Agency’s Constitutive Normativity: An Elucidation

1 Introduction

My aim in this paper is to provide a conceptual elucidation of the notion of constitutive norms/normativity, which is central to Constitutivism as a first-order theory of agency.

After introducing and clarifying the origins and scope of Constitutivism (Section 2) and Christine M. Korsgaard’s version thereof (Section 3), I present three possible interpretations (which are not meant to be exhaustive) of Korsgaard’s characterization of the notion of constitutive norms (Section 5). I endorse one of them—the Practical Interpretation—as the best way to make sense of the notion while accounting for some of the conclusions drawn from the discussion of the problem of defective actions (Section 4). For Korsgaard, the constitutive norms governing agency amount to the norms governing practical reasoning and morality. My contention here is that the constitutive normativity pertaining to human agency is not ‘homogeneous’ in kind. In other words, the point of this paper is to show that there are more norms governing agency than just the norms for practical reasoning—and, in particular, that agency necessarily involves social norms distinct from any notion of practical reasoning.

After a brief methodological interlude (Section 6), I draw on a theoretical framework on the metaphysics of normativity developed within the philosophy of biology to distinguish two varieties of constitutive normativity (socially-generated and non-practice-based) (Section 7), and end by providing a preliminary articulation of the socially-generated constitutive normativity inspired by the speech-act-theoretical account of the performative dimension of language (Section 8).

2 Constitutivism

Constitutivism is a philosophical position according to which certain normative claims of practical reason and morality—that is, about which actions an agent has reason (or ought) to pursue/perform—can be derived from the features that are essential to being a rational agent: 1 a creature capable of acting on the basis of reasons. Here are a few definitions:

A constitutive theory is a metanormative theory that purports to establish the objective validity and content of practical norms on the basis of the constitutive features of agency.2

Constitutive arguments attempt to establish the normativity of rational requirements by pointing out that we are already committed to them insofar as we are believers or agents […] to justify requirements anyone would have reason to accept [normative requirements] by explaining what it is to engage in a certain activity or be a certain kind of being.3

1 Unless otherwise specified, my use of the term ‘agent’ stands for ‘rational agent,’ as opposed, for example, to non-human, animal agents. In other words, the discussion ranges within the scope of the standard conception of agency, i.e., intentional and rational agency.


The insight, in a nutshell, is to ground content and authority of non-hypothetical practical standards in features viewed as distinctive to either agency or action.4

The above definitions are an expression of the prevailing line of interpretation of the constitutivist project informing current debates on constitutivism, which emphasizes the distinctively metanormative character of the project. In other words, constitutivism aims to show that the normativity of requirements of practical reason—namely their objectivity (non-contingency and universality), and practicality (action-guiding character and authority)—is grounded in principles proper to intentional human agency. The constitutivist project, however, originates from somewhat separate debates and theoretical concerns.

2.1 Constitutivism and Metaethical Constructivism

On the one hand, within moral and metaethical debates, constitutivism should be interpreted as a metanormative project akin to certain versions of constructivism.5 Metaethical constructivism represents a distinctive alternative to realist and anti-realist approaches to questions about the existence and nature of normative truths, and can be said to be advancing both a metaphysical and a normative thesis. The metaphysical thesis consists in a positive reply to the question about the existence of normative truths, and holds that “these truths are not fixed by facts that are independent of the practical standpoint, […] rather, they are constituted by what agents would agree to under specified conditions of choice.” As a first-order (moral) normative theory, constructivism holds that the normative principles we ought to accept/follow “are the ones that agents would endorse were they to engage in a hypothetical or idealized process of rational deliberation” 6—call this the normative thesis. In other words, the human capacities for self-reflection, evaluation, practical reasoning, are the legitimate and authoritative processes to appeal to in our quest for normative validation and guidance.

The distinctiveness of constructivism qua metaethical position lies therefore in a vindication of the objectivity of normative judgments via the appeal to the self-authenticating character of the evaluative procedure through which these judgements are arrived at, or constructed. The relation between constitutivism and certain varieties of constructivism consists therefore in the appeal, on the part of the latter, to constitutive norms of reasoning/evaluation/agency.7

Emphasizing the tight connection between constitutivism and constructivism, however, should not lead to underestimate the fact that they do not necessarily overlap. First of all, constructivism extends beyond the metaethical debate and as a matter of fact it originally came out as a position in political philosophy. Second, there are various accounts of constructivism

6 Bagnoli, op. cit., preamble.
7 Ivi., Section 7.3.
and constitutivism, and—as it has been noticed—certain versions of the latter might not be compatible with some versions of the former. In particular, the constitutivist strategy as we can find it articulated in recent debates is not part of early constructivist theories, notably the one exemplified by Rawls’ *A Theory of Justice*, which is a form of political constructivism, more modest than ethical constructivism.  

2.2 Constitutivist Strategies and Action Theory

On the other hand, constitutivism arguably represents one among many recent philosophical approaches which take action theory to have an explanatory priority over theories of practical reasoning. In particular, such approaches focus on establishing what it takes for something to be an action, and in delineating a distinction between full-fledged action and “mere activity” or “mere behavior.” With these distinctions in place, they make a case for the claim that in order for something to be an action at all, reasons for action must have a certain logical form. As a project in philosophy of action distinct from strictly metaethical and metanormative concerns, we can thus say that constitutivism has, first and foremost, a distinctive metaphysical ambition in that it provides an account of what intentional human agency is, and what its essential, or constitutive features are.

The crucial maneuver in the overall constitutivist strategy, however, is made at the very level of the first-order account of what, metaphysically, human agency is. It is here that the essentially normative character of constitutivist strategies comes into play—that is, with the claim that it is the nature of human agency itself that determines the standards for good or successful agency. In light of the distinctive metaphysical position advanced, I will therefore refer to, and focus on, constitutivism as a kind of metaphysical-cum-normative strategy in action theory.

To be sure, there are differences in the scope and strength of the conclusions of extant variants of constitutivism. Christine M. Korsgaard’s Kantian account, for example, identifies the constitutive standards of agency with moral principles—the hypothetical and categorical

---

8 Sorin Baiasu, “Constitutivism and Transcendental Practical Philosophy: How to Pull the Rabbit Out of the Hat,” *Philosophia*, doi.10.1007/s1106-016-9746-3 (2016), n. 1. While this is the case for *A Theory of Justice*, it is less straightforward with respect to the Dewey Lectures.

9 Bagnoli, op. cit., Section 1.


imperatives—qualifying it as a strong form of constitutivism. David Velleman, on the other hand, maintains that moral norms are supported by, but not categorically rationally required by, the constitutive aim of action, situating it as a more flexible account. Both, however, define agency in terms of a functional kind by claiming that action has a constitutive function, or aim, and both track down the nature and sources of ethical normativity by looking to the constitutive function of action. The very idea of a constitutive aim or function of action, in fact, carries normative consequences, since it sets standards of success with respect to the fulfillment of the aim/function itself.

The capacity to act for reasons at the core of the standard conception of agency is crucial for all the varieties of constitutivism on offer. For example, Korsgaard and Velleman share a similar approach to understanding where agency lies—that is, in the exercise of our distinctive agential capacities—even though they develop independent views of what these distinctive capacities are (the capacity for self-constitution, according to Korsgaard, and for self-understanding, to Velleman).

Despite their differences, all varieties of the constitutivist strategy can be said to display the following structural features: They provide (i) an account/theory of human action—i.e., propounding certain features as the constitutive aim/standards/principles of action—which, in conjunction with (ii) the claim about the inescapability of acting, is aimed at (iii) vindicating the normative authority of normative claims of practical reason from the theory of action/agency endorsed.

Unlike most of the extant literature on constitutivism, in the present work I will not engage in an evaluation of the prospects of constitutivism as a metanormative strategy (step (iii) of the argument). Rather, my focus will be on how constitutive norms work in the elaboration of a theory of agency and action, and, most importantly, on clarifying the status of the practical normativity pertaining to agency, as opposed to moral normativity. This focus will be discussed using Christine Korsgaard’s theory, according to which the normativity of rational agency and of morality are one and the same, and falling short of morality implies falling short of agency altogether, and vice versa.

3 Korsgaard’s Constitutivism

The scope of my analysis in the present work is restricted to Korsgaard’s theory of agency, which is arguably the most prominent and developed, as well as the most contentious, version of constitutivism on offer. Korsgaard’s constitutivist argument can be reconstructed as follows:

(1) Human beings must act
(2) In order to act, human beings must unify themselves into agents


14 Cf. Paul Katsafanas, “Constitutivism about Practical Reason,” in D. Star (ed.), The Oxford Handbook of Reasons and Normativity (Oxford: Oxford University Press, forthcoming), and the special issue on constitutivism and Kantian constitutivism of Philosophia (2016) for a similar characterization of the argumentative strategy common to the different varieties of constitutivism on offer. According to this characterization, constitutivist arguments might endorse a traditional, causal, view of action (e.g., Velleman) or not (e.g., Korsgaard). As stated in the previous section, the focus of the present work is on one example of the latter kind of theory.

(3) Human beings can only be unified into agents when they cohere with the principles of practical rationality

(C) So, human beings as unified agents must act in accordance with the principles of practical rationality

The premises of the argument relate to the “inescapability” condition on the one hand (premise 1), and to Korsgaard’s account of what, metaphysically, intentional human agency is, and what its essential, or constitutive features are, on the other (premise 2 and 3). As mentioned in the previous section, the core of the constitutivist strategy lies at the level of first-order *metaphysical-cum-normative* account of agency. This, accordingly, will be the level of concern of my analysis throughout this dissertation.

Among the reasons for focusing on the first-order account rather than on the metanormative strategy is the fact that, as it has been pointed out, a preliminary clarification of the first-order characterization of agency will prove to be crucial to the viability of the metanormative ambitions of the constitutivist project itself.

The main reason for addressing Korsgaard’s account is that it provides an explicit articulation of the notion of constitutive norms. Constitutive norms are not just principles regulating a pre-existing activity—they constitute, or define, while at the same time providing guidance for, the very activity one is engaged in by conforming to them. This is Korsgaard’s claim about the double nature of constitutive standards:

They are normative, because in performing the activities of which they are the principles, we are guided by them, and yet we can fail to conform to them. But they are also descriptive, because they describe the activities we perform when we are guided by them.17

Korsgaard’s account of agency has been accused of conflating the concept of action with that of a good action, and of not being compatible with the intuitive claim that in order for something to be subject to a norm, it should be possible for it to violate the norm.18 If these criticisms are correct, then on Korsgaard’s account any action must be a good action, and any defective action entails a denial of agency. This raises a number of issues for the messy, actual world within which we ordinarily act, and is especially problematic because the double nature of constitutive norms seems to involve a commitment to provide an account which is both normative *and* descriptive. In the real world, instances of ‘defective’ and ‘bad’ actions and agents abound, thereby making Korsgaard’s construal of the double nature thesis problematic in the face of the “error constraint” to which agency intuitively must be subject.

