writer fails to make himself clear to a reader, especially to as earnest and intelligent a reader as Ridge, the fault is the writer’s. When I wrote those passages, I did not appreciate how strong a hold moralism would have on many of my readers or how hard it would be for them to understand my view that (strictly speaking) “punishment” is a legal, not a moral, term. So, at the risk of seeming to repeat the obvious, let me repeat the point I think Ridge missed: morality as such does not require punishment—though it may have much to say about deterrence, reform, education, incapacitation, and so on. Only when morality is brought into a legal system, as a side constraint, does speaking literally of “[morally] deserved punishment” become possible. Anyone who understands this will, I think, have no trouble reading as I intended the passages Ridge cites as implicitly showing that I accept a literal sense of “deserved punishment” in “nature” (that is, in morality operating outside the criminal law).

Having read those passages my way, the same reader will have no trouble dismissing a number of Ridge’s criticisms that depend on assuming a natural scale of punishment. For example, it is not true (as Ridge claims) that any disagreement we may have about the outcome of an auction for licenses to commit crimes is no easier to resolve than disagreements about how to determine the relative overall moral wrongness that moralists take to be the first step to determining the morally deserved punishment. Determining the outcome of an auction for licenses of the sort I imagine is (as I shall soon show) a relatively well-defined task, one about which there has, as far as I know, yet to be any disagreement in print. In contrast, the moralist’s task is so ill-defined that more than two centuries of work by philosophers, lawyers, and social scientists has so far failed to produce agreement even on such a simple question as whether attempts doing no harm deserve less punishment than the corresponding complete crime.

3. Using the Model

That Ridge is a moralist may explain other features of his argument. Consider, for example, his “small fishing tribe” with only a “rudimentary system of law” (§IV). Ridge supposes that practice there provides an interesting counterexample to a theory that (as he admits) fits (reasonably well) the actual practice of fully developed (relatively just) legal systems. Since I am legalist, my first response to such an example is to dismiss it as irrelevant. Ridge owes us an explanation (other than moralism) of why a theory of punishment should concern itself with an example so far from the central cases of punishment. Insofar as Ridge’s tribe’s legal system is (as he says) “rudimentary,” it is also marginal to punishment theory. Because it is marginal, we are entitled to dismiss any conclusion about criminal justice Ridge might derive from it.

But I shall not do that here. The fishing tribe embodies another important error of method, a misunderstanding of how to use the auction model. Ridge’s fishing tribe is a thought experiment, not an actual case. It purports to undermine the fairness theory by showing that the market analogy itself yields results independent of unfair advantage. The tribe is supposed to embody a relatively simple situation in which the public goods at stake are all that distinguish one crime from another. The two crimes in question are moral wrongs only because prohibited, not moral wrongs whether prohibited or not. The unfair advantage taken is supposed to be the same. Yet, according to Ridge, the crime involving the more important public good (“staple”) deserves a punishment greater than the crime involving the less important public good (“caviar”). If he is right, we must conclude that what is deserved here, though criminal punishment, is independent of the unfair advantage the two crimes take. The unfair advantage theory must be incomplete.

If this argument were an appeal to intuition, it would commit an error already discussed. The appeal would be to our intuitions, not to those of the fishing tribe. Since our intuitions about what should be punished more seem an unlikely indicator of what should be punished more in a primitive fishing tribe, I think charity requires us to interpret Ridge’s claim that over-fishing staple is morally worse than over-fishing caviar as having some basis other than our intuition. Unfortunately, he does not say what that other basis is—and I have no guess. So, for purposes of argument, let us assume we somehow know that over-fishing staple is morally worse than over-fishing caviar. What follows about deserved punishment?

