

Alienation and the Metaphysics of Normativity: on the Quality of our Relations with the World

Jack Samuel

June 2022

Abstract

I argue that metaethicists should be concerned with two kinds of alienation that can result from theories of normativity: alienation between an agent and her reasons, and alienation between an agent and the concrete others with whom morality is principally concerned. A theory that cannot avoid alienation risks failing to make sense of central features of our experience of being agents, in whose lives normativity plays an important role. The twin threats of alienation establish two desiderata for theories of normativity; however, I argue that they are difficult to jointly satisfy.

Philosophy is to meet its need... by running together what thought has put asunder, by suppressing the differentiations of the concept, and restoring the feeling of essential being

G.W.F. Hegel, *Phenomenology of Spirit*

Our picture of ourselves has become too grand, we have isolated, and identified ourselves with, an unrealistic conception of the will, we have lost the vision of a reality separate from ourselves.

Iris Murdoch, *The Sovereignty of Good*¹

Introduction

Metaethical inquiry is at least partly a matter of making sense of ourselves, of the dimension of our lives that involves thinking and acting as moral agents. What we're doing matters to us because it is about us. I am interested in particular in two sets of potential consequences of accepting a metaethical theory: what it would mean to understand ourselves as the kinds of agents a theory envisions and what it would mean to understand our relations with one another through the theory's lens.²

I argue that metaethicists should be concerned with two kinds of alienation that can result from theories of normativity: alienation between an agent and her reasons, and alienation between an agent and the concrete others with whom morality is principally concerned. A theory that cannot avoid alienation risks failing to make sense of central features of our experience of being agents, in whose lives normativity plays an important role. The twin threats of alienation establish two desiderata for theories of normativity; however, I argue that they are difficult to jointly satisfy.³

In §2 I elaborate what I will call the threat of *normative alienation*: that a theory of normativity could leave agents estranged from the normative facts that the theory

¹(Hegel 1807, ¶7 (trans. slightly modified)), (Murdoch 1971, 46)

²In recent years metaethicists have along similar lines become increasingly concerned with the question of what it would mean for us if a theory of normativity were true. In contrast to conventional appeals to theoretical virtues, or to the consequences of supposedly more fundamental accounts of linguistic meaning or ontology, Parfit (2011*b*) (for example) famously claimed that if non-naturalism is false then nothing matters, and he and his colleagues have wasted their lives. Others have invoked a deep sense of *angst* that underlies the conviction that realism must be true (Blanchard 2020), or even the first-order moral consequences of philosophers accepting realism or expressivism (Hayward 2019). These debates are not new—as Hayward notes, he is entering a decades-old debate between Dworkin, Blackburn, and Williams *inter alia*. My sense, however, is that these sorts of considerations have recently begun to gain traction. See also (Bedke 2020, Zhao 2020). Though I won't engage directly with any of these arguments for or against metaethical positions, my aim is to establish a set of criteria motivated by a similar methodological orientation toward the theory-as-self-understanding.

³A theory of normativity, as I will use the term, consists in an explanation of what reasons are, and perhaps which ones there are, or of what normative facts are, and perhaps which ones are true. In what follows I will speak interchangeably about reasons and normative facts, or about normativity in general, depending on what fits best in context. Nothing, I hope, hangs on the distinction, even if it turns out that normative facts are not in the first instance facts about reasons, contra the 'reasons-first' orthodoxy.

explains. Here I draw on a few familiar literatures and argue that they express different flavors of the same underlying anxiety. In §3 I elaborate what I will call the threat of *social alienation*: that the normative structure of social relations envisioned by a theory of normativity would leave us estranged from one another.

The threat of normative alienation points toward a need to center the agent (the subject, the valuer, the reasoner, etc.) in a theory of normativity. The idea of ‘centering’ the agent will, for now, have to stand as a useful metaphor, buttressed by its application to familiar examples: constructivists, subjectivists, and quasi-realists all center the agent, in the relevant sense.⁴ As a first pass, the idea is that the agent is first in the order of explanation, or the order of conceptual priority. *Agent-centered* theories of normativity (typically though not necessarily antirealist) are well-positioned to explain what normative facts have to do with agents, but limit themselves to bringing others into view indirectly: as a consequence of accepting universal prescriptions, or as the content of a valuing attitude, for example.

The threat of social alienation, however, points toward a need to center the object of moral demands—the other—but the resulting *other-centered* theories of normativity (typically though not necessarily realist) will have difficulty accounting for the significance of normative facts to agents. Metaethical accounts suited to accommodate the role of others in our normative lives ground normativity in e.g. facts about concrete others, or the relations we stand in to them. But facts about our relationships to others, or the properties possessed by others, aren’t the right sorts of facts to ensure that we will have the right kind of connection to them.

If this is right, a theory of normativity suited to avoid both forms of alienation would paradoxically seem to need to center both the agent and the other. The tension can be resolved, however, by centering the constitutive relations between agents as such and others. To paraphrase Michael Thompson (2004, 346), metaethics must be able to record the special sort of dent that *others* themselves make on one’s *own* agency, on pain of leaving us in one state of alienation or another.

1 Alienation in General

A natural worry that’s worth addressing before I begin is that without some account of what alienation is, organizing the following problems under that heading will have diminished explanatory potential.⁵ It is not, after all, a stable or uncontested concept. In the most general use it is more or less synonymous with ‘separation,’ as in the ‘alienation’ of property rights through contract. Philosophers tend to use the term with a negative valence, as synonymous with estrangement—making strange. While alienating one’s property rights through contract is neutral or even good, being alienated from the products of one’s labor, from nature, or from God is bad. An alienated relationship with something is a defective form of that relationship. In its

⁴It is difficult to be more precise in advance of laying out the relevant features these views have in common, as I do in §2, but see the conclusion for more elaboration. To head off one likely misunderstanding, however, I do not mean it in the sense that is roughly synonymous with ‘agent-relative’ and contrasts with ‘agent-neutral,’ as in (Scheffler 1982).

⁵Thanks to an anonymous reviewer for encouraging me to address this worry.

most general form alienation is a problematic separation between a person (a subject, an agent) and something else, something from which we ought not to be separate.

Alienation and the critique thereof operate on a number of levels. In the first instance alienation may be a feature of a way of life or a mode of social organization, as when capitalism allegedly alienates workers from the products of their labor. But insofar as this kind of alienation is subject to critique (and not just material social or political intervention) it is because the alienated mode of social organization embodies an alienated conception of ourselves. We can thus critique in philosophy the underlying picture of the human, the person, the worker, as a way of making explicit the distortion of social organization it produces or reflects. This sort of critique appears in the work and interpretations of ‘continental’ figures like Hegel, Marx, Lucács, Heidegger, and Fromm (see (Schacht 1970) and (Jaeggi 2014) for survey and reconstruction). Capitalism may be (or at least require) a defective relationship between a person as producer and the product of their labor, and is thus a defective form of social reproduction, which embodies a defective picture of the nature of human agency.⁶

At one level of abstraction higher, but in more or less the same tradition, we might say that a *theory* alienates us insofar as it has a tendency to lead to our living alienated lives if we adopt it, or if it informs the cultural backdrop against which we live. On the other hand, we might say that a theory is itself alienating, or embodies alienation, insofar as it represents agents such that if we *were* the way the theory envisions us then we would be alienated, or insofar as it obscures, qua theory, that from which we risk being alienated. This use is probably more familiar in Anglophone philosophy, where worries about alienation are often associated with Bernard Williams or Peter Railton (see §§2.2, 3.1).

In my view, however, they come to the same thing: the alienation at issue is between a person and something from which persons are not properly separate, and it can be realized in a social relation, a mode of production, a theory that informs a social relation or mode of production, or a theory that holds itself out as giving us some insight into what kinds of things we are. Where there is in human life—the life of the metaethicist, and of those they imagine as their subjects of inquiry, for my purposes—a harmony or unity or cohesion or familiarity, an alienating theory imagines us as held apart from that with which we are in reality united and familiar. It makes those things strange to us, and if we could manage to truly understand ourselves as the theory encourages us to, we would suddenly be puzzled by the commonplace, unable to make sense of some important part of our own lives. That is the sense in which, as I will argue, a theory of normativity threatens to *alienate* us from it, by casting it as a strange and distant thing, rather than as something that suffuses or partly constitutes our experience of ourselves. And likewise, this is the sense in which a theory of normativity threatens to alienate us from one another, but making mysterious the grip

⁶See (Julius MS) for an interpretation of Marx along these lines. There is some reconstruction involved in attributing to Marx a concern for correctly conceiving of human agency, but for the sake of assimilating the Marxian critique of alienation into a larger story about the concept I trust that it’s sufficiently well-founded. Compare Axel Honneth, in his foreword to (Jaeggi 2014, vii, *emph. added*): “the concept of alienation... presupposes, for Rousseau no less than for Marx and his heirs, a *conception of the human essence*: whatever is diagnosed as alienated must have become distanced from, and hence alien to, something that counts as the human being’s true nature or essence.”

we have on each other.

2 Normative Alienation

If a normative theory is to offer a satisfying account of reasons it must be able not only to tell us what reasons are, and perhaps which ones there are, but what they have to do with us. It must be able to explain normative facts in a way that connects them to the individuals they are normative for in the right way. In doing so, it will avoid normative alienation.