4 Constitutive Norms and the Problem of Defective Actions

Constitutive norms, Korsgaard claims, are descriptive and normative at once. This type of double-edged sword is rare in philosophical theory, where the normative or the ideal is traditionally separated from the descriptive or the real. Most often, philosophers postulate an idea as a recognized and separated standard to aim at to improve practice. Accordingly, recognizing the mix of the normative and descriptive in agency already sets Korsgaard’s account off on a unique foot.

---


There seems, however, to be a problem with the thesis about the double nature of constitutive principles of agency. As Korsgaard says, since constitutive principles are normative, we can fail to conform to them,19 thus suggesting that they are ‘norms for action.’ But given that the principles are also descriptive, and a violation of a description is impossible, it would make one who fails to conform to them inherently defective as an agent. Since Korsgaard accepts the possibility of bad action, then, according to her characterization of constitutive norms, a violation would sanction the defective status of the action. In turn, a defective action would fail to constitute its author as an agent, up until the exclusion from the kind-membership individuated by the norms themselves. The descriptive character of constitutive norms seems to make them ‘norms for being.’ Agents, according to Korsgaard, are evaluative kinds: “the agent, just as such, is the kind of thing that can succeed or fail.”20 By failing to conform to the constitutive principles of the activity of self-constitution, one performs defective actions that make her less of a fully unified agent, and thus less of an agent—“a mere heap”:21

Since the function of action is self-constitution, […] bad actions, defective actions, are ones that fail to constitute their agents as the unified authors of their actions.22

To put the problem in a highly schematic form, Korsgaard’s characterization seems to present us with the following situation: constitutive norms of agency/action have a double nature—they are both descriptive and normative with regard to the activity of self-constitution. They describe what an action is (they are norms for being), and at the same time sanction as defective instances of actions that do not conform to them (they provide norms for action). In addition, given that the relation between an agent and her actions is not one of production, but rather of authoring and self-constitution, defectiveness of actions transmits to the agent, making her less of a fully unified, and therefore less of an agent. In Korsgaard’s words:

[…] if your action is unsuccessful and you do not bring about the state of affairs that you intended, it is not (or not just) the action that is ineffective. It is you that is ineffective. It is not as if you were effective in producing the action, but then the action, once out there on its own, failed, like a defective machine you have invented and then let loose on the world. The action is not your product: it was you that failed. An unsuccessful action renders you ineffective. Therefore a successful action is one that renders you effective. A similar point holds for autonomy […] Therefore a successful action, an action that is good as an action, is one that renders its agent both efficacious and autonomous.

It is the essential nature of action that it has a certain metaphysical property […] But in order to have that metaphysical property it must have a certain

19 The ‘error constraint’ is usually thought to be a requirement for a norm to be prescriptive, or action guiding; see Lavin op. cit. for a critical discussion of this assumption.
21 Korsgaard, Self-Constiution, Section 10.2.3; cf. also Section 8.2.4.
22 Ivi, Section 2.1.6.
normative property […] This explains why the action must meet the normative standard: *it just isn’t action* if it doesn’t.\(^{23}\)

According to Korsgaard, consistency, self-reflection, and self-unification are necessary conditions for unified agency, which are also requirements for intentional action.\(^{24}\) Autonomy and efficacy are the “essential characteristics of an agent,” and actions that fail to constitute the agent as autonomous and efficacious fail to constitute the agent altogether:

If you fail to follow the Kantian imperatives you will not be efficacious and autonomous, and then you will not be an agent. Your action constitutes you as an agent by being chosen in a way that renders you, the agent, efficacious and autonomous.\(^{25}\)

For the moment, this rough outline of the problem of defective actions should suffice, as it will become clearer through the discussion in the next sections. In the next section I am going to discuss cases of ‘disorders of agency,’ which are meant to show that while the agent’s actions might ‘misfire’ or fail to display Korsgaard’s required pattern of consistency, self-reflection and self-unification (and often, morality), her agency does not fade nor disrupt. Moreover, by relying on evidence emerging from therapeutic treatments of disorders of agency, I argue, *contra* Korsgaard, in favor of a clear distinction of the notion of *agential responsibility* from that of moral responsibility, claiming that only the former is necessary for agency.

### 4.1 Disorders of Agency

In mental health-care practice there is a category of conditions referred to as “personality disorders” (PDs).\(^{26}\) The term ‘personality’ stands for a set of traits that incline one to behave/act, think, and feel in stable ways with respect to certain circumstances, which make her ‘the kind of person she is.’ PDs are defined by the *Diagnostic and Statistical Manual of Mental Disorders* as:

A. An enduring pattern of experience and behavior that deviates markedly from the expectations of an individual’s culture. This pattern is manifested in two (or more) of the following areas:
   1. Cognition (i.e., ways of perceiving and interpreting self, other people, and events);
   2. Affectivity (i.e., the range, intensity, lability, and appropriateness of emotional response);
   3. Interpersonal functioning; and
   4. Impulse control.

B. The enduring pattern is inflexible and pervasive across a broad range of personal and social situations.

---

\(^{23}\) *Ivi*, Sections 5.1.4 and 8.1.3.

\(^{24}\) Cf. Arruda, *op. cit.*

\(^{25}\) Korsgaard, *Self-Constitution*, Section 5.1.3 ff.

\(^{26}\) Although I am keenly aware of the complex conceptual and methodological issues concerning the nature and classification of so-called ‘mental illnesses’ I will not engage in any of the relevant philosophical discussions here. PDs, however, are not categorized as ‘illnesses’ nor ‘diseases,’ but rather as a ‘disorders.’
C. The enduring pattern leads to clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The pattern is stable and of long duration, and its onset can be traced back at least to adolescence or early adulthood.
E. The enduring pattern is not better accounted for as a manifestation or consequence of another mental disorder.
F. The enduring pattern is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., head trauma).27

There are several features of PDs that make them a relevant case for a philosophical account of agency and for the present discussion, in particular. First, PDs generally, have what Hanna Pickard calls a “Janus-faced nature”—i.e., two distinct but closely connected characteristics. Beside causing pain and distress to the individual, the problematic traits and patterns of behavior involved in PDs also often involve impairment of areas of interpersonal functioning, as well as harm to others. For this reason, PDs are considered “a psychiatric condition that essentially involves a moral component”,28 and whose management poses unique scientific, philosophical, legal, clinical, and practical challenges.

Second, Pickard points to the fact that PDs lie on a continuum with ‘normal’ human personality—normal traits are “taken to the extreme in PD[s]”.29 The relationship of ‘continuity’ between PDs and normal traits seems to emerge from the close similarities with our ordinary experience and behavior evoked by the first-personal narratives of what it is like to live with PD. It is important to stress, however, that the what counts as ‘normal’ is always articulated in relation to the expectations of a given culture (cf. DSM’s characterization above), implying that a major role in the definition of disorders of agency related to PDs is played by societal and cultural factors.

Third, and most importantly, diagnoses of PDs are prompted by symptoms typically involving patterns of behavior (actions or omissions thereof) and attitudes, which play a crucial role in the maintenance of the condition:

They could be emotionally cruel, or extremely angry and threatening without just cause; they might self-harm or disengage from the Community without explanation, provoking high levels of anxiety in others concerned for their well-being; they might shirk their Community tasks and responsibilities, leaving others to pick up the work.30

Conversely, with the appropriate therapeutic, interpersonal, and societal support, people with PDs can successfully work to change their problematic behavioral patterns and traits.31 In other words, the possibility for change and recovery for people with PDs crucially involves the

31 Pickard, “What is Personality Disorder?” p. 182.
exercise of their agency within an appropriate therapeutic interpersonal environment. For this reason, and as a result of a clash of intuitions and beliefs she herself experienced in approaching the psychiatric practice, Pickard coined the term “disorders of agency,” thereby instigating an important shift in focus from ‘personality’ to ‘agency.’

Finally, unlike other kinds of disorders of agency (e.g., those associated with schizophrenia and Alzheimer), discussion of cases of PDs and their treatment in the context of therapeutic communities can be successfully conducted quite independently from any specific empirical hypothesis concerning the psychological architecture underlying the impairment (i.e., Sense of Agency/Sense of Ownership, and memory, respectively). By considering disorders of agency related to PDs, the focus is therefore prominently on the level of interpersonal/institutional contexts, relations, and practices, thereby providing a privileged insight into the conceptual and normative framework of agency and responsibility.

Along with the interpretation of these cases provided by Pickard and her specific proposal, my claim will be that cases of disorders of agency bear considerable significance for an understanding of the way we conceptualize the notion of responsibility and the relevant normative practices in non-clinical contexts. This represents a minimal level of normative reflection—that is, which contexts should we consider when we model the notion/concept of responsibility, whether those carry normative preferences (if so, which those are), and whether a homogeneous model would fit our practical needs for the concept of responsibility.

4.2 The Clinical Model of “Responsibility Without Blame”

Pickard’s research and analysis are of further pertinence to my assessment of Korsgaard’s account of agency for the following reasons. According to Pickard, clinical practice involving service users (i.e., patients) with disorders of agency seems to be at odds with both philosophical and ordinary beliefs about the close relation between responsibility and evaluative reactive attitudes, where reactive attitudes are considered the evaluative emotional responses elicited by the kind of attitude toward us manifested by other people’s behavior, like resentment, blame, praise, gratitude. The idea at the core of both the ordinary and the philosophical perspective, Pickard observes, is that responsibility is an essentially moral notion and that reactive attitudes are integral—either constitutively or, more weakly, via a relation of appropriateness—to the practice of holding people responsible. In other words, holding someone responsible characteristically involves morally evaluative, personal reactive attitudes.

Korsgaard seems to be on the same page with respect to the connection between responsibility and reactive attitudes. If we think about Korsgaard’s characterization of the two implications of the ordinary concept of agency—and of the linchpin between the two, the notion of responsibility—we can recall that there were two claims at play. The first, which I will refer to as the Identification Thesis, is about the intimate relationship between agents and their actions via the notion of personal, or practical identity—of which the actions are a manifestation or expression, thereby qualifying as her actions—since “the most essential part of the person is constituted by her actions”. The second claim states that the reactions to someone’s actions—the practical attitudes constitutive of treating someone as an agent—are

---

32 Now that this distinction is in place, I henceforth use the term ‘disorders of agency’ narrowly—or in Pickard’s sense, so to speak—to refer just to disorders of agency related to PDs.