Nothing—unless my theory, or rather the auction model embedded in it, yields a result inconsistent with the comparative moral judgment we just assumed to be true. Does it? Ridge claims it does. But his only argument for that claim is a single sentence: “It may well be that a license to catch extra caviar would fetch much more on an auction than a license to catch extra staple—caviar is very, very tasty and 5 fish is easily enough to keep one from going hungry (change the number if 5 doesn’t sound compatible with this, but I am assuming that stable are big fish!)” (§IV. “It may well be.” The argument seems to have this form: since our ignorance (or our under-description of the case) does not rule out X’s being true, X is true. The argument seems to be a non sequitur.

Underlying the non sequitur seems to be a misunderstanding of how to use the auction as a model. In the unfair advantage theory, the auction is not an appeal to intuition. It is, instead, a way to transfer methods from economics, methods allowing us to deduce a (partial) ordering of crimes from greatest to least. But the auction will allow that only if we construct the model so that economic theory applies to the questions we put. There must, for example, be a common medium of exchange. Without that, we cannot compare one sale with the next. Ridge’s fishing tribe, however primitive in other respects, will have to have money; whether denominated in dollars, shells, or the like. Barter will not do; barter generates only a series of independent trades. The auction will also have to have enough licenses and enough bidders to avoid collusion, to cancel the effects of individual idiosyncrasies, and otherwise to make the tribe’s market behave (more or less) like a free, fair, and efficient market (that is, a perfect market). Anything like that is impossible in a small tribe. Ridge’s fishing tribe is going to have to be much larger than he supposes and, given the sophistication running a large auction presupposes, much more sophisticated as well.

Ridge will have to make a number of choices as he constructs this large tribe, choices that might beg the question with which he began. So, for example, he must decide whether
the tribe will trade with others for staple. If the tribe can trade
caviar to other tribes and receive staple in return, the caviar
will, in effect, be equivalent to staple for purposes of feeding
the tribe. There will be one public good (“fish”) rather than
two (“caviar” and “staple”). Even masculinity will have to treat
them as (qualitatively) the same. To get the result he wants,
Ridge must imagine his tribe isolated. But, even assuming his
large tribe to be isolated will not guarantee Ridge the result he
is looking for. He will have to do much more staple there is.
If staple is abundant enough to replenish supplies however
much the tribe might fish, the tribe could license any amount
of staple fishing. Licenses being many, the price of each would
(as he claimed) be low, low enough perhaps for the license
for caviar to be higher. If, however, he assumes that caviar
is abundant but staple is not, the price of a license to fish caviar
might be higher than the price of a license to fish staple. Given
the importance of supply to price, Ridge will have a hard time
choosing a way to decide the number of licenses without
begging his question about the relative price of caviar and
staple.

I could go on. But I think you get the idea. Ridge has so
under-described his fishing tribe that nothing follows from it.
Some more fully described version of it might yield the result
that he claims, but some will not. He must carefully construct
his model to find out what is possible, justifying each
assumptions on which that construction relies. Among
considerations relevant to that justification is that the
assumption does not beg the question Ridge set out to answer.
Once Ridge has constructed the market, he has only to refer
to economic theory to deduce the (relative) price of licenses.
If the prices came out as he claims (and we continue to assume
that over-fishing staple should be punished more severely than
over-fishing caviar), he could restate his objection.

4. Between the Horns of Ridge’s Tri-Lemma

Such considerations led me to decide “poaching” would be
possible in my model but should be dealt with in a way separate
from the (ordinary) system of licenses. By “poaching,” I mean
violating my imagined society’s (second-order) rule, “No
person shall commit a [first-order] crime unless he has a license
to do so.” Ridge argues that I made the right choice when I
decided to handle poaching in this way because I faced a “tri-
lemma” the other horns of which were worse, but that I paid
a high price for my choice: I imagined a society that is radically
unjust ($§II$).

The tri-lemma Ridge supposes me to have faced is: (a)
licensing “poaching” in the way other crimes are licensed; (b)
punishing poaching outside the ordinary legal system
proportionally; and (c) punishing poaching outside the system
draconically. I believe I had at least one more option: I could
have assumed that poaching was impossible. That assumption
would both preserve the applicability of the market theory I
need and avoid any injustice that arises from punishing
poaching draconically. I chose not to make that assumption
for at least two reasons.