A normatively alienated agent would be one for whom normative facts were recognizably true, but irrelevant or obscure. They would be, so to speak, *mere* facts, like the fact of whether or not Golbach’s conjecture is true, or the fact of how many stars there are in a distant galaxy: suitable objects of curiosity but possibly unknowable, of no consequence to us in our ordinary lives, or both.

Moral facts cannot be facts like these, and this image of agency—mere receptivity to such facts—cannot represent ours. The first desideratum for a theory of normativity is that in its explanation of how normative facts can be true it contains an explanation of how they are normative *for us*.⁷

The threat of normative alienation appears in different guises: that normative facts could fail to be motivating, that they could fail to be acknowledged as authoritative, and that they could fail to be identifiable. Each of these concerns corresponds to a familiar debate in recent metaethics but they are generally not recognized as expressions of a more general anxiety.⁸ One thing that they do have in common, however, is that they underlie many of the familiar challenges to traditional forms of normative realism, and are offered in support of various agent-centered alternatives.⁹ This is, I argue, no accident. Traditional forms of moral realism threaten to vindicate the truth of normative facts at the expense of undermining the intimacy of their connection to agents. Realists themselves are of course typically untroubled by this, but many (most?) of us find the idea intolerable. We find it intolerable in different ways, and it is not always clear that it is the same thing we find intolerable. But, I argue, these debates represent more local ways in which philosophers have struggled to bring normativity closer to us, and thus have a common source in an implicit concern for something like

⁷This is, in a way, Kant’s demand to explain how reason can be *practical*—see e.g. *Groundwork* 4:395,448 and *KpV* 5:4,44–6.

⁸Shamik Dasgupta (2017) identifies the first and second guises of normative alienation as versions of the same desideratum, though he does not include the epistemic challenge or characterize them as a threat to properly conceiving of normativity’s relation to agents.

⁹“Agent-centered” and “realist” are not antonyms in my usage. Mark Schroeder’s Humeanism is a form of reductive realism about reasons that centers the desires of the agent in its explanation of what reasons there are and which ones exist (Schroeder 2007). Metaethical constructivism, Kantian (Korsgaard 1996; 2009) and Humean (Street 2008), is a paradigmatically agent-centered approach to metaethics, and is sometimes characterized as a form of *procedural* realism about normativity. Agent-centered metaethics contrasts rather with what I will sometimes call “traditional” forms of realism: nonnaturalist (Moore 1903, Ross 1930, Scanlon 1998, Parfit 2011a, Enoch 2011b, Shafer-Landau 2003) and naturalist (Railton 1986, Boyd 1988, Brink 1986, Sturgeon 1988), wherein the truth of normative facts is explanatorily independent of the agents for whom they are normative, and they become practical for agents only by being discovered (and perhaps further by being discovered in relation to the agent’s desires).

normative alienation.

If it were possible for us to be alienated from morality in the way that this anxiety concerns, morality would not be fit to play the role in our lives that it evidently does. The truth, reality, or objectivity of normative facts would have been purchased at the cost of their relevance.

2.1 The constructivist challenge: normative ‘grip’

It is common to characterize Kantian constructivism as an attempt to avoid naturalistic objections to traditional realism without losing the objectivity of moral talk (as noncognitivism is often thought to do). But it is in my view Korsgaard’s key insight that metaethics must avoid what I’m calling normative alienation.¹⁰ She argues that traditional realism leaves an explanatory gap between the normative facts and the agents for whom they are reasons. Realists allegedly hold that “we have normative concepts because we’ve spotted some normative entities, as it were wafting by” (Korsgaard 1996, 44). In other words, even if (contra Mackie) there were entities answering to the realist’s needs it would be a mistake to understand moral language as merely registering their existence, rather than having an essentially practical role. Insofar as such entities were just there, the furniture of the universe, it would be totally mysteri-

¹⁰In the last decade the conventional wisdom has consolidated around the idea that what speaks in favor of metaethical constructivism, if anything, is its ability to balance a handful of theoretical desiderata. Facing a stalemate between realism and antirealism, constructivism supposedly aims to recover the *objectivity* of moral facts from the prevailing noncognitivism of the mid-20th century, and to do so without running afoul of the *naturalistic* worries associated with critics of traditional (intuitionist) moral realism (e.g. Mackie 1977). What exactly objectivity comes to is a matter of dispute, but something like that there are normative facts, or facts about reasons, irrespective of what anyone in particular thinks; that our normative judgments or judgments about reasons are truth-apt and at least sometimes true; or that genuine cognitive disagreement about normative facts or facts about reasons is possible. Thus constructivism splits the difference, rendering moral facts genuinely objective while naturalistically respectable.

In other words, constructivism offers a way of being a naturalist (which we all want in a post-Mackie world) *and* a cognitivist (which we all want in a post Frege-Geach world), something thought difficult to pull off before the Tanner Lectures that became *Sources of Normativity*. Or at any rate, this, I take it, is the received view of what the problem is for which constructivism is supposed to be a solution.

Enoch (2011a, 324) summarizes this motivation for the Kantian constructivist nicely:

Many people are suspicious about more robust, non-procedural forms of metanormative realism. They think that there are serious metaphysical and epistemological worries (and perhaps others as well) that make such realism highly implausible. Nevertheless, going shamelessly antirealist also has problems. We seem to be rather strongly committed, for instance, to there being correct and incorrect ways of answering moral (and more generally normative) questions, and moreover our moral (and more generally normative) discourse purports to be rather strongly objective. Constructivism may be thought of as a way of securing goods realism (purportedly) delivers, for a more attractive price.

The metaphysical worries that Enoch gestures toward here are associated with ‘Mackie’s problem’. They express the suspicion that there could be entities answering to the traditional realist’s needs. This is *a* problem for which constructivism might provide an answer, but representing the dialectic this way ignores the internal motivation that Korsgaard offers.

In addition to Enoch, See (Tiffany 2006), (Shafer-Landau 2003), (Lenman & Shemmer 2012), and arguably (Smith 1999) and (Gibbard 1999) for this understanding of what motivates Kantian constructivism. In fairness, Korsgaard does cite something like naturalistic scruples as motivation in the preface to *Sources*—“the ethics of autonomy is the only one consistent with the metaphysics of the modern world”—but on my reading of Korsgaard this is not the central question for which constructivism is supposed to be an answer.

ous how they could get a grip on us, how they could address us *as agents*, how they could feature centrally in the exercise of practical reason. Constructivism proposes to explain normative facts in a way that connects them to the individuals they are normative for in the right way.¹¹

The task for constructivism is to account for both reasons for action and our capacity to act for reasons in such a way that their harmony is non-accidental. The demand to understand reasons in the first place arises out of the fact that insofar as we occupy the practical standpoint, we rely on them:

Normative concepts exist because human beings have normative problems. And we have normative problems because we are self-conscious rational animals, capable of reflection about what we ought to believe and to do. (Korsgaard 1996, 46)

As Scanlon (2014, 9) summarizes the worry on Korsgaard's behalf (though he is unpersuaded), "If a consideration's being a reason for a person is just another fact about the world... then the person could still be perfectly indifferent to this fact." The worry is that simply ascribing to certain facts a very special kind of property leaves mysterious why it should appear in our deliberation:

There are certain things that we ought to do and to want simply because they have the normative property that we ought to do or to want them (or perhaps I should say that they ought to be done or to be wanted). The synthesis between the oughtness and the action, or the agent and the oughtness—however that is supposed to go—cannot be explained. It is like a brute fact, except that it is at the same time an a priori and necessary fact. (Korsgaard MS, 2)¹²

The idea at the core of Korsgaard's project is that metaethics will leave us alienated from normativity if it doesn't offer an explanation for its connection to agents. Her

¹¹Bagnoli (2016) makes a similar point in arguing that the 'standard objection' to Kantian constructivism rests on a mistaken understanding of its basic claim to explain the bindingness of reasons in terms of the activity of reasoning.

¹²Cf. (Korsgaard 2003, 112):

If it is just a fact that a certain action would be good, a fact that you might or might not apply in deliberation, then it seems to be an open question whether you should apply it. The model of applied knowledge does not correctly capture the relation between the normative standards to which action is subject and the deliberative process. And moral realism conceives ethics on the model of applied knowledge.

Here Korsgaard follows Rawls (1980), for whom constructivism is explicitly an approach to theorizing normativity that begins with the the difficulty of finding a way to live together—an essentially practical project—rather than with the theoretical investigation of a special kind of truth: "The search for reasonable grounds for reaching agreement rooted in our conception of ourselves and in our relation to society replaces the search for moral truth interpreted as fixed by a prior and independent order of objects and relations, whether natural or divine, an order apart and distinct from how we conceive of ourselves" (519).

Realists like Scanlon and Parfit resist the idea that there is anything to be explained. It simply is the nature of the property of rightness, goodness, oughtness, or being a reason that insofar as we have the capacity for practical reason any bearer of the property is a fitting object for its exercise. As Scanlon (2014, 44) puts it, "it seems to me that no such further explanation of reasons need or can be given: the 'grip' that a consideration that is a reason has on a person for whom it is a reason is just being a reason for him or her."

solution is to center the agent, understood in terms of the reflective capacity to act for reasons, in the explanation of how there can be normative facts.