33 Korsgaard, Self-constitution, Section 5.5.1. Here is the full quote: “The intimate connection between person and action does not rest in the fact that action is caused by the most essential part of the person, but rather in the fact that the most essential part of the person is constituted by her actions.” I use the Identification Thesis quite vaguely to refer both to the metaphysical and the practical relation between an agent and her actions. In the next sections I distinguish between the metaphysical and the practical counterparts of the Identification Thesis, and thereby use it in a more restrictive way, to refer just to the practical counterpart.
evaluations/assessments, and as such, crucially involve standards of appropriateness. Korsgaard’s interpretation of the kind of appropriateness at stake is moral (as is the one proper to common-sense).\textsuperscript{34} I shall call this the \textit{Moral Evaluation Claim}.

The two claims work together in Korsgaard’s theory in that agency is characterized both in metaphysical terms (as the activity of self-constitution) and in practical terms (as choosing what to do, thereby taking responsibility for a specific practical identity that enacts a certain pattern of self-reflection and self-deliberation). Since the practical, for Korsgaard, is subsumed into the moral (the Categorical Imperative is \textit{the} constitutive principle of agency), then the fundamental dimension of evaluation for agency is the moral one.\textsuperscript{35}

Paying attention to therapeutic contexts, however, sheds light on the fact that the relationships among ‘failures’ of agency (including those involving morally relevant aspects), competent agency, responsibility ascriptions, and personal identity, is more complex than Korsgaard seems to have it.

Crucial for the treatment of disorders of agency is the undertaking, on the part of the practitioners, of a stance of “responsibility without blame.” In Pickard’s words:

\begin{quote}
Service users were responsible for their actions and omissions and accountable to the Community for them, but an attitude of compassion and empathy prevailed, and they were not blamed.\textsuperscript{36}
\end{quote}

For Pickard, blame is a complex emotion involving a negative affect towards the addressee, which is characteristically accompanied by a sense of entitlement to that negative reactive attitude in relation to the addressee’s wrongdoing. The notion of responsibility without blame captures the fact that in their therapeutic relations and interactions with practitioners, service users can be successfully held accountable for their actions without being morally evaluated. In other words, in these cases, usual components of holding someone accountable for her actions, such as ascribing blameworthiness and imposing consequences, remain—with the exception of the expression of moral blame. Still, without blame responsibility is preserved, and successfully goes through, despite or perhaps in virtue of that very omission.

Dissecting the elements at play in Pickard’s model of responsibility without blame reveals that Therapeutic Communities are interaction spaces between service users and mental health practitioners that focus on and encourage change in service users’ behavior. The therapeutic nature of these spaces lies in both the goal of the interaction (i.e., fostering change of maladaptive patterns of behavior; improvement of interpersonal functioning), and the type of relationship between the service users and practitioners (i.e., between a user and a provider of a health care service, which is a distinctively asymmetrical relationship). The specificity of the situation (in terms of goal and type of relationship) notwithstanding, “it is a presumption of treatment,” Pickard explains, “that service users have choice and control over their behavior and can therefore be asked to take responsibility for it, as we naturally say.” This presumption pervades and structures the interactions within Therapeutic Communities, in that it is made

\textsuperscript{34} The reactive attitudes which are supposed to be involved in the identity implication of the ordinary concept of agency are “normative or at least personal;” Korsgaard, “The Normative Constitution of Agency,” Section 1.2.

\textsuperscript{35} Cf. Korsgaard, \textit{Self-Constitution}, Section 8.5.2: “[…] integrity in the metaphysical sense—the unity of agency—and in the moral sense—goodness—are one and the same property.” Here Korsgaard refers to Kant’s and Plato’s arguments, highlighting the same pattern of establishing that a normative property (“universalizability in Kant’s argument, justice in Plato”) just is the counterpart of a metaphysical property (“autonomy in Kant’s argument, constitutional unity in Plato’s”). Her account follows the same strategy; in other words, Korsgaard endorses the “ancient metaphysical thesis of the identification of the real with the good;” (\textit{ivi}, section 2.1.1). See also \textit{The Constitution of Agency}, pp. 110-113.

\textsuperscript{36} Pickard, “Responsibility without Blame,” p.1135.
explicit through “the language of agency and responsibility.” What service users usually lack, is “full conscious knowledge of why they are behaving as they do, or what the full effects of their behavior on others may be,” as well as a full control over their behavior. Evidence from clinical treatment shows that service users routinely exercise choice and control over their behavior “when they have incentive, motivation, and genuinely want to do so.”

Control and conscious knowledge are therefore graded notions.

The fact that this presumption is operative and constitutes a necessary condition for effective treatment is crucial, as it means that agency exercised in clinical contexts shares the same features, in different degrees, of nonclinical/ordinary contexts. In particular, the parallel with the ordinary practice of holding people responsible is maintained in that judgements and ascriptions of blameworthiness are not suspended, and the demands (i.e., specific requests, and potential imposition of negative consequences) and expectations involved in participating in the practice of regarding oneself and holding each other responsible are not lifted either. What practitioners refrain from doing—thanks to appropriate training and cultivation—is to express blame towards service users (where the expression of blame might be considered a minimal level of acting on it). By taking a blame-free participatory stance—which is, importantly, different from Peter Strawson’s objective stance—practitioners enable an environment where service users are not absolved from responsibility, and their agency is positively affirmed, rather than denied.

Pickard’s contention is therefore that if the evaluation at stake in reactive attitudes is construed as moral in kind—thus involving moral standards in the ascription of agency and responsibility—to close of a connection between responsibility and morality is established; in particular, one that obscures the fact that responsibility is a distinctively agential concept, alongside, but not synonymous with or a part of, morality. The experience stemming from therapeutic treatments of disorders of agency seems to testify exactly to the conceptual and practical viability of the distinction between moral and agential responsibility, as the exercise of agential capacities can be enhanced specifically by refraining from the moral evaluation involved in blaming.

In light of the conceptual possibility and practical viability of the clinical stance of responsibility without blame, Pickard develops a conceptual framework that clearly distinguishes responsibility, blameworthiness, and blame. She distinguishes two varieties of blame: affective and detached. Affective blame involves a characteristic ‘sting’ and a moral connotation that represents a hindrance to the effective exercise and improvement of the service user’s agential capacities. On the other hand, what makes detached blame conducive to effectively securing service users’ accountability and change in behavior is that it is devoid of “a sense of entitlement to any negative reactive attitudes and emotions one might experience, no matter what the service user has done.”

What emerges from the clinical practice with sufferers of disorders of agency shows that while their participation in responsibility practices—i.e., being considered as responsible agents and held in relations of accountability—is indeed a necessary condition for them to effectively

37 Ivi., pp.1137-1138.


39 Pickard, “Responsibility without Blame,” pp. 1141, 1146. Pickard’s notion of ‘affective’ blame is similar to, but differs under important respects from, Angela Smith’s notion of ‘active’ blame. See ivi, p. 1141 n. 13.
be accountable and take responsibility for their actions, thereby exercising agential capacities, moral evaluation is not essential to it, and could actually be a hindrance. I therefore side with Pickard in maintaining that these concepts should be best understood in morally neutral terms.

Getting back to the assessment of Korsgaard’s proposal, this consideration seems to pose a serious challenge to the *Moral Evaluation Claim*—that is, morally evaluative reactive attitudes are not necessary for the practice of holding someone responsible. My suggestion is therefore that the *Moral Evaluation Claim* be turned into the *Agential Responsibility Claim*. While this move constitutes a seemingly minor adjustment to Korsgaard’s theory, it is one with big repercussions, in that a more nuanced account of agency—in particular, one that clearly distinguishes between moral responsibility and agential responsibility—would thereby broaden the scope of humans we can consider, and who can understandably consider themselves, as agents.

Elaborating on the elements pinpointed by Pickard as central to the presumption of treatment in clinical contexts, I will regard as *agential* those capacities in virtue of which an agent is responsible (without blame), or can take responsibility for, her behavior. Agential capacities include: i) conscious knowledge of what one is doing, ii) choice, and iii) a degree of control over one’s behavior. Accordingly, I propose the following claim:

**Agential Responsibility Claim**: treating someone as an agent is to treat her as accountable/answerable for her actions, which presupposes that she has knowledge of what she is doing, can exercise choice and a degree of control over her behavior.

It is this presupposition’s structuring of the interpersonal attitudes and reactions to one’s behavior that proves to be a necessary condition for exercising agency. In other words, it is not just the agent’s individual psychological arrangement that matters for agency, but rather the fact that any such capacity of the individual agent is actually acknowledged/recognized, and interacted with, within an interpersonal relation. What will be involved in actually treating someone as responsible will vary depending on the context, the nature of the relationship with her, along with a consideration of her past personal history. In any case, it will involve some concrete effects in terms of re-action/response, to the agent’s actions.

While corroborating the relevance of Korsgaard’s approach to responsibility as a practical relation for the notion of agency, treatment of disorders of agency indicates that those who in Korsgaard’s terms would be ‘mere heaps,’ (and therefore less of fully unified agents) would rather seem to be always already agents, and never cease to be such. The process of recovery for people with disorders of agency—whose actions/omissions would, according to Korsgaard, consistently fall short of making them *unified* agents—shows that any degree of realization of one’s agential capacities, as manifested in her actions, is sufficient to hold her accountable for her behavior, and for her to effectively take responsibility for those actions. A ‘disorder of agency’ does not make the person who is diagnosed with it a defective *qua* agent. The defect, then, would not be in her agency. Rather, it would be in the dysfunctional dynamics of the agent’s proximal relational environment, and in those resulting from her past history (including

---

40 Here I draw on Pickard’s characterization of what is presupposed by effective clinical treatment, *ivi*, p. 1141.

41 Pickard’s understanding of this notion is admittedly an intuitive one. By ‘conscious knowledge’ of behavior she refers to “the way we normally know what we are doing when we are doing when we do it. […] Normally, we have some knowledge of why we are acting, some knowledge of how we are acting, some knowledge of what we intend in acting, and some knowledge of what effects our actions have on the world” *ivi*, p. 1136 n. 5.

the complex, overlapping influence of many different structural societal factors). Overall, disorders of agency and their treatments point to the relational, interactional character of agency and responsibility, as well as their social dimension. The discussion in the present section is meant to challenge the very idea of *defectiveness* attached to the term ‘disorder,’ stemming from the placement of moral blame on an individual.

Alternately, we should consider agency as a complex and interactive capacity, the exercise of which crucially depends on interpersonal relationships and contextual factors—at any degree or stage of actualization. A characterization of agency in these terms would imply not just that a lack of appropriate interpersonal relationships and material opportunities might importantly impact on the development and exercise of the capacity of agency—which Korsgaard herself would certainly concede—but eventually lead to the following stronger claim:

**Interaction Claim:** agency itself is *constitutively* a function of one’s location within a network of social and interpersonal relationships, as well as of the quality of these relationships, as perceived by the agent.

