The Sociality of Agency

by

Jack Samuel

BA, New York University, 2008

MA, University of Wisconsin, Milwaukee, 2013

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Jack Samuel

It was defended on

April 6th, 2020

and approved by

Japa Pallikkathayil, Associate Professor of Philosophy

John McDowell, Distinguished University Professor of Philosophy

Michael Thompson, Professor of Philosophy

Kate Manne, Associate Professor, Cornell University Sage School of Philosophy

Dissertation Advisors: Japa Pallikkathayil, Associate Professor of Philosophy,

John McDowell, Distinguished University Professor of Philosophy

The Sociality of Agency

Jack Samuel, PhD

University of Pittsburgh, 2020

Practical philosophy is dominated by two pictures of human agency: the Kantian image of a rational, empirically unaffected will and the Humean image of instrumental reasoning from desires, aims, values, or interests. Because Kantian and Humean accounts of agency emphasize the individual over the social, they lack explanatory resources important to understanding how others can matter to us as agents in the right way. Insufficiently social conceptions of agency, I argue, risk depicting agents as alienated from one another, leaving mysterious how we can get a normative grip on one another. Taking inspiration from GWF Hegel and Iris Murdoch, I develop a conception of agency on which it constitutively depends on standing in relations of mutual recognition with other agents. I argue that a crucial part of becoming an agent is becoming reflectively self-conscious of our sociality—of the historically conditioned nature of mutually recognitive relations. A significant upshot of the picture of agency that emerges is a new perspective on individual moral progress. It's not the purely personal achievement suggested on the Kantian and Humean accounts of agency. Rather, individual moral progress is a political achievement, bound up in both the inner struggle to achieve moral clarity and the social struggle for an ethical form of life.

List of Tables

Preface

Last night the COVID–19 death toll crossed the 15,000 mark in the US, and is close to 100,000 worldwide. Unemployment claims continue to set records every week, nearly-completed academic job searches are being canceled, and many or perhaps most American universities have already announced hiring freezes for next year. (I expect to attend law school in the fall, having failed to secure academic employment.) I've spent most of the last month at home with Winnie, which, as a somewhat solitary grad student, has been less of a change to my day-to-day lifestyle than perhaps it ought to have been. My defense three days ago was remote, via videochat. The last in-person meetings I had with my advisors have retroactively become the last meetings I will ever have had with them as advisors.

I understand that upon defending one's dissertation it is common to feel an odd combination of overwhelmed and underwhelmed: the latter because, as with turning 30 having spent 11 months feeling 30 already, this just makes it official; the former because it nevertheless marks the end of something significant. This unusual backdrop may have heightened the surreality of the moment, as I found myself just now in an unexpectedly raw state making final changes based on John's proofreading notes on my manuscript. Being told that a comma is needed is not usually so poignant.

In any case, this is the preface to a dissertation, and not the beginning of a memoir. It felt absurd not to remark on the context. Having done so, some acknowledgements. Early work on what became this dissertation began as a paper for Karl Schafer, to whom I am grateful for feedback and support throughout. Bob Brandom was originally on my committee before the project shifted focus, and chapter 4 is largely a product of reading his work, attending his seminars, and meeting with him to discuss Hegel. I am grateful as well to the rest of the Pitt philosophy faculty, from whom I have learned so much, and for the faculty at the University of Wisconsin, Milwaukee, who helped me get here. Special thanks are due as well to Bill Bristow, for helping to keep the fire lit.

Both philosophy departments are also exemplary for their collegial and intellectually serious communities, and I have benefited tremendously from conversations with my peers

during seminars, Q&As, receptions, works-in-progress talks, dinners, board game nights, parties, reading groups, and coffee shop co-working dates. In particular I would like to thank the members of the Hegel reading group and the ethics reading group for helping me understand some of the core issues I engage with in my work. Discussing philosophy with them has been a highlight of my adult life so far.

For valuable discussions of and feedback on the ideas and work that became part of this dissertation I would like to thank Stephanie Allen, Paul Blaschko, Adam Blazej, Eric Brown, Sara Ĉopić, Daniela Dover, Josh Eisenthal, Maegan Fairchild, Andrew Flynn, Markus Gabriel, Elena Garadja, Chuck Golhaber, Barbara Herman, Nathan Howard, Ned Howells-Whitaker, Lucian Ionel, A.J. Julius, Dan Kaplan, James Kreines, Quill Kukla, Nick Laskowski, John Lawless, Kathryn Lindeman, Suzie Love, Barry Maguire, Analisa Paese, Aaron Salomon, Kirun Sankaran, Joe Saunders, Aaron Segal, Julius Sensat, David Shope, Alison Springle, Eric Swanson, Olúfẹ́mi Táíwò, Nandi Theunissen, Lynne Tirrell, Alnica Visser, Willem deVries, Kenny Walden, Dan Webber, Eric Wiland, Jack Woods, the Pittsburgh feminist philosophy reading group, the graduate students and faculty at UCLA and Cornell, and the participants in the Bonn University summer schools.

For not only feedback on written work and ongoing discussions that shaped me as a philosopher, but also personal support and encouragement when I needed it the most, I would especially like to thank Tom Breed, Liam Bright, Sophie Cote, Eleanor Gordon-Smith, Mark Lance, Christa Peterson, and Francey Russell, as well as my family and friends outside of philosophy, who kept me tethered, and Winnie, who is a very good girl. (Special shout-out to Rich's noon hoops crew, Ironsmith Tattoo, Pitt GSOC, and the Braddock mechanics—it's good to have a life outside of work.)