2.2 The Humean challenge: motivation

Perhaps the most familiar expression of anxiety about normative alienation, though it does not present itself in these terms, is the ‘Humean’ challenge to motivational externalism about reasons. ‘Internal reasons theorists’ hold that it is a necessary condition on something’s being a reason for an agent that it stands in some relation to motivational facts about her. Exactly what relation and exactly what kind of motivational facts vary, but the underlying thought is that if it is not *possible* (for some sense of possibility) for an agent to be motivated by something then it cannot be a reason for her.

Internal reasons theorists do not generally frame their position in terms of avoiding alienation. Insofar as Hume held a view like this it followed from his more basic metaphysical commitments, and in the recent literature internalism is sometimes framed as an analysis of reasons or reasons-talk, where it is part of the very idea of something’s being a reason that it is related to one’s motivations in a certain way.¹³ However, I suspect that the enduring appeal of the position depends at least in part on anxieties (explicit or implicit) about alienation: if there were ‘external reasons’ then they could fail to be motivating, but reasons must be capable of motivating us, so there could not be external reasons. In other words, external reasons, if there were any, would be distant from us in a way that they could not be while still playing the role that we take them to in our lives. Railton (2009, 171) glosses the basic idea similarly, bringing out the dimension of this debate that corresponds to what I’m calling normative alienation:

Absent a link between moral judgment and motivation, ethics might as well be speculative metaphysics. What else could account for the distinctive way in which moral judgments are normative—‘action guiding’—for the agent who makes them?

This interpretation of the impulse underlying the Humean challenge finds support in Williams’s inaugural contribution to the debate (Williams 1979). There he argues against the possibility of external reasons on the basis that if there were any they could not motivate us. He accepts that external reasons correspond to something in ordinary language but denies that there could be any because they would be unfit to play an explanatory role that he thinks reasons must: “If something can be a reason for action, then it could be someone’s reason for acting on a particular occasion, and it would then figure in an explanation of that action” (Williams 1979, 106). That there could not be external reasons because if there were they could not enter into the explanations of agents’ actions is plausibly an expression of an anxiety about normative alienation: if there were any external reasons, they would be (at least sometimes) irrelevant to us, and this cannot be.

¹³On Hume’s metaphysical internalism, see (Schafer 2016). On internalism as an analysis of the concept of a reason, see Finlay (2007; 2009).

Read in the context of Williams's larger body of work this interpretation gains further plausibility. One of Williams's persistent concerns is to vindicate a non-alienated conception of agency. What this amounts to for him is that as agents we are defined by our projects, values, and commitments, in a way allegedly threatened by utilitarian and Kantian moral theory.¹⁴ His work is animated by the conviction that things are going wrong if we conceive morality as the business of some isolable, rational part of the soul, whose task is to discover what reasons there are out there in the universe.

In his iconic "one thought too many" thought experiment he notes that an agent who reasoned that it was permissible to save their drowning spouse over a stranger will have already gone wrong in posing the question, rather than being moved directly by the recognition that it is their own spouse. To think that settling the question of what to do requires transcending the embodied particularity of oneself as an actual agent, in search of facts commanding objectivity or universality, is to lose one's grip on oneself.

At the level of moral theory Williams insists on bringing ethics 'closer' to the agent, preserving an intimate connection between who we are as distinctive agents and what we have reason to do, even if it means opting for a moderate form of moral nihilism. In this connection his denial that there could be reasons that fail to enter into the explanations of agents' actions appears to be part of a larger effort that cuts across the putative distinction between moral theory and metaethics: an effort to make normativity *human*, to restore its connection to us.

It should not be controversial that avoiding alienation by humanizing moral theory is a persistent concern for Williams. I hope that I have made it plausible that he is concerned with a form of alienation not only where he explicitly invokes it as a problem for moral theory but in his moral psychology, that is, that at least for Williams reasons-internalism is a part of his campaign to avoid alienation. This does not prove that the Humean challenge in general is really about avoiding alienation: there may be some internal reasons theorists for whom avoiding normative alienation is at most a welcome but unimportant subsidiary benefit. Nevertheless, the Humean challenge *can* be understood as an expression of an anxiety about alienation, and it is this connection to a deep philosophical impulse, more than technical problems about the analysis of language, that I suspect explains its perennial appeal. Insofar as metaethics is, as I have suggested, in the business of helping us to make sense of ourselves, it makes sense to worry that external reasons, if there were any, would be troublingly disconnected from our lives.

2.3 The epistemic challenge

Probably the least remarked-on guise of normative alienation is its epistemic one. A theory of normativity that vindicated the truth of normative facts but allowed that they were epistemically distant from us would leave us intolerably estranged from them. It is sometimes claimed that normative facts must be knowable for agents *in virtue of being agents*, that there must be a "non-accidental connection between the

¹⁴See (Williams 1973; 1976a;b).

normative truth and our faculties for forming normative beliefs” (Schafer 2015*b*, 709). Less controversially, we need some explanation for the knowability of normative facts in order not to be epistemically alienated from them.

As Thomas Nagel, himself a realist, puts it:

The connection between objectivity and truth is therefore closer in ethics than it is in science. I do not believe that the truth about how we should live could extend radically beyond any capacity we might have to discover it (apart from its dependence on nonevaluative facts we might be unable to discover). (Nagel 1986, 139)

This generates a familiar challenge to traditional realists, namely that they can offer no explanation for why, if the truth about how we should live is simply out there, this knowledge is possible. Like most realists Nagel is content not to offer one, but advocates of agent-centered approaches to metaethics generally—and constructivists in particular—tend to emphasize not only that we should want such an explanation but that there are special obstacles realists face to offering one.¹⁵

In her classic argument against realism and in favor of Humean constructivism Sharon Street (2006), for example, appeals to the knowability of normative facts as something that realism cannot explain. If normative facts were radically mind-independent it would be at best a matter of luck that we were able to track them with our normative judgments. Street relies on the perhaps controversial premise that humans come by our evaluative attitudes largely as a result of evolutionary forces, but the claim can be stated more generally: presumably insofar as we are natural creatures our evaluative attitudes are susceptible to empirical explanation, and such explanation will be independent of the truth of the corresponding normative facts. Thus, realists must be able to explain the relationship between whatever causal forces such empirical explanations invoke (evolutionary psychological or otherwise) and the truth of the relevant normative facts: a challenge that Street argues no realist can meet.

Street’s own view, Humean constructivism, holds that normative facts are determined for each agent by her own normative judgments, and thus are knowable through the activity of making them explicit and bringing them into coherence. Kantian constructivism as well can boast a ready explanation for their knowability for agents as such: that it is the exercise of practical reason that determines them.

Constructivists are not the only ones to press this challenge. Mark Schroeder (2007, 170) notes that ‘irrealists’ of different stripes can easily account for normative knowledge, and that reductivists in particular take this to speak in their favor.¹⁶ Given that realists find it especially difficult to do so, Schroeder notes that “the main divide among realists between reductivists and non-reductivists used to be characterized as the dispute about whether intuitionism is true.” (Schroeder is relying on a different taxonomy of metaethical theories, but in contrast to non-reductive realism, against which he presses a version of the epistemic alienation worry, the views he identifies as incurring no special epistemic burden are agent-centered in my terms.) In other words, the fate of non-reductive realism depends on realists’ ability to defend their rejection of

¹⁵See (Schafer 2015*a*) for a discussion of this point.

¹⁶Compare (Harman 1977).

having to explain the possibility of moral knowledge, over and above merely asserting it. As with the challenge to explain normativity's 'grip' on us as agents, traditional realists tend to respond to the puzzle of how moral knowledge is possible simply by claiming that it is. Or anyway, this is how anti-realists and reductive realists tend to see things.

For those who find it mysterious or even occult that we should see normative facts as essentially knowable, without further explanation, concerns about moral epistemology put pressure on approaches to metaethical theory that do not center the agent as the bearer of practical knowledge. To accept a traditional realist account of the explanation of normative facts while remaining skeptical of the realist's non-explanation of their knowability would leave one in a state of epistemic alienation, convinced that there were normative facts but with no way of discovering what they were.

2.4 The solution: agent-centered metaethics

These classic objections to realism reflect related anxieties: anxieties about the possibility that we could have reasons to which we were motivationally indifferent, reasons whose relevance to our activity of reflective self-determination was at best coincidental, or reasons of which we could in principle be systematically unaware. If it were possible for reasons to be like that, they would be totally estranged from us. These more local challenges to traditional moral realism are thus expressions of a sense that morality cannot be alien to us, and that a theory of normativity must come along with an explanation of how it can be ours.

The threat of normative alienation calls for a theory of normativity that brings it closer to us, intermingling it with the messy, embodied, and perhaps contingent features of human life with which we each individually have the most direct familiarity. The resulting proposals all center the individual agent in their derivation of normativity, emphasizing desires, values, preferences, or the embodied capacity to practically self-determine, as in some sense foundational to the explanation of how there could be such a thing as normativity at all. In the next section, however, we will see that in bringing normativity closer to ourselves we risk losing our moral grip on one another.

3 Social Alienation

In the last section I argued that several familiar challenges to traditional metaethical realism can be understood as expressions of a more general underlying anxiety, an anxiety about the possibility that morality could be alien to us. A theory of normativity that failed to grapple with this fact would fail to capture something important about the experience of being a moral agent. Though not everyone is moved by all or even any of these challenges I take it that I can help myself at least to their plausibility.