The hypothesis central to the reframing of constitutivism as a theory of agency that I pursue in this work intends therefore to build upon and strengthen Korsgaard’s practical approach to agency and responsibility.

4.3 Socially Displaced and Distorted Agency

By challenging the categorization of ‘disorders’ of agency as instances of defective agency, the previous sections were meant to make a case for distinguishing agency as a capacity from its actualization in an agent’s actions. The connection between agential capacities and the practice of holding someone responsible for their actions has been emphasized by introducing the Interaction Claim. Taken together, these two moves are designed to shift the focus from conceiving of the Identification Thesis (regarding the relation between an agent and her actions) as a metaphysical relation (of self-constitution), to a practical relation—i.e., taking place within interpersonal and social interaction.

A further instructive case to illustrate the interactive/relational character of agency, and therefore to support the Interaction Claim, is that of agency exercised in contexts of oppression and violence. In the Disorders scenario discussed above, the role of interpersonal relations and practices of accountability was showed to be a necessary condition for the agent to effectively develop and exercise their agential capacities. In the Displacement scenario presented in this section, the Interaction Claim gains (indirect) support by looking at the negative impact that societal relations might have on individual agents’ ability to actualize their agency. In these cases, agency operates despite/against/within structures of power and oppression which systematically distorts the intentions of some agents. As a result, the characteristic practical

---

43 See Korsgaard, *Self-Constiution*, Sections 1.4.7 and 2.4.2.

44 For example, the cases of feral children: being isolated from other people prevents them from having agency in any usual sense of the word. Another example is that of solitary confinement, that has proven to have disintegrating/destroying effects on sense of self (see Gallagher (2014, section 3) for an overview), which I would expect to extend to agency as well.

45 Think, for example, of the importance of trust and trustful relationships for adequately developing and exercising one’s agential capacities, or of the (enhancing/diminishing) effects connected with interiorizing certain (positive/negative) social labels.

46 See *infra*, Section 5.3.
relation of ‘authorship’ between an agent and the actions she performs, is bent to the effect that the actions’ locus of significance is displaced from their agents.

The Displacement scenario represents a problem for Korsgaard’s theory because a consistent failure in the agent’s efficacy—i.e., the “success in actually performing the action, in doing something that counts as an action”47—would eventually fail to make her an agent at all, if anything by disempowering the agents’ potential for agency via the internalization of derogatory labels and other mechanisms of social discrimination and oppression.48 The interpretation of the Displacement scenario that I favor in the following section, on the contrary, accounts for the ‘displacement effect’ without obliterating the agency of the disenfranchised individuals. The explanation I will provide for how that is the case relies on the Interaction Claim and calls for a further investigation of its connection with the Identification Thesis. Before getting there, let me introduce the cases of displaced agency.

The process through which agency gets displaced from an agent has been characterized as a form of “social authoring:”

As an example of social authoring of action, we can refer to the 2005 media coverage of Hurricane Katrina, which fixated on reports of looting in flooded New Orleans (Sommers et al. 2006). Two photographs published by different news agencies captured the public’s attention: One photo features white-skinned people traveling through the flooded area carrying food, and includes the caption, “Two residents wade through chest-deep water after finding bread and soda from a local grocery store after Hurricane Katrina came through the area in New Orleans, Louisiana” (Agence France Press 2005). The other image is nearly identical except the subject of the photo is a black man, and this time the caption reads, “A young man walks through chest deep flood water after looting a grocery store in New Orleans” (Associated Press 2005). Defining their actions as “finding” and “looting,” the captions diverged in how they narrated the same apparent intentional action. Although the caption for the white-skinned agents abstained from normative explanations about their actions, the caption for the black agent affirmatively criminalized him and his actions, and institutionalized that characterization by reporting it as news. […] Some have contended that the divergent captions are a consequence of different standards for the two news agencies that published their respective photos and captions. However, analysts have shown that the captions were consistent with the media’s overall pattern of racialized characterizations of individuals surviving the aftermath of the hurricane (for example, Sommer et al. 2006). […] 

Scenarios such as the looting/finding example highlight how intention is not just authored by the agent, but is also socially authored through others’ discernment and translation of that action. The term social authoring is meant

to convey a relationship of production between “observer” and “act.” I distinguish between social authoring and “social reading” of an act. To read an act is to apprehend an existing meaning, but to author an act is to create something new. When facilitated by reasoning designed to reinforce and rationalize systems of domination, social authoring relies on and further entrenches an institutionally sanctioned distortion of the intentions of some agents.49

Alisa Bierria’s example and analysis quoted above provides us with an important theoretical perspective on the social dimension of agency. Korsgaard’s acknowledgement of this dimension can be found, in a minimal form, in her account of responsibility as a practical relation of mutual reciprocity entered by the agent, inasmuch as she is engaged in the activities of practical reasoning and of exchanging reasons. Significantly, Korsgaard’s claim that the practical relations entered by participating in practices of mutual accountability seems to imply that those relations are inherently symmetrical, and actualized by default in virtue of the agent’s acting in conformity with the principles of practical reasoning.

The phenomenon of socially displaced agency, however, reveals that it is not just the normative relations among the agent’s mental states, established through the activity of practical reasoning and the exchange of reasons (i.e., not just her individual process of self-constitution) that matter for agency. What is relevant for the successful performance of an agent’s action is also the uptake on the part of other agents—that is, a relation of dependence on others for the action to be understood as having a certain meaning, and to take the corresponding intended effect.50 The uptake is in turn a function of the normative relations among the agent and her social and institutional environment, and those relations might also be such that the authoring process of the agent’s intentional actions is displaced from her and subject to distortion—the meaning of her actions is thus defined away from her. In the finding/looting case, it is a feature of the agent’s identity as a member of a racialized group that triggers a criminalized labelling of the action, thereby revealing the structural character of racism as a form of discrimination and oppression—that is, embedded, interiorized, reproduced, reinforced, and even legitimized, within social structures.

To be sure, the process of social authoring of an action is not inherently bad. If anything, it is key to understanding the role played by collective intentionality for both individual and shared/collective agency. However, the potential for, and persistence of, mismatches between the action as intended by the subject who performs it, and the action as attributed to her (or, as Bierria claims, authored away from her), raises the question of what the relation between the reading and the authoring of an action is, and calls for a closer examination of the distinction between the exercise of agency, attributions/ascriptions of agency, and their reciprocal connection.

How do these questions impact on the Identification Thesis that I have identified as central to Korsgaard’s account? How do individual and collective agency relate to each other? How does morality—or better, lack thereof, in terms of various forms of structural injustices pervading and structuring agents’ field of possibilities for action—connect to agency, after all? All of these issues are deserving of their own dedicated study in the future.


50 Here I am relying on a parallel between agency and linguistic communication, where the notion of uptake and reciprocity on the part of an audience are considered a condition for the successful performance of an illocutionary act by a speaker. Cf. infra, Section 8.
For present purposes, the most pressing question prompted by cases of socially displaced/distorted agency via social authoring is whether social uptime is necessary for agency. Are cases of displaced agency ‘failures’ of agency? Bierria is very straightforward in claiming that although the phenomenon of social authoring displaces some agents’ agency, it does not ‘disable’ nor ‘erase’ it: While individual agency is sanctioned/validated by social uptake, it does not depend on it. Characterizing the phenomenon in terms of a ‘failure’ of the oppressed subject’s agency, Bierria contends, is inaccurate and misleading:

Disenfranchised agents are ‘doing their part’ to reflect, reason, visualize, anticipate —intend— and meaningfully act according to their intention [...] The failure in these cases does not emerge from these actors’ agency, but lives in the distortion of these subjects’ intentions by others as well as the social and political conditions that legitimize those distortions.51

Bierria’s proposal is then to drop “a binary or scaled model of agency that gauges subjects as having more or less, abled or disabled, or successful or failed agency,” and adopt instead a “heterogeneous framework” of agency, which acknowledges different kinds of agency. The kind of agency exercised by an agent—which, Bierria proposes, can be hegemonic, alien, or insurgent/resistant/subversive—depends on how factors pertaining to the social dimension “position her in relation to others when practicing intentional action in the social sphere”.52 Whether hegemonic, insurgent or otherwise, I will explore the hypothesis that the exercise of any kind of agency involves some sort of interpersonal uptake or recognition (be it that of a minority, down to just another person).

Without following Bierria’s proposal of adopting a heterogeneous framework of agency, her analysis of the phenomenon of displaced agency is important in its own for the aims at issue. It helps clarify that human agency is an inherently relational/interactive phenomenon, and that the reciprocal position of the interacting agents is crucial. Overall, Bierria’s analysis corroborates the conclusions drawn from the case of disorders of agency. That is, that agential power and status are constitutively sensitive to the position an agent occupies in the relevant network of societal and institutional normative relations. These relations provide the architecture,53, or infrastructure—so to speak—within which agents come to be and operate. Individual agents, on their part, do, however, also play an active role in creating, maintaining/reinforcing, or challenging and reforming their infrastructure.

5 Elucidating Constitutive Normativity

An intended strength of Korsgaard’s constitutivist project is its aiming at bringing together metaphysical as well as psychological aspects of normativity. Korsgaard’s use of the terms ‘normative’ and ‘normativity,’ however, oscillates between different meanings. In order to clarify which sense of ‘normativity’ is at stake at which stage in her argument, let us start by looking at what kind of role the specific stage or step plays in the overall constitutivist argument.

Having set aside any intention to assess the meta-normative ambition of Korsgaard’s project in its foundational aspect—that is, the vindication of normative authority of normative claims

52 Ibid.
of practical reason—my aim here is to clarify the relevant sense of normativity pertaining to
the first-order level of the constitutivist theoretical nucleus advanced by Korsgaard. This
constitutivist theoretical nucleus amounts to the *metaphysical-cum-normative* claim about
action/agency having a constitutive aim/function, which sets normative principles/standards
(premises 2 and 3 in the argument reconstructed in Section 3), along with the claim about the
inescapability of acting (premise 1).

Now, the interpretation of the Disorders scenario endorsed in the previous section, if sound,
shows that the Moral Evaluation Claim at play in Korsgaard’s account is highly contentious.\(^5^4\)
I have claimed that a notion of *agential* responsibility as distinct from moral responsibility
carries more predictive and explanatory power regarding the exercise of agency. To recall,
Korsgaard claims that the constitutive principle of agency is the moral law.\(^5^5\) If my conclusion
that human beings can and do ‘unify’ themselves into agents *outside* the moral law is correct,
then premise 3 of Korsgaard’s constitutivist argument is false. The constitutive normativity
proper to agency does not *per se* entail the normativity of morality.