First, I did not want to make an assumption about poaching
that might invite irrelevant questions. Saying poaching is
impossible would have done that. If poaching is impossible in
the imagined society, why are not all crimes? What is so
different about poaching that people would not engage in it
when they may do so without punishment? And so on.

Second, I prefer to keep my model as close to the familiar
as possible. Punishing poaching draconically outside the system
of licensing does that. I may thereby have introduced some
injustice into the system. But, since the system is constructed
so that poaching is likely to be rare (the effect of draconian
punishment combined with relatively cheap licenses), I cannot
see why Ridge thinks the resulting society would be radically
unjust—or even why it would overall be less just than his
society or mine. For the ancients, Draco’s Athens exemplified
not injustice but severe justice. Here again Ridge owes us an
argument.

Of course, Ridge is right that my results would be the
same whether the imagined society is relatively just or radically
unjust ($§II$). Economic theory does not presuppose just
institutions (though it does presuppose institutions stable
enough and fair enough to make market exchange rational).
He is also right that my argument would be stronger insofar
as it did not rely on a sub-argument attempting to show that the
society imagined for the auction is relatively just. I have
nonetheless preferred to argue that the society is relatively
just for at least three reasons. First, I believe the society in
question is relatively just (as I just explained). Second, we are
much more familiar with relatively just societies than with
radically just ones. Thinking of the society as relatively just
allows us to use our intuitions in constructing the model.
Catching errors in reasoning is therefore easier. Third, I do not
want readers to reject my model because it is “unrealistic”
(though it is). The fewer hard-to-swallow assumptions I
introduce, the easier it is for theorists to at least consider using
the model to answer questions about how much to punish.

The assumption that an adequate theory of just punishment
should track conclusions drawn from a radically unjust society
is one of those hard-to-swallow assumptions I would rather
not ask readers to accept.

Neither of these reasons separately, nor even the two
together, is decisive. Construction, like design, is an art in which
even equally good solutions may differ much. I worry only
when equally good solutions yield importantly different results.
Like previous critics, Ridge has yet to show that there are
equally good models of unfair advantage yielding importantly
different results. That there are equally good models yielding
the same results is, of course, evidence for the robustness of
the results, as is Ridge’s acknowledgement that the
conclusions I draw from the auction I imagine are independent
of the justness of the society imagined.

5. Deterrence

Ridge concludes his paper with a problem. On the one hand,
he has (he claims) destroyed the foundation of the fairness
theory. On the other hand, he admits, “[The] theory does
seem to do an impressive job of predicting actual criminal
penalties” (Conclusion). While predicting actual criminal
penalties is a reason to accept the theory only insofar as actual
penalties are just (or, at least, justified), the actual penalties
the fairness theory predicts are, on the whole, just (or, at least,
not obviously unjust). We are then led to ask how a theory so
wrong in its foundations could be so right in its results.

Ridge’s answer is that “the theory tracks considerations of deterrence
rather than considerations of proportionality” (Conclusion).
This answer, though showing more insight into the fairness
theory than most critics show, makes the last error I shall
consider here. The answer assumes that deterrence and
proportion in punishment are not only unrelated but also
consistent. I have relied on “deterrence rather than proportionality.” The assumption, common in
discussions of punishment, is doubly mistaken.

The first mistake is obvious. For utilitarians at least,
deterrence and proportionality are so far from being
inconsistent that deterrence is the determinant of proportion
in punishment. Punishment should be scaled to give
the maximum of deterrence for the least cost in human suffering.
For utilitarians, the contrast is not between deterrence and
proportionality but between deterrence and moral desert. For
utilitarians, moral desert is irrelevant to deterrence or relevant only in some minor way (as in assessment of character when sentencing).

Ridge’s second mistake is less obvious. He assumes that, for retributivists at least, deterrence and proportionality must be unrelated. To let considerations of deterrence affect proportion is (he seems to say) to give up deserved punishment in favor of a utilitarian (or other consequentialist) theory. In fact, that is not so. We must distinguish two sorts of deterrence: (what we might call) empirical and conceptual.