In this section, however, I will raise a different kind of challenge, one that reflects a different kind of anxiety: that moral theory might represent us to ourselves as estranged from one another. Corresponding to this anxiety is the second desideratum for a theory of normativity: to explain how it can be that we are morally related to

concrete others, and thus to avoid what I call *social alienation*.

This desideratum has gone largely unrecognized and is difficult to formulate using ready-to-hand conceptual resources.¹⁷ As a first pass, the challenge is to ground the essential sociality of morality.¹⁸ Much of morality involves responding to the grip we have on each other. Agent-centered theories run the risk of erasing this distinctive grip, making agency out to be a matter of individuals following rules, recognizing reasons, or standing in relation to themselves (the relation of reflective distance, for example). This give us at best an indirect way of recognizing other people. To begin to bring this worry into view, I return to Williams.

3.1 Alienation in 20th century moral theory

A persistent theme in Williams's work is that ethics must account for the ways that we are shaped as distinctive agents by our projects, commitments, and values. To the extent that moral theory alienates us from these parts of our lives, it presents an image of the moral agent in which we cannot recognize ourselves. However, while the examples that Williams uses to motivate his objections typically feature important social relationships, his diagnosis of alienation interiorizes the problem, making it an individual, psychological defect, and not a social one.

Utilitarianism, for example, is a threat to an agent's integrity because "it can make only the most superficial sense of human desire and action" (Williams 1973, 82), and it "alienates one from one's moral feelings" (104). What goes wrong in the 'one thought too many' case is that the husband appeals to an explicit deontic order, thinking a judgment about duty or rules is a necessary intermediary between his affection and how he ought to act. Moral theory, he worries, "treat[s] persons in abstraction from character" (Williams 1976a, 19), making us out to be nothing more than a "locus of causal intervention in the world" (Williams 1973, 96). The 'one thought too many' case highlights a disconnect between moral theory and human life, realized in an agent's (in)ability to properly understand himself in relation to another. However, Williams's understanding of alienation and integrity points toward achieving internal, psychological unity (something like virtue) as the solution.

The contrast comes out more clearly in the work of two contemporary critics of alienation in moral theory: Michael Stocker and Peter Railton. Michael Stocker's (1976) central case involves someone explaining their choice to visit a friend in the hospital by appealing to the duties of camaraderie, and Peter Railton (1984) responds to a worry about someone regarding their spouse as a mere source of utility. For Stocker,

¹⁷The concern has gone largely unrecognized, but not entirely. Aside from Iris Murdoch, who I discuss in what follows, some others who I think are onto something like this worry include Kate Manne (2013; 2017), Michael Thompson (2004; 2012), Kenneth Walden (2012; 2017; 2018; 2020), and Kieran Setiya (forthcoming). As in the previous section none of my antecedents have explicitly identified social alienation as something to be avoided, but I think their interventions can be profitably understood, along the same lines as mine, as taking the sociality of morality seriously in a way that has metaethical implications.

¹⁸Social alienation is a problem for morality specifically. It may turn out in the end that the best theory of normativity implies that all normativity is social (cf. Brandom 1994, which argues that the normativity of meaning is social). But it is not a demand on a theory of normativity that it explain the sociality of all normativity, only that it explain normativity in general in a way that doesn't rule out the essential sociality of morality.

‘moral schizophrenia’ consists in a disunity of one’s motivations and values. “One mark of a good life,” he claims, “is a harmony between one’s motives and one’s reasons, values, justifications” (1976, 453). If moral theory is to help us understand what it is to live a good life, it must be able to make sense of how such harmony is possible. For Railton alienation involves our affective selves coming apart from our rational, deliberative selves: “there would seem to be an estrangement between [an agent’s] affections and their rational, deliberative selves; an abstract and universalizing point of view mediates their responses to others and to their own sentiments” (Railton 1984, 137). Both critiques are motivated by noting a defective form of sociality, allegedly due to adopting an alienating moral theory, and both diagnoses identify psychological disunity as the problem, and psychological unity as the solution.

Unlike Williams and Stocker, Railton hints at something like the problem of social alienation as I conceive of it—estrangement between oneself and another—as an equally important dimension along which moral theory can be alienating, and one from which the psychological is not cleanly separable. He notes that “we should not think of John’s alienation from his affections and his alienation from Anne as wholly independent phenomena, the one the cause of the other” (Railton 1984, 138).

In setting out a blueprint for responding to the problem of alienation he highlights the role that relationships with others must be allowed to play:

First, we must somehow give an account of practical reasoning that does not merely multiply points of view and divide the self—a more unified account is needed. Second, we must recognize that loving relationships, friendships, group loyalties, and spontaneous actions are among the most important contributors to whatever it is that makes life worthwhile; any moral theory deserving serious consideration must itself give them serious consideration. (Railton 1984, 139)

He cautions against “the picture of a hypothetical, presocial individual” by which philosophers have become distracted, leading to the (un-argued for) assumption that self-concern is natural and requires no explanation of the kind that concern for others is taken to require (168). He points toward the need for a solution that captures the importance of “participation in certain sorts of social relations—in fact, relations in which various kinds of alienation have been minimized” (147), and insists that the starting point for ethics must be the “situated rather than pre-social individual” (171).

However, Railton ultimately leaves the problem under-theorized. If there is a social dimension to these cases that has been mostly ignored, what demand does it place on the theorist? Here I only have the space to offer a sketch of an view that I elaborate elsewhere.¹⁹ The key upshot is that avoiding social alienation—achieving social integrity, to re-purpose Williams’s distinction—requires that in our ethical self-awareness we account for the significance for us as agents of others as *external*, as *particular*, and as *subjects*—as each an individual reality, separate from oneself.²⁰ We

¹⁹See [redacted].

²⁰ We can see the distinction more clearly by reflecting on an analogous puzzle about the epistemology of perception: that concerning how we can have perceptual experience of the world itself, and not merely of our inner representations of it. Not everyone agrees that this is something to be achieved, but those that are concerned with the threat of being trapped behind the ‘veil of ideas’ (perceptual alienation from the

must be able to make sense of ourselves, that is, as responsive to others themselves, not just to rules for conduct that make reference to others in their application conditions; to particular others, not just to abstract idealizations of others as representative rational agents, persons, and so on; and to others as subjects, and thus potentially responsive to us. I refer to a form of moral self-awareness that meets these conditions as the achievement of ‘practical openness to the other.’²¹ My practical openness to another isn’t separable from the other’s being practically open to me. If it were we would each only be open to one another as to a third person, one we each see as bearing a special normative property rather than as standing in relation to ourselves.

Integrity, for Williams, is a matter of an agent’s moral thought and action staying close to everything else that makes her her. Social integrity, as I’ve been sketching it, is a matter of one’s moral thought and action reflecting mutual practical openness to others. If socially alienated moral knowledge is the mere apprehension of one’s reasons or the rules by which one is bound, socially integrated moral knowledge is an awareness of others as such. The threat of social alienation in ethics is of a kind of normative solipsism. To avoid social alienation is to account for what Iris Murdoch characterizes as, “the extremely difficult realisation that something other than oneself is real” (Murdoch 1971, page).

The phenomenon of practical openness to the other is in my view tragically under-theorized, and this is not the place to attempt a project of that scope. With a hazy idea of the problem in view, in order to give a sense of the stakes I’ll offer an example of how it manifests in a set of issues in normative ethics: the phenomenon of ‘directedness’. Recognizing another as the object of a directed obligation is a case of practical openness to another, and one a proper understanding of which is threatened by agent-centered metaethics.

3.2 Directedness in ethics

An obligation is ‘directed’ when it is owed to someone in particular. Perhaps we are all obligated to give to charity, but we do not owe it to any particular charity to give to them. We are also obligated to keep our promises, but in each case we owe it to the promisee. Directed obligations are generally thought to correlate with or be identical to claim-rights, so another way to put the point would be that no particular charity has a claim on our beneficence, but each time we make a promise we grant to the recipient a claim to our performance. When we violate a directed obligation we do not merely do something wrong but *wrong* someone in particular: the one to whom the obligation is owed. The one who is wronged is thus in an important sense the *victim*, not merely the *occasion* of wrongdoing.²²

world) tend to emphasize both externality and particularity as important features of worldly objects qua worldly. See (Brewer 1999, Martin 2006, Travis 2004, McDowell 1983).

The phrase ‘individual reality, separate from oneself’ is a patchwork of two different phrases Murdoch (1971) uses: her gloss on Simone Weil’s concept of attention as a “just and loving gaze directed upon an individual reality” (33) and her characterization of the object of moral awareness as “a reality separate from ourselves” (46).

²¹To paraphrase John McDowell’s (1994) slogan that avoiding what I called perceptual alienation requires epistemic ‘openness to the world’.

²²This way to refer to the distinction is due to (Thompson 2004, 340).

Directed obligations constitute the core of morality. They reflect what Wallace (2019) calls the ‘moral nexus’ that joins concrete persons, equally real. Being aware of and responsive to standing to others in a moral nexus is an important way, if not *the* fundamental way, of being practically open to one another. The moral nexus is a basic social relation that arguably cannot be explained in terms of reasons, rules, and putatively more normatively fundamental self-relations. A metaethics without the resources to capture the moral nexus risks theorizing away the sociality of morality.