It is important to stress once again that the step in the argument that carries the normative
weight/import/implications is made in the characterization of what, metaphysically, agency is
(roughly, premise 2). Having ruled out moral normativity, what does, then, constitutive
*normativity*—i.e., the normativity of constitutive norms/principles governing the constitution
of agents—consist in? What is its nature, and what is the kind of necessity attached to it? In
what follows, I provide three possible interpretations of Korsgaard’s characterization of
constitutive norms of agency in order to assess the plausibility of those interpretations in light
of the analysis of cases of supposed defective agency conducted in Section 4.

The constitutivist theoretical nucleus of Korsgaard’s account—as exemplified by premise 2:
“In order to act, human beings must unify themselves into agents”—amounts to the following
two claims. On the one hand, there is the claim that (i) agency/action is a functional kind, or
that the function of action is to constitute the one who exercises it/perform it into a unified
agent, and denotes the special (metaphysical) relation between the agent and her actions: a
relation of self-constitution. In other words, there is no (unified) agent preexisting her actions.
On the other hand, we have the claim that (ii) the essential feature of an action is whole-person
attributability, or its being authored; where being authored/whole-person attributability refers
to the (practical) relation between an agent and her actions as mediated by the notion of
personal, or practical identity, which denotes the agent’s engagement in practical reasoning and
deliberation.

According to what I have identified as the *Identification Thesis*, someone’s actions,
Korsgaard claims, constitute her most essential part, and are the expression/manifestation of
her practical identity, or capacity for normative self-government. With this rough recap of
Korsgaard’s constitutivist theory of agency/action in place, let us turn to the task of clarifying
the notion of constitutive normativity operating within it.

Korsgaard puts forwards an elaborated account of the psychology and metaphysics of agency,
revolving around the notion of constitutive principles, or norms. Let start, then, with a look into
what constitutive norms are. According to Korsgaard’s most recent characterization of

\(^5^4\) I.e., a moral failure does not imply a failure of agency; bad/defective actions do not transmit failure to the agent
*qua* agent.

\(^5^5\) Cf. Korsgaard, *Self-Constitution*, pp. xii-xiii: “the kind of unity that is necessary for action cannot be achieved
without a commitment to morality. […] the only way in which you can constitute yourself well is by governing
yourself in accordance with universal principles which you can will as laws for every rational being. It follows
that you can’t maintain the integrity you need in order to be an agent with your own identity on any terms short of
morality itself. […] a commitment to the moral law is built right into the activity that, by virtue of being human,
we are necessarily engaged in: the activity of making something of ourselves. The moral law is the law of self-
constitution, and as such, it is a constitutive principle of human life itself.”
constitutive norms, they “arise from the nature of the object.” The idea, she specifies, breaks up in two parts:

**Constitution Requirement (C):** “unless the object conforms to the standard, it ceases to be the kind of object that it is.”

**Self-constitution Requirement (SC):** “the object makes itself into the kind of object that it is by conforming to the standard.”

This characterization of constitutive norms is meant to capture what for Korsgaard is a crucial feature of constitutive normativity—agency's liability to standards is part of its constitution. In other words, the source of constitutive norms of agency lies in the nature of agency itself. The characterization is also meant to encode the double nature of constitutive norms/principles—i.e., the fact that the norms have both descriptive and normative implications with regard to the activity they are norms of. They constitute the activity they govern, thereby describing the conditions for the possibility of agency itself; and they are in force, thereby exerting a normative role, for the individual, under pain of kind-membe exclusion. This double nature is the hallmark of constitutive normativity.

At this point, having exposed the supposed cases of defective agency as instances of genuine agency in Section 4, we are left with the task of determining if Korsgaard’s characterization of the constitutive norms of agency can be interpreted in a way that accounts for the conclusions reached at the end of those cases. If there is an interpretation, then we merely need to take it up, along with the suggested changes to the normative theory of agency suggested in Section 4, to be able to include those with disorders of agency and those whose agency is subject to social displacement as agents proper. If not, we will have to suggest alterations to her more foundational characterization of the constitutive norms of agency to cover the appropriate anomalies. In what follows I present three possible interpretations of Korsgaard’s characterization of constitutive norms reconstructed above.

### 5.1 The Metaphysical Interpretation

According to what I will call the Metaphysical Interpretation, the constitutive principles of agency are metaphysical principles describing the formal principle of agency—that is, self-determined efficacy. As a result, kind-membership (i.e., whether someone is an agent and genuinely exercises agency) depends on a certain relation between her and her behavior being in place—that is, by the fact that actions are those pieces of behavior instantiating a certain metaphysical property (autonomy/constitutional unity) that reflects on the agent. The kind of necessity associated with constitutive norms, under this interpretation, would amount to metaphysical necessity.

This interpretation is supported by Korsgaard’s deliberate pursuit, within her constitutivist project, of a characterization of the metaphysics of normativity, which results in several distinct

---

56 Korsgaard, “How to Be an Aristotelian Kantian Constitutivist,” Section 1.2.

57 Cf. Korsgaard, “The Normative Constitution of Agency,” Section 1.3: “Agency is liable to a standard of success and failure from the inside, that is, by virtue of its own nature: such liability is part of its constitution.”

58 The normative character of constitutive norms has been targeted by the ‘Why be an agent?’ objection, exemplified by Enoch, *op. cit.* Undertaking a reconstruction and assessment of the objection is beyond the scope of the present work, but represents an important issue to account for given constitutivist strategies’ claims about the inescapability and non-optionality of agency. For a defense of constitutivism against the objection see Ferrero, “Constitutivism and the Inescapability of Agency,” and Ferrero, “Inescapability Revisited,” *manuscript.*
claims. These are: her endorsement of the ‘classical view’ that normative principles are principles of the unification of the manifold; the claim about the interdependence of the metaphysical and the normative properties essential to the nature of action; and, finally, her statement that the appropriate metaphysics for the concept of activity is a Platonic metaphysics. Besides inheriting many potentially problematic issues associated with each of the individual claims supporting it (which I am going to overlook, here), the major problem with the Metaphysical Interpretation is that it is inherently unstable, in that it tends to collapse on the Individual Psychology Interpretation (presented below).

Korsgaard takes agency’s metaphysical property (autonomy/constitutional unity) to ultimately amount to the agent’s psychic unity, issuing from the process of self-reflection, and resulting in her maxims/principles of choice displaying a certain normative property (universalizability/justice). The way the formal principle of agency is given specific content, Korsgaard claims, is through the contingent practical identities, which—by providing the agent with reasons for acting—determine a functioning self. However, as mentioned, on this account the metaphysical necessity associated with constitutive norms, would collapse into a sort of psychological necessity. Let us turn, then, to the Individual Psychology Interpretation.

5.2 The Individual Psychology Interpretation

According to the Individual Psychology Interpretation, the constitutive norms of agency describe the psychological structure that an individual agent must have to be able to act at all, or to author an action. This was, in a nutshell, the idea behind Korsgaard’s version of the authorship view. As we have seen, according to Korsgaard, exercising agency amounts to being engaged in a self-guiding activity—that is, an essentially conscious activity—even though the explicit awareness that one is engaged in such an activity does not necessarily have to be available to the agent. And yet, the compliance to agency’s constitutive norms is crucially mediated by the agent’s own conception of herself as an agent. This is where Korsgaard’s notion of practical/personal identity, and the normative role associated with it—i.e., the fact that it provides reasons for acting—come into play.

If we interpret the constitutive norms of agency along the Individual Psychology Interpretation, however, we have to face the following predicament—compliance (to any degree, and at any level of conscious awareness) with the constitutive norms of agency would be not necessary for an individual to be engaged in the activity of self-constitution because less than fully unified selves are always already agents. Compliance with the norms of agency under the Individual Psychology Interpretation, would also be not sufficient for successfully exercising agency, since perfectly unified selves without appropriate external/contextual/societal backup are not able to act as they intend.

To recall, the conclusion reached in discussing the Disorders scenario was that what proves crucial to the exercise of someone’s agency is her being held in interpersonal relationships, being treated as legitimate terms in relationships of accountability independently of the psychological quality (i.e., the realization of higher order self-reflection capacities, ensuing in a more unified self) their actions might display. If anything, it is through inclusion and participation in practices of accountability that the relevant psychological organization can be implemented. As to the Displacement case, the conclusion reached was that even when agents

59 Korsgaard, Self-constitution, Sections 2.1.1 and 8.1.3.
60 Ivi, Sections 8.1.3 and 10.2.3.
61 This is also David Velleman’s view. For a characterization of the authorship view, see Millgram, “Practical Reason and the Structure of Actions,” Section 2.
successfully intend and meaningfully act according to their intentions, the normative relations between them and their environment can be such that the characteristic relation of authorship they entertain to their actions is systematically displaced/distorted, and yet, their agency is neither disabled nor erased.

What these scenarios show is that an interpretation of constitutive norms of agency focused on the Individual Psychology is therefore partial, and urges us to look beyond the individual agent’s psychological arrangement. In particular, both cases of alleged defective agency point to the necessity to look at the individual’s interpersonal and social relationships. Let us, then, consider a final possible interpretation of constitutive norms—the Practical Interpretation.

5.3 The Practical Interpretation

According to the Practical Interpretation, the constitutive norms of agency describe the practice of agency. Exercising agency means to participate into a practice—a rule-/norm-governed activity—whose dynamics/functioning can be described through the constitutive norms.

The Practical Interpretation of constitutive normativity is textually well supported, as well. Korsgaard’s takes herself to be engaged in providing a practical account of responsibility. She follows the Strawsonian strategy of accounting for the notion of responsibility in terms of an agent’s being in a network of interpersonal relations, qua member of a certain community. This practical account was explicitly propounded in opposition to metaphysical approaches to responsibility, and, plausibly, to merely psychological ones. It is a practical account because it refers to the normative practice of holding each other responsible for our actions, characterized by interpersonal relationships, attitudes, and expectations towards others qua persons. Participating in such a practice amounts to inhabiting the perspective of a practically deliberating agent and entering practical relations of reciprocity with other agents—i.e., treating them as responsible persons and committing oneself to be so treated—through the exchange of reasons which makes joint deliberation and commitment possible. Inasmuch as one first-personally engages in and evaluates practical reasoning and deliberation, constitutive norms set standards of evaluation and specify what is permissible, forbidden or prescribed for being acknowledged and being held in the practice, thereby playing a markedly normative function.