Actual or probable deterrence is an empirically contingent (“external”) relation between announced penalty and the actual conduct of potential criminals. The penalty is justified in prospect if, and only if, it probably deters enough (and all else seems equal). The penalty is justified in retrospect if, and only if, it does indeed deter enough (and all else is equal). The only way to know whether the penalty does deter enough is to enact it and see what happens. There is nothing in the concept of empirical deterrence to guarantee that a penalty will have any deterrent effect whatever.

That is one way to understand deterrence, but not the only way. We can also understand deterrence as a deduction from certain assumptions (an “internal” rather than empirical relation between penalty and conduct). When we ask about the justification of the criminal law, we ask about the justification of a system of rules that treats those subject to it as if they were rational agents able (and willing) to choose (in part) on the basis of the penalties announced in advance. Insofar as they are rational, potential criminals will take the penalties of the criminal law into account. They will treat the system of penalties just as they would any other system of incentives or disincentives. All else equal, they will always prefer the crime with the lower penalty to one with a higher penalty and prefer to avoid many acts because the penalty for committing them is too high. This is not an empirical relation between penalty and conduct, any more than are claims about what rational agents will do in a perfect market. But the relation is one of deterrence—and may explain why so many people can declare their belief in the deterrent effect of stiffer penalties while acknowledging the absence of empirical evidence for that deterrent effect.

I therefore can (and do) agree with Ridge that the fairness theory works because of its connection with deterrence—without giving up my retributivism. I need deny only that the deterrence in question is empirical. If I am right that it is not, then the theory’s success at predicting practice, especially the practice of systems of criminal justice we regard as relatively just, is strong evidence for the theory. Of course, I should have something to say about why the theory is so good at predicting the penalties of actual systems of criminal law when everyone knows that criminals are not, on the whole, particularly rational—and legislators, not particularly given to accurate application of philosophical theories. But, in this, I am no worse off than economists who must explain why their theories of rational behavior fit consumers and firms that, though perhaps more rational than most criminals, are still far from perfectly rational.

Endnotes

1. All references in the text are to Ridge’s essay in this issue, by section.


3. For a philosophically more damning example of this error, see Michael Davis, “Nozick’s Argument For the Legitimacy of the Welfare State,” Ethics, 97 (April 1987): 576-394.


5. Ibid., 238.

6. Ibid., 71.

7. I have made this point before. See, for example, ibid., 73-77; or my Justice in the Shadow of Death (Lanham: Rowman and Littlefield, 1996), 10-21.

RECENT ARTICLES OF INTEREST

—Abstracts—


Finkelstein maintains that it may be plausible to consider being exposed to a risk of harm (which she calls “risk harm”) as itself constituting a harm (understood as a setback to a legitimate interest), even if the harm to which one has been exposed (the “outcome harm”) never occurs. Just as being given a chance of getting a benefit is a good for a person even if the benefit does not ultimately occur, so likewise one is worse off for having been exposed to a risk of harm, even if it does not materialize. The sense in which one is made worse off by such exposure is not merely psychological; it would be incorrect to say that, if the harm never occurs, there has been no loss other than a psychological setback. Rather, it is an “objective setback” to one’s interests. If a person is exposed to a risk of developing cancer, then, even if the cancer never develops, he is worse off, he is doing less well in life, in belonging to a group having a higher probability of developing cancer.