One way a metaethics might run this risk would be by purporting to directly entail a normative ethical theory with no room for directed obligations at all (say, act-consequentialism). More subtly, a metaethics might entail that directed obligations are not really directed. Along these lines, Aleksy Tarasenko-Struc argues that Korsgaard is committed to the view that obligations apparently owed to others are in fact owed to ourselves. Because Korsgaard grounds all normative authority in the constitutive ability of agents to bind themselves, he argues, all obligations are ultimately grounded in this self-relation: “The problem is that she embraces an egocentric conception of authority, on which we originally have the authority to obligate ourselves whereas others only have the authority to obligate us because we grant it to them” (Tarasenko-Struc 2019, 1). There will always be an unbridgeable explanatory gap between obligations to oneself and those apparently owed to another.

From the fact that Korsgaard grounds obligations to others in obligations to oneself it does not obviously follow that obligations to others are illusory. They would be derivative, but a derivative obligation may bind all the same. The worry is that it may not bind *in the right way*—that is, that an obligation that derives ultimately from the individual requirements of self-constitution will turn out not to be a genuine instance of being bound by another, but only appear so. The explanatory challenge for Korsgaard is to explain how an obligation that derives from an obligation to oneself will not turn out, on careful inspection, to be merely an obligation *to* oneself that *concerns* another, depending on how the derivation is fleshed out.

If I promise myself I’ll smile at strangers more, my promise concerns a stranger walking by (whose presence triggers a smile), but is not owed to them: if I am feeling grouchy, I do not wrong the stranger by failing to smile (except perhaps by rudeness). I act wrongly *vis-à-vis my promise to myself* in my conduct concerning the stranger. An obligation genuinely owed to another is not like this. It is an opportunity to do right by another or to wrong them, not just to do right or wrong.

One way Korsgaard might try to get around this problem is to hold that our authority over ourselves can be transmitted to others. On this view, I can have directed obligations to others because other people can exercise the power, which I have transmitted to them, to bind my agency.²³ In other words, rather than exercising my ability to obligate myself by binding myself to do something concerning another (smiling at strangers), I could somehow *transfer* that authority to another, to be exercised by them, thereby obligating me.

It’s not clear that the idea of such a voluntary transfer of authority can work. The trouble isn’t that authority can never be genuinely transferred: if one party with

²³See (Korsgaard 2009, 189–91), cf. (Tarasenko-Struc 2019, 85–87). Another strategy Korsgaard could pursue, but appears not to, would be to invoke the distinction between the content and justification of a norm, like promise-keeping. I address this approach as a general matter below—see §3.3.

authority—say, the president—appoints an official to oversee the activity of a third party, the third party will for all practical purposes answer directly to the official. One could argue that there remains a sense in which the third party ultimately is obligated only to the president, with an official as a normative intermediary, but there is surely a recognizable sense in which the official's orders obligate the third party directly.²⁴ However, if the president appoints an official to oversee *himself*, on the authority of his own office, he can only ever *appear* to obey the official, for the moment the official issues an apparent command the president does not wish to follow, he can simply withdraw the grant of authority, proving the transfer to have been illusory all along.²⁵

What is needed in the special case of voluntarily transferring one's own authority over oneself to another is some way to ensure that, once transferred, the authority cannot be voluntarily withdrawn. If we model the transfer of authority on the idea of a promise to oneself to obey another we won't get that, since it is characteristic of promises that the promisee has the ability to release the promisor (this is why the idea of a promise to oneself is suspicious to begin with). But if we can find a different model on which a power to obligate oneself can be transferred, such that when done it cannot be voluntarily undone, we will still have to confront the worry that whatever it is that prevents it from being withdrawn will require an independent source of authority, one that finds no place in Korsgaard's theory.

Supposing, however, that a genuine, voluntary transfer of authority is possible on Korsgaard's account, it will leave us with an unsatisfying asymmetry: that others have only as much authority over us as we grant them is not much of an improvement over having obligations concerning others but owed to oneself. As Tarasenko-Struc concludes, with an analogy to the classic 'problem of other minds,' "just as a person's wince might be thought to directly reveal that she is in pain, the fact of her pain may likewise be thought to directly make a claim on us, where the validity of this claim in no way depends on our having validated it or on our having granted her the authority to make claims on us more broadly" (88).

Tarasenko-Struc does not—and I do not mean to—assume that if the ultimate ground of a duty is a fact about an agent (rather than another subject), then that duty cannot be genuinely directed at another subject. The heart of the argument is that if the explanatory ground of a theory of obligations is a self-relation, more must be said about how a self-relation can generate a self-standing self-other-relation. Korsgaard's own strategy is not promising. It does not follow that the trick cannot be accomplished, but working through the Tarasenko-Struc's argument can provide a vivid example of how

²⁴Tarasenko-Struc (85-86) makes a similar distinction, between *discretionary* and *original* authority, and notes that for discretionary authority to be genuine authority it must presuppose a prior grant of original authority, which again Korsgaard cannot explain. The following argument runs parallel to his, though in slightly different terms.

While the official's orders plausibly obligate the third party directly it does not follow that the third party owes performance to the official—see the discussion of the example of private law enforcement below in §3.3—but my point here is that even if we assume that in a trilateral case we get something approximating genuine transfer of authority we face a special difficulty where the original source of authority is the one putatively obligated by that same authority once transferred.

²⁵Compare *United States v. Nixon*, 418 U.S. 683, 706–07 (1974), holding that the president cannot be permitted to determine the extent of his own executive privilege vis-à-vis a special prosecutor, at the risk of collapsing a limited privilege into an absolute immunity.

things can go wrong with accounting for the sociality of morality—how metaethics can lend itself to a form of social alienation. It can at least bring into view the shape of the problem, and put some pressure on agent-centered theorists of normativity to say more about how the self-other gap can be bridged.

Importantly, the problem is generated by the Kantian constructivist theory of normativity: the explanation the Kantian provides for the truth-aptness of normative facts entails that those facts have a certain structure. They are ultimately facts about how we stand in relation to ourselves, and not about how we stand with respect to others. Other forms of agent-centered metaethics run a similar risk, if not a greater one: if normative facts are ultimately explained in terms of agents' desires or other psychological states it is even more difficult to see how to recover the status of the other as the one who stands to be wronged.

One way to put the general worry is that the *reason* relation that forms the basis of normativity has argument places for the fact (or consideration) that is a reason, the agent for whom it is a reason, the action it is a reason to do, and perhaps the context, but not for the other, the one to whom a directed obligation is owed. The other may have a corresponding reason for a reactive attitude associated with being wronged, and thus the directedness of the reason would be at least partly accounted for as a psychological correspondence.²⁶ But to account for directedness in terms of merely corresponding reasons is to hold obligors and obligees at a normative distance from one another: the difference between having a reason to ϕ and owing it to someone in particular to ϕ is not that the other happens to have a specific attitude, but that one thereby stands to the other as witnesses to the same relational fact. The rights correlative to duties do not just happen to line up with them; they are inextricably linked. They are different perspectives on the same moral nexus between persons—indeed they are often claimed to be the very same fact expressed in two different ways.²⁷

We might try to accommodate this feature of directed obligations by putting the duty or right in the 'fact' argument place: [that A owes it to B to ϕ] is a reason for A to ϕ , and the very same fact is also a reason for B to (e.g.) resent A if A does not ϕ , and the same pair of reasons could be described in terms of the fact [that B has a claim-right against A that A ϕ], which is after all the same fact. This will only push the problem back a step, however; A 's reason to ϕ and B 's reason to resent A if A does not ϕ will be constituted by a common fact (a fact about A 's duty i.e. B 's right), but A will not be normatively related to B in virtue of having this reason, in which B only features as part of the content (like a movie features in my prudential reason to see it—more on this example below in §3.5), rather than as a normative relatum. Metaethics must do more than generate the reasons associated with directedness if it is to fully vindicate the importance of recognizing another as standing to one in a relation of right.

For one person to owe a directed obligation to another is for them to recognize

²⁶Darwall (2012; 2006) uses reactive attitudes and the standing to hold them to explain directedness, but it is not clear whether he is in fact *reducing* directedness to this correspondence. He claims that the concepts of authority, accountability, obligation, and the second person, as well as of attitudes like blame and the reasons or standing to hold them, come together in a circle. He is thus not reducing relational concepts like obligation to monadic or psychological concepts like attitudes, but using all of them to explicate the others.

²⁷E.g. Gilbert (2004) and Wallace (2019).

the other as the bearer of a claim against them, which is to recognize the other as recognizing them as owing a directed obligation, and so on. Contained within the self-consciousness that one stands in a juridical relation of this kind with another is at least the implicit recognition of the other as recognizing oneself. (Of course some bearers of rights and obligations are unaware, so it does not follow from one person's having a right against another that the other is similarly self-conscious, but the logic of directed obligations involves at least unrealized mutual recognition.) This is what mutual practical openness comes to in the realm of rights, and it is what metaethics needs to explain at the risk of leaving us socially alienated.