By focusing on agency qua interpersonal activity/practice, the Practical Interpretation seems well suited to positively account for the conclusion drawn from Disorders and Displacement cases. In particular, it fits in with the relational definition of what it is to be an agent exemplified by my Agential Responsibility Claim. The issue remains open, however, about the provenance (biological, psychological, cognitive, rational, conventional, social, moral, …) of constitutive norms of agency, and therefore about whether constitutive norms of agency are ‘homogeneous’ in kind. In light of the discussion of the cases of alleged defective agency, there are prima facie reasons for hypothesizing that they are not. Notably, the normativity pertaining to agency is not exhausted by a single kind of normativity (i.e., rational normativity) and features instead a fair

---

62 Although Korsgaard’s and other Strawsonian accounts concern ‘moral’ responsibility, I take their ‘practical route’ to defining the notion to coherently hold also for what I have labelled as ‘agential responsibility.’

63 I see the metaphysical turn in Korsgaard’s latest work to be a departure from the practical character of the constructivist enterprise she takes herself to be engaged in. For an interpretation of the evolution of Korsgaard along this trajectory—i.e., seeing her latest account as expressing a metaphysical, rather than practical, route to agential notions—see Robin Zheng, “Attributability, Accountability, and Implicit Bias,” in M. Brownstein, & J. Saul (eds.), Implicit Bias and Philosophy. Structural Injustice, and Ethics, Vol. 2 (Oxford: Oxford University Press 2016). Zheng’s focus is on moral responsibility.

amount of conventionality and practice-based norms, along with possibly being subject to other sorts of (non-practice-based) normative constraints—including milder, descriptive forms of normativity.

In other words, even if principles of practical rationality are constitutive—to varying degrees—of human agency, other conditions might equally qualify as ‘constitutive’ of (internal/essential to) intentional human agency, as the following sections will illustrate. If this is the case, the indigenous variety of constitutive norms/normativity, along with the consolidated philosophical tendency to consider different kinds of normativity as irreducible to one another, calls for further investigation in the notion: Are different kinds of norms all equally constitutive of agency? Are they constitutive of the same aspects of agency? The coherence of a constitutivist account of agency hinges on disentangling the different varieties of normativity associated with the notion of constitutive norms.

6 On the Metaphysics of Agency and Normativity

In the above section, I considered three possible interpretations of Korsgaard’s characterization of constitutive norms. I did this because clarifying what kind of norms constitutive norms are, or what kind of normativity constitutive normativity is, represents a crucial step for assessing the viability of Korsgaard’s constitutivist theory of agency. The result of the comparison revealed that the Practical Interpretation is the most promising to positively account for the alleged anomalous cases of Disorders discussed in Section 4, while at the same time being consistent with the spirit of Korsgaard’s overall philosophical project, conceived of in terms of a practical enterprise, as well as preserving the central features of the Individual Psychology Interpretation. In the remainder of this paper I will undertake a conceptual clarification necessary to take positive steps towards a reframing of Korsgaard’s constitutivist theory of agency and action which fully embraces and elaborates on the Practical Interpretation.

6.1 Scope and Methodology

First of all, I shall tackle some methodological issues to make clear how I operate in providing such a clarification, following Amie Thomasson’s distinction among three different kinds of questions asked in contemporary metaphysics. First, Thomasson identifies existence questions: questions about what does and does not exist. Answering such questions produces an ontology. Existence questions might be, in Carnap’s sense, internal—i.e., questions “asked within (or using) a linguistic framework,” whose “answers may be ‘found either by purely logical methods or by empirical methods, depending upon whether the framework is a logical or a factual one’”—and external—“questions about the existence or reality of the system of entities as a whole”. Second, there are relational questions, targeting how many ‘levels of reality’ are there and in what kind of relation (of reduction, supervenience, truthmaking, grounding) they stand.

65 I am assuming that natural facts display a form of descriptive normativity. If this assumption holds, loci classici of the irreducibility claim would then be David Hume’s ‘is/ought’ problem and G. E. Moore’s ‘open question’ argument. More on ‘natural normativity’ framework infra, Section 7. For a basic distinction of varieties of normativity between soft and robust kinds irreducible to one another, see, for example, Leo Zaiber & Barry Smith, “Varieties of Normativity: An Essay on Social Ontology,” in S. L. Tsohatzidis (ed.), Intentional Acts and Institutional Facts (Dordrecht: Springer, 2007).


to each other. Finally, modal questions ask what “properties an object must have to be of a certain type, [...] about what it would take for something of a given sort (or for an individual) to exist: about the existence conditions for things of various kinds.”

Taking up Korsgaard’s project of providing a constitutivist theory of human agency amounts to answering the modal question ‘What would it take for someone to be an agent?’ restricted to this/the actual world (or nearby possible ones). In so far as the question asks for identity conditions, it is also an existence question, specifically, of the internal kind. The question is asked within the linguistic and conceptual framework of the ‘manifest image’ of human agency—i.e., the conceptual framework in terms of which human beings conceive of themselves as agents. As to the relational questions, the constitutivist theory of agency I am after is non-reductive: both the psychological (in a broad sense, to encompass cognitive, conative and volitional elements) and the normative dimensions of agency are equally recognized as parts of reality. While no attempt is made to spell out their reciprocal metaphysical relation and argue for it, I endorse a commitment to a naturalistically respectable framework, broadly conceived. Inasmuch as I am concerned with the metaphysics of agency and normativity in my distinction of different dimensions of agency in Section 9, these distinctions should be borne in mind.

6.2 The Metaphysics of Agency

An adequate clarification of the metaphysics of constitutive normativity requires us to take a step back to Korsgaard’s characterization of the metaphysics of agency—that is, to the claims that (i) the function of agency/action is self-constitution, and that (ii) the essential feature of action is whole-person attributability/being authored. These two claims can be considered as two sides of the same coin (i.e., the Identification Thesis).

The claim that the essential feature of action is whole-person attributability/being authored is, at least prima facie, easily relatable to the Practical Interpretation of constitutive normativity sketched above. On one reading, if the essential feature of an action is its attributability, then the very concept of action and the normative standards associated with it inherently/necessarily involve interaction with other agents and the existence of a practice. Beyond the minimal (largely sub-personal and pre-conceptual) sense of agency and self-efficacy associated with our interaction with the physical environment, the existence of some other—even if just mediated by the use of language or other cultural tools—represents the background of meaning against which we can recognize ourselves as agents on the richer conceptual level we are concerned with here. The constitutive normativity at stake in this case is distinctively social. Spelling out the metaphysics of agency in terms of social practices places us in the irreducibly normative province of social ontology. Recasting the Identification Thesis through the lenses of the Practical Interpretation, I will refer to the second claim as its practical counterpart, or ITP.

How to make sense of the claim that the function of agency is self-constitution in strictly metaphysical terms (i.e., the metaphysical counterpart of the Identification Thesis, or ITM), without falling into the highly controversial terrain proper to any functional claim invoking

---

68 Ivi, preamble.

69 By naturalistic framework I mean a non-reductive methodological naturalism—i.e., the idea that the phenomenon at issue is liable to an explanation in thoroughly naturalistic terms—and, in a broader sense, the idea that attention should be payed to empirical information about the agents’ situatedness within social structures and practices—in other words, the material circumstances of agency.

70 The claim that the existence of a practice is presupposed—or, in John Rawls’ (1955) words, “stage-setting” (“Two Concepts of Rules,” The Philosophical Review, Vol. 64)—for agency qua social activity is compatible with the possibility of cases of authoring or self-attribution of an action in isolation/when no one else is around.
teleological notions, is however, far from straightforward. This is especially true in relation to a commitment to naturalistic respectability, with which Korsgaard’s account is admittedly supposed to be compatible.\(^{71}\)

The rationale behind the appeal to the notion of function is that it provides a principled way for identifying entities. What an entity is, amounts to its functional organization, and its functional organization is such that it engenders standards of success and failure internal/indigenous to the entity itself. Korsgaard relies on Aristotle’s characterization of living things in terms of self-maintaining forms, whose existence depends on their being engaged in the very activity of self-maintenance and continuous reproduction of their specific functional organization.\(^ {72}\)

Are the normative standards engendered by self-constitution as the function of action the same as those involved in the practice-based constitutive normativity? A practice is a rule- or norm-governed activity in which certain behaviors count as fitting and others are out of place. In other words, a practice is characterized by normative standards whose violation/deviation from the norm warrants a negative evaluation, disqualification/exclusion, or sanction/penalty. As mentioned, this kind of normativity seems to be distinctively practical (norms for action).\(^ {73}\)

In contrast, given that the notion of proper function is appealed to in the attempt to pin down the identity conditions of an entity, a violation of those conditions would make it a defective/malfunctioning entity, or another kind of entity altogether (norms for being). In this case, the functional organization of an entity sets the conditions for its existence and the standards for its evaluation, whose normative character is non-practice-based, yet seemingly constitutive of agency as well.

To explore how these different kinds of constitutive normativity interweave into the metaphysics of agency, I shall look into the Autonomous Systems Account of functional normativity, and use the taxonomy of the varieties of normativity it provides as a framework against which some aspects of the constitutive normativity of agency might be tentatively understood.

### 7 The Autonomous Systems Account of Normativity

Within contemporary debates in the philosophy of biology, the Aristotelian idea of functional normativity has seen a resurgence. This is especially the case as it pertains to the organizational structure of living systems. One such account is the Autonomous Systems Account (ASA) of functional normativity.\(^ {74}\) The ASA looks at normativity as a pervasive phenomenon “inherent

---

\(^{71}\) For an overview on the lively debate around teleological notions in biology, see Colin Allen, “Teleological Notions in Biology,” in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (1996/2003, Winter 2009 ed.). Korsgaard’s declaration of compatibility between her defense of teleology and Darwinian approaches can be found in *Self-constitution*, Section 2.3 “In Defense of Teleology,” see in particular section 2.3.1. See Alvaro Moreno & Matteo Mossio, *Biological Autonomy: A Philosophical and Theoretical Enquiry* (Dordrecht: Springer, 2015), Introduction, for a historical reconstruction that traces back to Kant the use of the notion of teleology and purposiveness in the “holistic and circular organization of biological systems,” or self-organization, and contextualizes the “Kantian-inspired organicist ideas” in the most recent research paradigms in biology.

\(^{72}\) For Korsgaard’s arguments see *Self-constitution*, Section 2.3 (“In Defense of Teleology”), and *The Constitution of Agency*, Essay 4 (“Aristotle’s Function Argument”).


\(^{74}\) Wayne Christensen, “Natural Sources of Normativity,” *Studies in History and Philosophy of Biological and Biomedical Sciences*, Vol. 43 (2012), and Wayne Christensen & Mark Bickhard, “The Process Dynamics of
in the organization or form of living systems, specifically in the form that generates their unity and hence explains their existence”.