There is support in the law of torts and crimes for the notion of risk harm, particularly in cases in which, without this notion, it is difficult to understand why compensation or punishment are prescribed. For example, there is a case that, though denying recovery for increased risk of developing cancer, nevertheless allowed recovery for the costs of medical testing to identify early onset of the disease. This, Finkelstein says, would make no sense unless plaintiff were thought to have been harmed even were no cancer to develop. Likewise, Finkelstein argues, so-called market-share liability is best understood as compensating those who have been exposed to the risk of harm, and the idea may also be at work in some toxic tort cases, where suits are permitted at the time of exposure and before any outcome harm materializes. The criminal law also responds to risk harm—for example, in the crime of reckless endangerment. Risk harm may also help explain the law’s treatment of attempted crime, for if “exposing someone to a risk is to harm him, and if . . . criminal liability [is] primarily concerned with harm infliction, then punishing attempts can be explained by the Risk Harm Thesis . . .”

Finkelstein is cautious in her conclusions. She acknowledges that the notion of risk harm may not ultimately prove defensible and that there are other possible explanations.
of the legal situations she has identified as illuminating the role of risk harm in the law. Furthermore, she concedes, even if risk harm is a defensible notion, it may be that “there are reasons not to import that insight into the law.”


This article discusses a small but important issue within legal theory: the role of theories of meaning and reference within jurisprudence, and whether having the proper theory of meaning and reference can help resolve persistent issues regarding legal indeterminacy. The article focuses on recent work of Michael Moore, while also discussing the writings of Nicos Stavropoulos and David Brink, all of whom have argued for a change in our understanding of law and legal interpretation based on philosophical theories about language and meaning. Their works urge that terms, even legal terms, must be understood according to their true nature, as determined by the world, rather than by the intentions of lawmakers. The article presses a position that requires a partial dissent from these views: that because of the nature of law (or the nature of law in a democracy), legal interpretation must at least sometimes focus on the lawmakers’ intentions, even (or especially) when those intentions diverge from the usual meaning of terms.


In 1984, Congress passed the Sentencing Reform Act, in which it listed acceptable purposes of punishment and charged the U.S. Sentencing Commission with determining, based on these purposes, lower and upper bounds of permissible criminal sentences in the federal courts. Congress’s purpose was to rationalize sentencing practices. This is a difficult task, and it is unclear whether the Commission has succeeded, or even can succeed, in devising sentencing rules that reflect a consistent and coherent philosophy of punishment.

There are four sorts of considerations the Commission is directed to take into account in establishing sentencing ranges. Two of them are central in establishing the initial sentencing range. These are the defendant’s prior criminal record and the seriousness of the offense for which he or she has now been found guilty. There is a schedule of scores, or points, for numbers of past convictions and severity of offense, and the total sets the initial sentencing range. Considerations of “substantial assistance”—that is, discounts for defendants who help law enforcement agencies—and family circumstances—the extent to which punishing a defendant would harm third parties, such as defendant’s family—enter at this point to reduce the range of permissible sentences. Each of these is frequently used by judges in their sentencing decisions.

Rappaport argues that no single theory of punishment adequately explains all of these elements the federal sentencing guidelines, though utilitarian-deterrence theory seems to be at work with regard to most of them. A hybrid theory (utilitarianism-cum-just-desert theory) cannot succeed in providing a fully principled basis for the sentencing guidelines. This does not mean, Rappaport says, that the Commission must adopt a thoroughgoing utilitarian rationale, but it does mean that without this it will be left with some degree of arbitrariness.


Philosophical aesthetics rarely has been brought to bear on an understanding of the law, but interest has grown in recent years, perhaps influenced by the flourishing law and literature movement. Aesthetic issues also have been recognized in obscenity, copyright, historic preservation, and other zoning issues.

Philosopher Brian Butler first surveys various ways in which aesthetic considerations have been taken into account in the law, including the impact of popular culture on courtroom expectations, the critique of legal reasoning from the perspective of literary narrative forms, and the use of metaphor in Ronald Dworkin’s “chain novel.” He then critiques a series of philosophical proposals by legal scholar Pierre Schlag. Of special concern to Butler is Schlag’s apparent assumption that the aesthetic is non-rational and can only be grasped through “imaginative empathy.” Butler also rejects a hierarchical assumption in the law that pure reason is paramount and any appeal to aesthetics or “beauty” is of lesser importance. Butler argues that legal reasoning is not hierarchical, but rather consists of “multiple, overlapping, intersecting types of reasoning.” Aesthetics, he insists, does not render our cognitive and rational faculties “irrelevant.” Instead, we should recognize that differences “between argument form and aesthetic form” are matters of degree, not of type. In doing so, he at least keeps open the door to a more rational, less dismissive consideration of aesthetics in the law.