3.3 Two-level accounts of directedness

One strategy available to Korsgaard or, for that matter, any other agent-centered metaethicist, would be to invoke the distinction between the content and justification of a norm like promise-keeping. Thus the fact that *A* owes it to *B* to keep her promise can be explained by the role that promise plays in, for example, the integrity of *A*'s agency *so long as the promise itself is an entity partially constituted by B*. The content of a norm (that a promise is directed at *B*) and the justification of that norm (that you need to follow it to successfully constitute yourself as an agent) operate at different levels.²⁸ This kind of two-level theory, often associated with contractualism or rule-utilitarianism,²⁹ is usually criticized on the grounds that higher-level theories that generate rules without directedness built in get the extensions wrong, failing to reliably pick out correlative rights holders, or that in the particulars they fail to actually *explain* the correlativity of rights and duties altogether.³⁰

But there are reasons to worry that in principle no such theory of directedness can succeed in vindicating it. Here the question is not about evaluating an action recommended by a practice (as in the original Rawls argument), but a relation of authority putatively established by it. But because authority is a higher-order moral concept rules and practices are transparent when it comes to authority in a way they are not when it comes to reasons for action: it is one thing to say that a rule or practice can create reasons, and another altogether to say that a rule or practice can establish basic, and not merely conventional, relations of authority and accountability.

Two-level accounts can create a fiction in which the rule has a structure the underlying normative theory lacks, but to see whether it is more than a fiction we need to look at whether the underlying normative theory can make sense of the structure. Suppose that the lawmakers of a legitimate political authority delegate enforcement power to a private party well-positioned to track malfeasance—say, Google, with its immense surveillance apparatus. (And suppose—however implausible—that the legislators are right to do so, perhaps because it is an important issue and the state can't deal with it alone, and the procedure through which Google will enforce the law doesn't violate any civil liberties.) When a Google auditor knocks on your door to ask you a few questions, you may be obligated to answer, even morally obligated, and within the fiction established by the law you may have to act as if you owe this

²⁸Cf. Rawls 1955.

²⁹E.g. Scanlon 1998 and Hooker 2011, respectively.

³⁰See e.g. Wenar 2013, Gilbert 2004, Woods 2016.

duty to the Google auditor. But if you refuse, you may not wrong the auditor, or Google itself—you may wrong the state, or your fellow citizens, or perhaps no one at all. Figuring out the party to whom you truly owed the obligation (if any) requires going outside of the convention to see how the relevant authority (political, moral, legal) works and under what conditions (if any) it can be legitimately transferred to a third party.³¹

So, for example, in the promissory case I discussed above part of what it means to say that the promissory obligation is directed at *B* is to say that *B* is the bearer of not only the correlative claim right but the *power* of waiver. It may be that we take ourselves to be bound by rules that by convention stipulate some other person as the obligee, but that does not establish a genuine transfer of authority over our actions. In order to see whether on a given theory this is possible we need to ‘pierce the veil’ of the convention and see whether the underlying account of authority is compatible with transferring it, or only with agreeing to act as if we have. That is precisely the move that Tarsenko-Struc targets under the guise of a transmission-of-authority principle, as I have just reviewed.

3.4 Social alienation and agent-centered metaethics

The agent-centered metaethical theories that we saw provide the resources to answer the challenge of normative alienation face special difficulties in accounting for the sociality of morality. These views explain moral facts starting with attitudes or capacities indexed to the individual, or from the first-person perspective. They thus come along with certain commitments about the kinds of facts moral theory can rely on: principally, facts about individual agents, or facts about oneself. Insofar as they aim to capture the sociality of morality, in the sense I’ve been discussing here, they are in the position of trying to reconstruct relational facts out of individual-agent facts, and it is not clear that this can be done. They may be able to recover the reasons associated with directed obligations, but if they do so by making such reasons out to be psychological facts about individual agents, or explained in terms of self-relations rather than social relations, that will not be enough.

The Kantian, for example, begins with facts about the nature of agency as such. Then, in attempting to derive substantive moral facts, she has to somehow generate facts of the right kind. That is, she has to generate facts suitable to bring others into view in the right way and explain the moral nexus that (for example) joins bearers of correlative rights and duties. While my discussion of social alienation in moral theory is in some important respects heterodox, under some description this is an aim that Korsgaard herself endorses. She holds that there is a role for sociality in the characteristic exercise of agency: reflecting on essentially public reasons, or responding to the call of another. Even on her own terms it’s not clear that her conception of agency is up to the task of grounding the sociality she appeals to in lecture 4.2 of *Sources*, a discussion which appears out of nowhere. More broadly, the sense in which agency is social for Korsgaard is, so to speak, inside-out. What it is to be an agent is such as to potentially stand in recognitive relations with others, if any others show up:

³¹For a longer discussion of what is essentially the same point in a legal context, see (Murphy 2020).

the reflective relation that one stands in to oneself as an agent (the ‘second-person within’, as she puts it in (Korsgaard 2007)) is generalizable. By her own lights, then, relations to others are not built into agency. What the above discussion of directedness suggests is that Korsgaard’s theory is inadequate to vindicate the irreducible sociality of morality, and it is this structural feature of her theory that I suspect explains why. There is widespread skepticism regarding Kantian constructivism’s ability to make good on its explanatory ambitions, and the gap between its agent-centered explanatory structure and the sociality of morality provides a compelling diagnosis. Explaining sociality in morality is a desideratum that at least some agent-centered approaches to metaethics recognize, and they are not set up to have a natural way of doing so.

3.5 The solution: other-centered metaethics

The demand to appreciate the significance of others as external and particular, and as subjects themselves is realized in the demand to fully appreciate the directedness of certain moral requirements. There is an important sense in which at least some of the time what morality consists in is not recognizing oneself as having a reason or bound by a law but recognizing and responding to the other qua other.

In one sense the upshot of this discussion is somewhat trivial: moral facts are, at least some of the time, facts about particular others, and the relations we stand in to them. But what I have been trying to bring out is that this is not just a matter of the content of normative facts, but of their form. The other must show up in practical thought *in the right way*. Consider the reason I have to see a movie I’m likely to enjoy. The movie shows up in an account of what I have reason to do. But when I reflect on the reason I have to respect the bodily autonomy of the person sitting next to me on the bus, she appears in my practical thought in a different way from the movie I’m likely to enjoy, or she ought to if I am fully appreciating her as an individual reality.³²

Social alienation is thus a problem for metaethics insofar as it is in part concerned with how normativity works, about its structure, and further insofar as many theories of normativity seem committed to ruling out any way for us to play the right sort of role in the normative lives of one another. While constructivist, subjectivist, relativist, and

³²In something like the way that there is a formal difference between the way a *de re* thought relates to a referent and the way a *de dicto* thought relates to the same one, perhaps we should say that my thought of an other qua other relates me to her in a way that my thought of a movie qua potential source of pleasure does not. Some philosophers have sought to capture this distinction by insisting on the importance of *second-personal* thought in ethics (most famously probably Darwall (2006)), and though I quibble with the assimilation of this difference to one of grammatical person I am inclined to endorse something like this line. See also Zylberman (2014) and Haase (2014) for attempts to push the discussion of the second person in a direction similar to the one I’m trying to go here. The discussion of the second person that gets the closest to what I’m after appears in Moran’s characterization of the relationship between parties to successful communication:

The relevant incorporation of another perspective on one’s act and including that in one’s own understanding of it is not the same thing as taking an “outside” perspective on what one is doing, something that each of the parties could do separately. The speaker does not imagine a third-person perspective on her act but rather a second-person one, that of her addressee; in adjusting her performance to this perspective she is not speaking so as to be overheard by an observer, but rather inhabiting the perspective of a *shared* participant in a practice, *the shared consciousness of what they are doing together*. (Moran 2018, 144, *emph. added*)

other agent-centered approaches to metaethics can claim some success in addressing normative alienation it is more traditional forms of realism that are better-positioned to provide the resources for addressing social alienation.

Existing realist metaethical theories may not be able to accommodate irreducible directedness without substantial revisions. As we saw above a ‘reasons-first’ realism of the kind associated with Parfit and Scanlon runs into the difficulty that the reason relation lacks an argument place required to account for the other person that stands to one in the relation of duty and right, and thus risks theorizing away the relationality of directed obligations.³³ An emphasis on one species or another of normative facts—facts about fittingness or value or the good—leaves on similarly ill-equipped to make out the fundamentality of the moral nexus that joins an agent and the other. Such theories deliver impersonal facts about the world that feature in specifying an agent’s relation to possible actions, attitudes, or aims, but are not obviously relevant to an agent’s relation to another.

But the basic realist strategy of taking whatever normative ethics delivers and promising to vindicate it by augmenting the ontological inventory (or, in more quietist flavors, but granting the legitimacy of a certain quasi-metaphysical discourse), is in principle perfectly consistent with taking directed obligations and the moral nexus they saturate is a primitive feature of reality. Whatever discourse of duties, rights, and sociality emerges the realist can simply affirm as a description of how things really are. If that means positing a new kind of metaphysical relation, so be it.³⁴

That moral thought is at least sometimes thought of another, and that this difference is more than one of merely which singular terms appear in a reason-stating sentence, suggests that a metaethics adequate to capture the sociality of morality will be somehow *other-centered*. The explanation for how we come to have moral reasons will have to revolve around other creatures, how things are with them, and how they stand with respect to us. This is no real challenge for traditional realists, who can accommodate any constraint on what the normative facts must be like by saying of those facts, “yes, and they are simply true, no further explanation required.” But as we will see in the conclusion, agent-centered approaches to metaethics struggle to meet the same standard, and thus to address the threat of social alienation.