Within this framework, the normativity proper to human agency and rationality are grounded in a more basic kind of normativity, shared with non-human forms of autonomous systems such as other animals and simple biological organisms.

A system is autonomous when it actively contributes to generate the conditions of its own unity and ongoing persistence, by building an infrastructure—a “persistent, relatively stable structure that shapes more dynamic system-maintaining processes, with the cell membrane of living cells being a paradigm example”. The infrastructure supports and contributes to the system’s self-perpetuation, often by playing a regulative role in the organization of the system. The autonomy of a system is thus defined as relative to the system as a whole, and in relation to the relevant infrastructure. The resulting functional organization generates normative constraints to which the system as a whole is subject. As a consequence, these normative constraints represent, minimally, but fundamentally, the conditions that must hold in order for the system to exist.

An autonomous system’s ontology, in turn, is specified by the different levels at which the regulative processes and interactions contributing to the system’s existence take place. The ontology of complex autonomous systems (like rational agents) is specified in terms of “hierarchically structured forms of organization,”—the autonomous system, the cognitive agent, the social individual, the person—“each of which impose normative constraints”. The system’s ontology generates a normative cascade ranging from fundamental persistence norms, up to general agency and cognitive norms, skill domain norms, and individual-specific cognitive norms. For example, to be a competent cognitive agent, an agent needs working memory, reasoning, and higher order emotional regulation, along with other kinds of processes and activities and the associated variety of functional norms.

In light of the theoretical tools provided by the ASA, the case of human agency can be analyzed as follows. Human agents are complex autonomous systems made up of a network of interdependent processes (physical, physiological, cognitive, affective, conceptual, metacognitive, volitional, social, etc.). There are thus many levels at which human agents’ self-maintenance/functioning is carried out, and for each level there are specific relations of dependence on different kinds of processes and on the relevant infrastructure.

The consequence of such a view for the first-order constitutivist theory I am concerned with is that we are provided with a naturalistically oriented way of cashing out the notion of functional organization and its normative character as the source of the unity and existence conditions of a living system. Functional normativity in the sense specified by the ASA might therefore be a legitimate candidate for understanding the notion of constitutive normativity in the non-practice-based sense outlined above (i.e., constitutive norms as norms for being).

7.1 Varieties and Sources of Normativity

An important feature of the ASA is that it allows us to distinguish between different normative perspectives. Specifically, a case is made for conceptualizing normativity as extending beyond the realm of rational agency, and admitting of the normative perspective of the autonomous systems considered as whole, alongside the normative perspective of individual persons.

Normative Function,” The Monist, Vol. 85, No. 1 (2002); the present discussion relies on the former. See also the ‘organisational approach’ in the study of biological autonomy developed by Moreno & Mossio, op. cit.

75 Christensen op. cit., p. 104; Cf. ivi, p. 105 n. 1, for references to the current debate on functional normativity.

76 Ivi, p. 106.

77 Ivi p. 110.

24
Different kinds of normative evaluation (/normativity) may be thought of as having a parallel structure, insofar as the following elements are identifiable:

a) a normative perspective (persons, autonomous systems) for which things matter,

b) the nature/basis of mattering/significance (relevance to the person, relevance to autonomous systems), and

c) the mechanism through which entities with normative perspective respond to the normative facts constituted by the respective relevance relations (rationality, regulation). 78

The ASA provides arguments in favor of a naturalist grounding of the normativity of personhood and rationality—that is, it aims at providing an explanation of normativity which is “consistent with the natural emergence of th[is] phenomenon”. 79 It does so by situating it within the more basic kind of normativity pertaining to autonomous systems, addressing therefore the question about the origins of the normativity proper of rationality and personhood. The grounding relation between different kinds of normativity might be interpreted, however, not just in terms of ‘origins,’ but also of ‘constitution,’ according to which, persons are not just descended from autonomous agents, they are autonomous agents: a person is constituted as a certain kind of autonomous agent in the base sense of autonomy, and this makes an important contribution to the normativity of personhood. 80

The relevance of the ASA for the present purposes is neither in its metanormative implications, especially foundationalist ones, nor in the answers it provides to metaphysical questions of the relational kind described above. Rather, it helps us advance our understanding of what the constitutive normativity proper to agency amounts to.

Gathering treatments of normativity across diverse philosophical literature, the ASA distinguishes different kinds of normativity: on the one hand there is descriptive normativity, a minimal kind of normativity which refers to the possibility for a system to depart from the norm, but does not evaluate nonconformance as ‘bad,’ or ‘wrong.’ 81

On the other hand, evaluative normativity involves a comparison between actual and alternative states, and comes in different degrees of strength/robustness, corresponding to:

1) valuations (axiology): a certain state being ‘good,’ or ‘better/worse than’ another;

2) prescriptions (deontic theory): stating how things ‘ought’ to be;

3) constitutive norms: ‘specify rules which must hold if something is to exist;’ ‘are per se non-evaluative, though they can inform evaluations in


80 Christensen op. cit., p. 108.

81 Descriptive normativity would be the one pertaining to natural states of affairs. Cf. supra, n. 65.
conjunction with other information, such as an agreement (perhaps tacit) to play by the rules’—usually involving ‘paradigmatic performance norms.’\textsuperscript{82}

The rich taxonomy of normativity delineated by the ASA, along with the generic structural features shared by different kinds of normative evaluation, provide us with a basic set of theoretical tools to solve some of the difficulties met in providing a coherent interpretation of Korsgaard’s attempt to bring together metaphysical and psychological aspects of normativity. After this detour on the metaphysics of normativity, I shall provide a provisional characterization of constitutive normativity and its heterogeneity.

8 Constitutive Normativity, Regained

Agency has a dual nature, roughly paralleling the nature-nurture debate: it is in part due to our social practices, and in part to our biological, cognitive, and psychological constitution—our being autonomous agents in the terms specified by the ASA. The constitutive norms pertaining to agency can therefore be thought of as having a parallel dual nature: \textit{socially generated}, or \textit{practice-based} constitutive normativity, and \textit{non-practice-based} constitutive normativity.

While a full exploration of the distinction between natural and social roots of (constitutive) normativity exceeds the possibility of adequate exploration here, it is worth pointing to some recent work by Katharina Nieswandt\textsuperscript{83} that might help clarify some of the points I am making. By providing a compelling interpretation of G. E. M. Anscombe’s writings, Nieswandt sets the stage for the development of a rich and coherent Anscombean metaethical framework, grounded on the distinction between two kinds of necessity. Specifically, she draws on Anscombe’s distinction between a “practice-internal or ‘conventional’ necessity” (i.e., “the necessity imposed by a rule, a right or a promise”), and a “practice-external or ‘Aristotelian’ necessity,” according to which human goods, or standards for human flourishing, are grounded in human nature.

Nieswandt’s elaboration of the Anscombean framework is interesting for present purposes because it is yet another approach aimed at elucidating how different kinds of norms—i.e., norms of distinct sources and with different kinds of necessity associated to them (biological, practical, legal, ethical, etc.)—equally pertain/contribute to the constitution of distinctive kinds of entities such as human agents. In particular, it is the individuation of the specificity of the practice-internal, or conventional necessity, that is relevant here, as I arrive at similar conclusions while drawing from a different theoretical approach. In line with the Anscombean distinction between two kinds of necessity, my proposal is therefore meant to be noncommittal about metaethical and evaluatively normative issues. My aim, once again, is to provide an elucidation of the notion of constitutive normativity relevant for a first-order constitutivist theory of agency.

8.1 Non-Practice-Based Constitutive Normativity

How does the ASA relate to the constitutivist theory of agency I am concerned with here? Insofar as we consider agents as autonomous systems—which represent one possible normative perspective from which we can consider an agent’s ontology—the norms governing the different processes (physiological, psychological, cognitive) contributing to the system’s

\textsuperscript{82} Christensen \textit{op. cit.}, p. 105.

functional organization are constitutive in that they represent the unity and existence conditions for the system considered as a whole. This is a first, non-practice-based way of interpreting constitutive normativity, which amounts to functional normativity as specified by the ASA.

Since the characterization of an autonomous system constitutively incorporates a reference to the infrastructure utilized by the system for its self-maintenance, the wholeness of the system and its identity are possible just against the background of the infrastructure and circumstances sustaining it. Given the processual and dynamic character of a system’s functional organization, a human agent *qua* autonomous system, is an inherently dynamic entity, which actively contributes to its own constitution. Being an agent and exercising agency amounts to being engaged in processes and activities in interaction with a rich infrastructure. I am following here an important conceptual distinction operated by theorists of biological autonomy, who identify two interrelated, and yet conceptually distinct, dimensions of biological autonomy: the constitutive one, which largely determines the identity of the system; and the interactive one, which, far from being a mere side effect of the constitutive dimension, deals with the inherent functional interactions that the organisms must maintain with the environment. These two dimensions are intimately related and equally necessary. […] In particular, the emphasis on the interactive dimension implies […] that autonomy should not be confused with independence: an autonomous system must interact with its environment in order to maintain its organisation. […] This is what grounds the agential dimension of autonomy. 84

The interpretation of the ITM that I am favoring here understands agency as an activity of self-constitution which is inherently situated and interactive, thereby providing a footing for the Interaction Claim advanced in Section 4. A consequence of this interpretation, then, would be that it is a matter of constitutive normativity in a non-practice-based sense—i.e., independently from any sort of necessity established within the practice itself, or by it—, that human beings socially organize themselves, and act in the context of practices they themselves create/establish. The kind of necessity associated with non-practice-based constitutive normativity would be something analogous to the necessity of natural-historical judgments, or ‘Aristotelian categoricals’ (or Aristotelian necessity, in Anscombe’s sense), of the form “The S is/has/does F,” and “S’s do F”. 85

Human agency is thus constitutively social, in that social structures and practices are not just metaphysically (logically) prior to individual agency, but also empirically and practically necessary to it, as they provide the material and interactive conditions for individual self-awareness and selfhood. There is in fact a consistent body of empirical evidence (coming from clinical, cognitive, developmental psychology, as well as neurophysiology) supporting the hypothesis of a close interdependence between human subjectivity, and the embodied and interpersonal dimensions of human cognition and meta-cognition, which informs rich strands of research in philosophy of mind and action. 86

84 Moreno & Mossio, op. cit., p. xxviii.
8.2 Socially-Generated Constitutive Normativity

For human beings, kind-membership to agency has both natural and social roots. As autonomous systems and cognizers, human beings are born with agential capacities that are actualized and operated in the context of relationships (material, interpersonal, social), and have effects (physical, psychological, cognitive, normative) on their environment (natural and social), on themselves, and others. As stated in the Agential Responsibility Claim, the notion of an agent is inherently relational. Agents constitute themselves qua social individuals, within social practices and interactions.