“Is it possible to say, on first principles, what is wrong with ‘something for nothing’?” The recent transition from AFDC to TANF highlights an old distinction between the “deserving” and the “undeserving” poor. TANF, with its limitations on public assistance for those (including single mothers) who are able to work, reflects public antipathy to providing guaranteed incomes for those who can but do not work. The question Wax addresses in this article is whether such a view is compatible with liberal political theories and their conceptions of a just society, the basic principles of which would seem to endorse unconditional basic incomes. Despite this, however, Wax finds that liberal theorists are hardly forthright in support of unconditional basic incomes as a matter of principle, and frequently “smuggle in” assumptions that seem to support aspects of desert that they officially reject.

Wax examines some prominent theories of social justice to determine whether they support unconditional basic incomes or instead support work requirements. So-called “luck egalitarianism” holds that if a person’s position in society, including his or her ability to hold a job, are beyond his or her control, social support is appropriate, but not otherwise. This apparently desert-based idea has difficulty, though, dealing
with the apparent inability of the economy to create work for everyone. Furthermore, in the market there is no obvious correlation between desert and reward (some get much less than they deserve, and others much more), making it difficult to object to getting something for nothing on the ground that it is not deserved. Another critique of work requirements comes from feminists who “disparage the logic of providing compensation only for labor market work,” believing instead that much work is performed outside of the market and that many people outside of the market deserve society’s support. It is simply wrong to think that all work is done in the market and that all activities outside the market constitute leisure.

Wax next reviews a number of contractarian theories of justice. Both John Rawls and Ronald Dworkin, she maintains, are equivocal on the issue of unconditional basic income. Though Rawls disagrees with using desert as a basis for resource distribution, it has been argued that he has been unclear about who qualifies as being among the least advantaged who must be benefited by the difference principle, and in later writings he includes leisure as a primary good to be distributed by a society, suggesting that this can replace income transfers for some people. Dworkin too, Wax believes, is equivocal on the issue. The design of the initial equal allocation of resources, followed by an auction, that he constructs, and the unclarity of the distinction between “option luck,” for which one is responsible, and “brute luck,” for which one is not, make it doubtful whether Dworkin has anything to say as a matter of principle about getting something for nothing. Another theorist, Philippe van Parijs, does defend guaranteed basic incomes, though his theory too, Wax maintains, rests on question-begging premises.

These political theories, Wax concludes, are at odds with most people’s political psychology, in particular with the widespread belief that those getting something for nothing are freeloaders who do not do their share, or even make an effort to do so, and who thus exploit productive members of society. Such sentiments are likely to have an evolutionary explanation, having “emerged as a feature of human psychology or culture from a competitive process that pitted individuals or groups with different behavioral strategies against one another in round after round of competition.” They require us to decide whether reason or natural moral sentiments should be primary.
Civility and Its Discontents
Civic Virtue, Toleration, and Cultural Fragmentation
Edited by Christine T. Sistare

“An extremely impressive volume integrated around issues at the core of many, if not most, debates in contemporary Anglo-American political and legal philosophy. It will undoubtedly be useful to scholars, teachers, and students working in philosophy, law, and political science.”—David Ingram, author of Group Rights: Reconciling Equality and Difference

“The discussions taken up by the contributors to this volume—including such distinguished voices as Joan McGregor, Patricia Smith, and Wade Robinson—reach to the nature of civil society, the character of democracy, the relationship between the demands of individual liberty and those of the common good, and the role of law and government policy in weaving the fabric of a social whole.”—from the Preface

An AMINTAPHIL volume. 320 pages, Cloth $39.95, Paper $19.95