³³Nagel (1970) is an interesting case of a realist who comes close to explicitly setting for himself a goal like what I describe as avoiding social alienation—what he calls “practical solipsism”—but his focus is on recovering motivation and normative grip, rather than on explaining how his view can accommodate anything like directedness in particular or irreducible sociality in general. In other words, the challenge he sets for himself is to address normative alienation so he offers little by way of directly accounting for social alienation. Given that his metaethics is reasons-first and his primary route to avoiding practical solipsism is through publicity, rather than anything in the neighborhood of practical openness to the other, he is more in Korsgaard’s position than the generic ‘realist’ I am imagining here, who faces the opposite problem. (Perhaps this should not be surprising, as he, like Korsgaard, associates his view with Kant.)

³⁴This suggestion is not meant to be dismissive. As with the analogous problem of perceptual alienation I allude to in fn.20, where the direct realist answer is to simply insist that when we open our eyes in a well-lit room it is the objects in it that we see (i.e., to which we are perceptually related) without positing any mediating representations, I think it is in perfectly good order to insist that the self-other relations disclosed through practical openness to the other are just as real as anything else. The limitation of quietist realism is, as far as I’m concerned, that the agent-centered approaches are right to worry about normative alienation; it is not metaphysical scruples that pull me in their direction, but a dissatisfaction with an unexplained connection between the other so disclosed and the self as open to them.

Conclusion

Avoiding normative alienation urges making some concession toward agent-centered approaches to explaining normativity. But any explanation of what reasons an agent has that derives them from facts about her will risk having started in the wrong place to ever bring the other into view as an individual reality. To start with an individualistic account of the source of normativity and wind up with a full-throated vindication of normative facts as facts about concrete others appears to involve crossing a gap. Theories of normativity that define themselves by the task of accounting for the significance of the other-qua-other, however, risk having started in the wrong place to ever bring the resulting normativity close enough to the individual agent to avoid the threat of normative alienation.

The attempt to reckon with normative alienation pulls in the direction of agent-centered metaethics (typically though not exclusively irrealist, broadly construed), while the attempt to reckon with social alienation pulls in the direction of other-centered metaethics (typically nonnaturalist realism).³⁵ It is difficult for a theory of normativity to avoid both normative alienation *and* social alienation, but not impossible.

Supposing that a satisfyingly non-alienated theory of normativity must be in some sense agent-centered *and* other-centered, it won't do simply to impose the conjunction of the two constraints. There is at least a superficial tradeoff, in that, to take the metaphor a bit literally, the theory can have one center or the other, but not both. Working out how these constraints can co-exist involves getting clearer on what it would mean to "center" the agent or the other in a theory of normativity—something that up until now I have expressed largely by example. What is the sense in which Humeans "center" the agent as a bearer of desires or values, or that Kantians "center" the agent as a bearer of the capacity for practical reason, in their explanation of how there can be normative facts?

It is tempting to reach for metaphysical notions like 'grounding' and 'fundamentality', but in this case I think their use obscures more than it reveals. Yes, desires are explanatorily fundamental for the Humean, and the capacity for practical reason grounds normativity for the Kantian. But nothing in this metaphysical gloss entails that normativity can't have more than one partial ground, that more than one thing cannot be fundamental. Yet it remains unclear how one's own desires *and* the individual reality of another could be at once fundamental to the explanation of a given moral fact, other than by stipulation. What is needed is not the mere conjunction but a synthesis, a self-other relation wherein the other-qua-other is invoked in an understanding of what it is for the self to be a self.

These remarks are programmatic at best, but rather than attempting to develop them in any detail at this late stage of the argument I'd like to close by considering a couple of positive proposals for how this could be done, coming, respectively, from either direction. First, from agency to sociality.

³⁵Strictly speaking it may be that the threat of social alienation is better understood as pulling in the direction of other-centered *ethics*, but that other-centered ethics is hard to square with agent-centered metaethics, and rather easy to ground in nonnaturalist realist metaethics.

I've used Kantian constructivism as a stalking horse throughout this paper, largely because there is so much that it gets right. Korsgaard in particular begins with the insight (not original to her, but one that she centers in her own story) that even if we could make sense of the 'queer' entities Mackie has long been taken to cast doubt upon, their mere existence would not be enough, unless we had some explanation of how they could get a grip on us. Further, she takes on board more or less the social aims I've argued are necessary.

In my view she does not have the explanatory resources to reach them. She begins with an individualistic conception of agency, one articulated in terms of an individual agent's capacities, capacities in turn understood through the form of law. Laws, on this picture, are universal generalities. In applying a law to oneself, one arrives at an instance: if we all ought to ϕ , then *I* ought to ϕ . Where is the other in this picture? The generality of a law hints at the logical possibility of another, but the law would still be a law if I were the only one around for it to bind. Korsgaard begins with this individualistic conception of agency and attempts to derive a picture of morality that has a deep social structure, in which we are responsive to the calls of others, who simply by speaking reshape the normative space in which we deliberate. This project is generally regarded as a failure.

The solution, it seems to me, or at least a solution, would be to build sociality into the story at the ground level: agency. Conveniently, for those of us who look to the history of philosophy to discern the movement of ideas, as Korsgaard clearly does, this suggestion has already been articulated by Kant's own successors in the tradition of German idealism: Fichte and Hegel. Both argue, in different ways, that self-consciousness—which marks the distinction between animal locomotion and rational action—depends on standing in relations of mutual recognition with other self-conscious creatures. Such a view is independently motivated, in ways I do not have the space to consider here, but for present purposes the appeal is that it has the potential to fund a constructivist theory of normativity that could both explain the grip reasons have on agents and the grip agents have on one another.

What about the other direction? The way to address normative and social alienation beginning with other-centered realism and recovering the connection between normativity and individual agents, I want to suggest, is by taking a cue from Iris Murdoch. I argue elsewhere³⁶ that we can read Murdoch as looking for a way of locating normativity in the world—in particular in historically-conditioned social relations between concrete individuals—rather than in the attitudes or choices of the agent, while at the same time holding that getting oneself in a position to be responsive to it is itself an achievement of agency.

Murdoch's is in some ways the paradigm of what I've called a other-centered metaethics, in that, as I noted above, for Murdoch the key element in morality is seeing others clearly, escaping fantasy and self-focus and getting directly in touch with the individual reality of others. However, I argue that for Murdoch it is equally important to emphasize that the development of a distinctive practical standpoint on the world is something that we continuously and actively cultivate and revise, and

³⁶See [redacted].

is thus in an important sense the realization of individual agency. That moral self-awareness is, for Murdoch, awareness of how one stands with respect to concrete other persons addresses social alienation, and that arriving at this form of self-awareness is something we struggle to do explains what the reality of others has to do with us, thereby addressing normative alienation.

Whether through the Hegelian strategy, the Murdochian strategy, some combination of the two, or some other approach altogether, metaethics has its work cut out for us in capturing the sociality of morality and its connection to individual agents.

References

- Bagnoli, Carla (2016), “Kantian Constructivism and the Moral Problem.” *Philosophia*, vol. 44, no. 4:pp. 1229–1246.
- Bedke, Matthew S. (2020), “A Dilemma for Non-Naturalists: Irrationality or Immorality?” *Philosophical Studies*, vol. 177, no. 4:pp. 1027–1042.
- Blanchard, Joshua (2020), “Moral Realism and Philosophical Angst.” In *Oxford Studies in Metaethics* vol. XV, Russ Shafer-Landau, ed.
- Boyd, Richard (1988), “How to Be a Moral Realist.” In *Essays on Moral Realism*, Geoffrey Sayre-McCord, ed., pp. 181–228, Cornell University Press.
- Brandom, Robert B. (1994), *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Brewer, Bill (1999), *Perception and Reason*. Oxford University Press.
- Brink, David O. (1986), “Externalist Moral Realism.” *Southern Journal of Philosophy*, vol. 24, no. S1:pp. 23–41.
- Darwall, Stephen (2006), *The Second-Person Standpoint: Morality, Respect, and Accountability*. Harvard University Press.
- (2012), “Bipolar Obligation.” *Oxford Studies in Metaethics* vol. VII, pp. 333–.
- Dasgupta, Shamik (2017), “Normative Non-Naturalism and the Problem of Authority.” *Proceedings of the Aristotelian Society*, vol. 117, no. 3:pp. 297–319.
- Enoch, David (2011a), “Can There Be a Global, Interesting, Coherent Constructivism About Practical Reason?” *Philosophical Explorations*, vol. 12, no. 3:pp. 319–339.
- (2011b), *Taking Morality Seriously: A Defense of Robust Realism*. Oxford University Press.
- Finlay, Stephen (2007), “Responding to Normativity.” In *Oxford Studies in Metaethics* vol. II, Russ Shafer-Landau, ed., pp. 220–39, Clarendon Press.
- (2009), “The Obscurity of Internal Reasons.” *Philosophers’ Imprint*, vol. 9:pp. 1–22.