Human agency as the activity of self-constitution (keeping Korsgaard’s central claim, in its practical counterpart ITP) takes therefore the form of concrete, embodied, situated social practices—crucially, through conceptually mediated practices of self- and other-ascriptions, which carry further, discursively and socially-generated normativity. Socially-generated normativity ranges from the norms governing acts of promising, commanding, consenting, pardoning, etc., to actions requiring more complex and structured social institutions, such as voting, marrying, and baptizing—all of which have normative states of affairs (consisting of commitments, rights, obligations, licenses, etc.) associated with their performance.

By socially-generated constitutive normativity, I mean the normative dimension associated with human practices, which I will refer to as the performative dimension of agency. The performative dimension of agency is roughly modelled on the performative dimension of language as characterized in Speech Act Theory (SAT) developed within the philosophy of language, and beyond, notably by feminist philosophers in a social account of language use.

A ‘speech act’ (or ‘illocutionary act’) is a term of art that refers to the “performance of an act in saying something as opposed to the performance of an act of saying something”, thereby pointing out that language is used not just to describe how things are, but to do things:

The use of language is a sort of social action, consisting of the production and reception of utterances. People do things with words—this is action; and (though this may at first seem a strange way to put it) they do things to one another—this is social action: They tell one another things, or ask them things, or try to persuade them of things, or whatever.

---


87 See Mariam Thalos & Chrisoula Andreou, “Of Human Bonding: An Essay on the Natural History of Agency,” Public Reason, Vol. 1, No. 2 (2009), for an evolutionary informed account of the distinctiveness of homo sapiens as a bonding species—i.e., capable of collective agency. Bonding, the authors claim, is a “universal of mammalian life.” They emphasize the hyper-sociality of our species to indicate how cooperation, rather than self-interest and individualism, should be understood as the default theoretical assumption.


90 Austin, op. cit. pp. 99-100.

91 Hornsby, op. cit. pp. 129-130.
As such, the term ‘performative’ refers to the production of normative states of affairs—“that is, states consisting of the conjunction or disjunction of an agent with a modal predicate belonging to the deontic kind”\(^{92}\)—associated with the performance of speech acts \textit{qua} social actions. By extension, and more broadly, with the term ‘performative’ I refer to the acts of doing or omitting to do something, generally identifiable by means of the specific changes/effects/outcomes in the agent’s environment that their performance brings about.\(^{93}\)

The normativity proper to the performative dimension of agency is socially-generated and conventional in that it is associated with the production of conventional effects. These effects are “made possible by the social frame and brought about thanks to the kind of agreement between [interactors] about what is being done, which we may call uptake,” and the mark of their conventionality is their “defeasibility,” i.e., “the liability to being annulled in particular circumstances.”\(^{94}\)

This kind of socially-generated normativity is constitutive in the sense that the activities governed by constitutive norms do not exist apart from those very norms. In Korsgaard’s descriptive-cum-normative sense of the term, the activities are at once defined, or constituted, and regulated by them. By illuminating their defeasible character—i.e., that their constitutive function “is only exercised against a background of intersubjective agreement”\(^{95}\)—the speech-act-theoretical analysis endorsed here is a good candidate for understanding the practice-based (yet non-moral) counterpart of constitutive normativity that emerged from the Practical Interpretation advanced above.

In my view, the gist of a constitutivist theory of agency consists in an explanation of the complex organization of processes and practices by means of which human agents can recognize themselves as inherently autonomous, yet social entities. In other words, as individuals occupying the characteristic first-personal (singular and plural) agential perspective of deliberating and evaluating agents, who are autonomous in a distinct manner, richer than the biological sense of the term, and therefore closer to Korsgaard’s use. Such a project is grounded in an elucidation of how different kinds of normativity hold together, and the present work is meant to produce a first step in that direction.\(^{96}\)

### 8.3 Dimensions of Agency

I distinguish three different dimensions of human agency: psychological, cognitive, and performative. In introducing these distinctions, it is important to emphasize that, \textit{qua} theoretical

---


\(^{93}\) For the distinction between behavior and performance (negative and positive) in an Austinian framework, see Marianna Ginocchietti, \textit{Actions: A Plea for an Ordinary Framework} (PhD dissertation: Università degli Studi di Trieste, 2016). Generally speaking, the ‘problem’ of action individuation necessarily points to broader issues of social ontology, which, however, deserve an investigation of their own.


distinctions, they are not supposed to be explanatorily exhaustive categories to understand the phenomenon of human agency, but rather useful tools to analyze some aspects of its complexity in theoretical isolation/abstraction from each other.\(^97\)

By the *psychological* dimension of human agency, I refer to the psychological processes and sub-personal mechanisms underlying the performance and (sub-agentive) control of an action, on the one hand, *and* to the personal-level experience of intentional actions available to the agent’s agential awareness, on the other.

Empirical studies on the role of consciousness in the initiation and control of actions investigate the hypothesis of a sub-personal psychological system, the so-called ‘comparator model of motor control.’ According to several studies,\(^98\) that system is partially responsible for the distinctive phenomenology of agency, or the sense the agent has that she is the author of her actions. These inquiries are meant to shed light on the complex workings of agency, characterized by the interplay between the agent’s conscious awareness and rational control, and sub-personal processes.

In other words, I take the term ‘psychological’ to apply to the self-conscious dimension distinctive of human subjectivity and phenomenological experience, independently of the degree of conscious awareness and cognitive penetrability on the part of the agent’s higher level cognitive and meta-cognitive capacities.

The *cognitive* dimension is distinct from, but functions as a bridge between, the psychological and performative. It encompasses an agent’s non-conceptual (procedural) capacities for regulation and monitoring of her own mental processes and bodily actions,\(^99\) for directed attention, as well as higher order conceptual and reflective capacities for epistemic and practical rationality.

The *performative* dimension pertains to the interpersonal and social practices, which rely on, and gain meaning from, the societal infrastructure within which these practices take place. I use the notion of *societal infrastructure* as an umbrella term, to include the set of cognitive, conceptual, and organizational tools—the paradigm case being language, up to the social constructions/categories, beliefs and attitudes constituting a culture—, various coordination devices (i.e., conventions, rules, institutions), and collective epistemic resources making up the social world.

Distinguishing these three dimensions helps provide a better articulation of the scope of a theory of agency, the focus of which can be on:

1. **agency as a capacity** (agental capacities/powers), exercised by concrete agents, to which different constitutive norms (both non-practice-based and socially-generated) apply, depending on whether we (third- or second-personally) consider them as
   a. biological (autonomous) systems

---

\(^{97}\) The methodological assumption guiding this distinction is that an investigation into each dimension separately from the overall explanation of their reciprocal connection, and, broadly speaking, from addressing relational questions, metaphysically speaking (see *supra*, Section 6.1), is a worthwhile theoretical task in itself.


b. cognizers  
c. social individuals  
d. persons

2. **agency as a social activity** (a set of procedures, practices, institutions), actualized in the concrete *actions performed* (actual performances and reactions—either positive or negative—executed, what is done)

3. **agency from within** the embodied (first- and second-personal) agential standpoints

My proposal is to bring the performative dimension of agency into the picture, to emphasize that intentional action is social action. A social action in the minimal sense proper of the performative dimension, is an action whose identification and attribution takes place within the context of a structured practice—even though not necessarily in an institutional context. The agent’s social and interpersonal environments provide not only the interactive conditions for individual self-awareness and self-experience\(^{100}\) crucially involved in the exercise of agency itself, but also the very conceptual tools to frame one’s and others’ intentional actions ‘under a description.’\(^{101}\) Dialogically, we learn the meaning that our actions ‘make’ in the public sphere. The performative dimension I am pointing to represents the *meeting ground*\(^{102}\) where the authoring of intentional actions takes place.

9 Conclusion

In this paper I have analyzed Korsgaard’s characterization of constitutive norms by providing three possible ways to interpret it. I have endorsed the Practical Interpretation—based on Korsgaard’s own practical account of responsibility—as the best candidate to account for, at least in principle, the conclusions drawn in Section 4, where I made a case for considering cases of alleged defective agents as genuinely exercising agency. I concluded that the constitutive normativity proper to agency does not *per se* entail the normativity of morality, and claimed that Korsgaard’s criteria for agency, if we look at the phenomenon of agency in merely individual psychological terms, are neither necessary nor sufficient for capturing why those cases legitimately count as genuine instances of agency.

---


Even highlighting the social aspects of Korsgaard’s account will not be enough to overcome those objections.\textsuperscript{103} Taken together, this means that there is room for an understanding of constitutive normativity/norms of agency in practical, yet non-moral, terms.

While giving prominence to norms of practical reasoning, the Practical Interpretation also pointed out that constitutive normativity is not exhausted by them. Insofar as exercising agency means participating into social practices, there are other sorts of norms (notably, social norms such as those governing the social meaning of actions) which might qualify as constitutive of agency, just as norms of practical reasoning do.

Under the hypothesis of an indigenous variety of constitutive norms/normativity, I turned to what I identified as the core of Korsgaard’s constitutivist theory—the Identification Thesis—to understand what notion of constitutive normativity it relies on. I first focused on the metaphysical counterpart of the Identification Thesis, or ITM—that is, Korsgaard’s appeal to the functional claim that the function of action is to constitute the one who exercises it/performs it into a unified agent. I introduced the Autonomous Systems Account (ASA) of functional normativity, and suggested that functional normativity—i.e., as the unity and existence conditions for a system’s self-maintenance and development in connection with its infrastructure—be considered as a non-practice-based variety of constitutive normativity.

As to the practical counterpart of the Identification Thesis, or ITP, I spelled out a socially-generated or practice-based variety of constitutive normativity, modelled after the normativity proper to the performance of speech acts \textit{qua} social actions as specified by Speech Act Theory (SAT). This preliminary specification of the performative dimension of agency was functional to some analytical distinctions prompted by my proposal that agency should be conceived as socially constituted. While for Korsgaard relationality and the social dimension of agency enter the picture with the relations of reciprocity and mutual accountability that the already extant agent enters in when exchanging reasons and engaging in practical deliberation (in isolation and with others), for me the practice itself is a key element for understanding agency in the first place. An agent is constituted \textit{with} others. In other words, on my view agency itself is constituted in the shared space where the authoring of intentional actions (more or less reasoned) gets recognition by the relevant participants in the practices of exchanging reasons and accountability for actions.

\textsuperscript{103} In Korsgaard’s account, the dialogical character of normative reasons—i.e., that a properly normative reason must be shareable and essentially public—and the relational notion of responsibility as answerability are the elements that are meant to ensure/secure the social dimension of the activity of self-constitution.