- Gibbard, Allan (1999), "Morality as Consistency in Living: Korsgaard's Kantian Lectures." *Ethics*, vol. 110, no. 1:pp. 140–164.
- Gilbert, Margaret (2004), "Scanlon on Promissory Obligation: The Problem of Promisees' Rights." *Journal of Philosophy*, vol. 101, no. 2:pp. 83–109.
- Haase, Matthias (2014), "For Oneself and Toward Another: The Puzzle About Recognition." *Philosophical Topics*, vol. 42, no. 1:pp. 113–152.
- Harman, Gilbert (1977), *The Nature of Morality: An Introduction to Ethics*. Oxford University Press.
- Hayward, Max Khan (2019), "Immoral Realism." *Philosophical Studies*, vol. 176, no. 4:pp. 897–914.
- Hegel, G.W.F. (1807), *Phänomenologie des Geistes*. Reprinted as *Phenomenology of Spirit* in 1977, A.V. Miller, trans., London: Oxford University Press. Citations are to paragraph numbers in that edition; quotations are based on Miller's translations, though occasionally modified.
- Hooker, Brad (2011), "Promises and Rule Consequentialism." In *Promises and Agreements*, Hanoch Sheinman, ed., pp. 235–252, New York, USA: Oxford University Press.
- Jaeggi, Rahel (2014), *Alienation*. Columbia University Press.
- Julius, A.J. (MS), "Suppose We had Produced as Humans."
- Kant, Immanuel (1785), *Grundlegung zur Metaphysik der Sitten*.
- (1788), *Kritik der praktischen Vernunft*.
- Korsgaard, Christine (1996), *The Sources of Normativity*. Cambridge University Press.
- (2003), "Realism and Constructivism in Twentieth-Century Moral Philosophy." *Journal of Philosophical Research*, vol. 28, no. Supplement:pp. 99–122.
- (2007), "Autonomy and the Second Person Within: A Commentary on Stephen Darwall's *The Second-Person Standpoint*." *Ethics*, vol. 118, no. 1:pp. 8–23.
- (2009), *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.
- (MS), "Normativity, Necessity, and the Synthetic a Priori a Response to Derek Parfit." Unpublished manuscript retrieved from <http://www.people.fas.harvard.edu/korsgaard/Korsgaard.on.Parfit.pdf> on 4/29/2019.
- Lenman, James & Shemmer, Yonatan (2012), "Introduction." In *Constructivism in Practical Philosophy*, Jimmy Lenman & Yonatan Shemmer, eds., Oxford University Press.
- Mackie, J. L. (1977), *Ethics: Inventing Right and Wrong*. Penguin Books.

- Manne, Kate (2013), "On Being Social in Metaethics." *Oxford Studies in Metaethics* vol. XII, vol. 8:p. 50.
- (2017), "Locating Morality: Moral Imperatives as Bodily Imperatives." *Oxford Studies in Metaethics* vol. XII.
- Martin, Michael G. F. (2006), "On Being Alienated." In *Perceptual Experience*, Tamar S. Gendler & John Hawthorne, eds., Oxford University Press.
- McDowell, John (1983), "Criteria, Defeasibility, and Knowledge." In *Proceedings of the British Academy, Volume 68: 1982*, pp. 455–79, Oxford University Press.
- (1994), *Mind and World*. Cambridge, MA: Harvard University Press.
- Moore, G. E. (1903), *Principia Ethica*. Dover Publications.
- Moran, Richard (2018), *The Exchange of Words: Speech, Testimony, and Intersubjectivity*. Oup Usa.
- Murdoch, Iris (1971), *The Sovereignty of Good*. Routledge.
- Murphy, Liam (2020), "Purely Formal Wrongs." In *Civil Wrongs and Justice in Private Law*, Paul B. Miller & John Oberdiek, eds., pp. 19–39, Oxford University Press.
- Nagel, Thomas (1970), *The Possibility of Altruism*. Oxford Clarendon Press.
- (1986), *The View from Nowhere*. Oxford University Press.
- Parfit, Derek (2011a), *On What Matters Volume I*. Oxford University Press.
- (2011b), *On What Matters Volume II*. Oxford University Press.
- Railton, Peter (1984), "Alienation, Consequentialism, and the Demands of Morality." *Philosophy and Public Affairs*, vol. 13, no. 2:pp. 134–171.
- (1986), "Moral Realism." *Philosophical Review*, vol. 95, no. 2:pp. 163–207.
- (2009), "Internalism for Externalists." *Philosophical Issues*, vol. 19, no. 1:pp. 166–181.
- Rawls, John (1955), "Two Concepts of Rules." *The Philosophical Review*, vol. 64, no. 1:pp. 3–32.
- (1980), "Kantian Constructivism in Moral Theory." *Journal of Philosophy*, vol. 77, no. 9:pp. 515–572.
- Ross, W.D. (1930), *The Right and the Good*. Oxford University Press.
- Scanlon, Thomas (1998), *What We Owe to Each Other*. Belknap Press of Harvard University Press.
- (2014), *Being Realistic about Reasons*. Oxford University Press.
- Schacht, Richard (1970), *Alienation*. Doubleday.

- Schafer, Karl (2015a), "Realism and Constructivism in Kantian Metaethics : Realism and Constructivism in a Kantian Context." *Philosophy Compass*, vol. 10, no. 10:pp. 690–701.
- (2015b), "Realism and Constructivism in Kantian Metaethics : The Kantian Conception of Rationality and Rationalist Constructivism." *Philosophy Compass*, vol. 10, no. 10:pp. 702–713.
- (2016), "Hume on Practical Reason: Against the Normative Authority of Reason." In *Oxford Handbook of David Hume*, Paul Russell, ed., Oxford University Press.
- Scheffler, Samuel (1982), *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Clarendon Press.
- Schroeder, Mark (2007), *Slaves to the Passions*. Oxford University Press.
- Setiya, Kieran (forthcoming), "Other People." In *The Value of Humanity: A Re-Evaluation*, Sarah Buss & Nandi Theunissen, eds., Oxford University Press.
- Shafer-Landau, Russ (2003), *Moral Realism: A Defence*. Oxford University Press.
- Smart, J. J. C. & Williams, Bernard (1973), *Utilitarianism: For and Against*. Cambridge University Press.
- Smith, Michael (1999), "Search for the Source." *Philosophical Quarterly*, vol. 49, no. 196:pp. 384–394.
- Stocker, Michael (1976), "The Schizophrenia of Modern Ethical Theories." *Journal of Philosophy*, vol. 73, no. 14:pp. 453–466.
- Street, Sharon (2006), "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies*, vol. 127, no. 1:pp. 109–166.
- (2008), "Constructivism About Reasons." *Oxford Studies in Metaethics vol. III*, pp. 207–45.
- Sturgeon, Nicholas (1988), "Moral Explanation." In *Essays on Moral Realism*, Geoffrey Sayre-McCord, ed., Cornell University Press.
- Tarasenko-Struc, Aleksy (2019), "Kantian Constructivism and the Authority of Others." *European Journal of Philosophy*.
- Thompson, Michael (2004), "What is It to Wrong Someone? A Puzzle About Justice." In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, R. Jay Wallace, P. Pettit, S. Scheffler, & M. Smith, eds., pp. 333–384, Clarendon Press.
- (2012), "You and I." Given as a talk at a meeting of the Aristotelian Society; vol. cxii no. 3.
- Tiffany, Evan (2006), "How Kantian Must Kantian Constructivists Be?" *Inquiry*, vol. 49, no. 6:pp. 524–546.

- Travis, Charles (2004), "The Silence of the Senses." *Mind*, vol. 113, no. 449:pp. 57–94.
- Walden, Kenneth (2012), "Laws of Nature, Laws of Freedom, and the Social Construction of Normativity." *Oxford Studies in Metaethics* vol. VII.
- (2017), "Mores and Morals: Metaethics and the Social." In *Routledge Handbook of Metaethics*, Tristram McPherson & David Plunkett, eds.
- (2018), "Morality, Agency, and Other People." *Ergo*, vol. 5.
- (2020), "Reason and Respect." In *Oxford Studies in Metaethics* vol. XV, Russ Shafer-Landau, ed.
- Wallace, R. Jay (2019), *The Moral Nexus*. Princeton University Press.
- Wenar, Leif (2013), "Rights and What We Owe to Each Other." *Journal of Moral Philosophy*, vol. 10, no. 4:pp. 375–399.
- Williams, Bernard (1973), "A Critique of Utilitarianism." In (Smart & Williams 1973).
- (1976a), "Moral Luck." *Proceedings of the Aristotelian Society*, vol. supplementary volume L:pp. 115–35. Reprinted in (Williams 1981); citations are to page numbers from that volume.
- (1976b), "Persons, Character, and Morality." In *The Identities of Persons*, A. O. Rorty, ed., University of California Press. Reprinted in (Williams 1981); citations are to page numbers from that volume.
- (1979), "Internal and External Reasons." In *Rational Action*, Ross Harrison, ed., pp. 101–113, Cambridge University Press.
- (1981), *Moral Luck*. Cambridge University Press.
- Woods, Jack (2016), "The Normative Force of Promising." *Oxford Studies in Normative Ethics*, vol. 6:pp. 77–101.
- Zhao, Michael (2020), "Meaning, moral realism, and the importance of morality." *Philosophical Studies*.
- Zylberman, Ariel (2014), "The Very Thought of You." *Philosophical Topics*, vol. 42, no. 1:pp. 153–175.