Last, my committee, all inspiring philosophers and generous and supportive mentors. I feel incredibly lucky to have had the chance to work with you.

1.0   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. We must be able to understand ourselves as the kinds of agents a theory envisions, and to understand our relations with one another through the theory's lens, at the risk of leaving us with a sense of alienation.

As Kantians emphasize, one way that metaethics can be alienating is by conceiving of agents as potentially estranged from the normative facts that it is designed to explain. It must be able to account for the grip that reasons have on us. The threat of this kind of *normative alienation* urges placing the agent (the valuer, the reasoner, the creature with desires) at the center of any explanation of normativity. In this dissertation I argue, however, that philosophical work on agency is dominated by pictures too austere and abstract to account for the ways that it materializes in a particular social contexts, and that its ethical significance is tied to the social nature of its exercise. This points toward another, undertheorized kind of alienation that metaethics must resist: alienation between an agent and the concrete others with whom morality is principally concerned. The threat of *social alienation* reveals a need to emphasize the significance of moral subjects, external to the agent and equally real. A theory of normativity suited to avoid both forms of alienation would thus paradoxically seem to need to center both the agent and the subject. The tension can be resolved by appealing to a more social conception of agency.

In chapter 2 I develop the worry about social alienation in more detail, by revisiting Bernard Williams's alienation objection to modern moral theory and Peter Railton's response. Williams's alienation objection gets at an important way in which a satisfying account of agency will need to make sense of the rich complexity of our lives. However, he saw in moral theory a threat only to the internal integrity of the agent. The feature of human life from which moral theory threatens to alienate us he theorizes as a part of ourselves, thus painting a picture of us as problematically disconnected from feelings, desires, passions; from our 'affective selves'. Williams captures one way in which moral theory risks theorizing away

1

important parts of agency, but the other persons to whom we stand in social relationships that likewise partly constitute the realization of our agency drop out of Williams's story.

In his response to Williams Railton suggests that a complete answer to the problem of alienation will involve not only making room for the integrity of the agent, but a kind of *social* integrity as well. Railton brings the puzzle into view but does little by way of elaborating what a solution would look like, hinting only at the need for one already implicitly contained in Williams's original objection. When we route the normative significance of others and our relationships with them through normative and psychological statuses that attach solely to individual agents, I argue, we again lose our grip on the lived social reality that gave rise to anxieties about alienation in the first place. A satisfying account of agency will need to make room for what I call 'genuine ethical contact' with others, both as concrete objects in the world external to ourselves and as subjects who can recognize us reciprocally.

In chapter 3 I review three standard objections to metaethical realism—the Humean challenge to explain motivational internalism, the constructivist challenge to explain the rational relationship between reasons and agents, and the epistemic challenge to explain how normative facts can be knowable—and argue that they reflect an underlying anxiety about normative alienation. I then draw out the tension between addressing normative and social alienation, and sketch two different approaches to navigating it, the first inspired by Hegel and the second by Murdoch.

Drawing on G.W.F. Hegel, I argue in chapter 4 that standing in historically-conditioned relations of mutual recognition with other agents is a constitutive element of agency. I first develop a conception of mutual recognition as a normative attitude that places another with respect to oneself, such that the mutuality of mutual recognition is something over and above mere symmetry of attitudes between agents: it is a form of social self-consciousness. This formal characterization of constitutive recognitional relations suggests a form of metaethical constructivism that can answer some of the objections against its Humean and Kantian rivals.

Drawing on Iris Murdoch, I argue in chapter 5 that agency is not only actualized in the moments of choice and action but is constantly at work in cultivating and improving our unique practical standpoints, through which we make contact with a normatively saturated

2

world. I then argue that for Murdoch practical standpoints are conditioned by social context, and that a crucial part of cultivating and improving them is the struggle to become reflectively self-conscious of our sociality. The aim of agency, according to Murdoch's conception of moral progress, is to transcend the socially alienated standpoints through which we encounter a world obscured by ideology and self-focus, and to develop a standpoint through which we can become self-conscious of our social relations and the political forces in which they are embedded.

The picture of agency that emerges is at once ethical and political, bound up in the inner struggle to achieve moral clarity and the social struggle for an ethical form of life.

The essays that compose this dissertation are each self-standing and can be read apart from the others. However, given the systematic ambitions of the dissertation as a whole, for each to be self-standing requires considerable substantive overlap. In a few places the argument in §4.3.3 draws on the upshot of §2.4. Most significantly, §3.3 summarizes the central theme of chapter 2, and §4.2 draws heavily from chapter 3. In a few places whole sentences or even paragraphs may be common between two chapters, as a result of the somewhat unwieldy process through which they were all completed over the course of two years—I have tried to minimize verbatim repetition, but I have not entirely succeeded.

If some substantive overlap is unavoidable and innocent, and some verbatim repetition only mildly irritating, most troubling may be the apparent dissonance between versions of what I attempt to present as common ideas. To my great frustration a general statement of the problem of social alienation remains elusive. I take it that the version presented in chapter 2 points in the same direction as the summary in §3.3, and for that matter the shorter summary in §4.2.2 and the sketch of a similar idea in §§5.3.1–5.3.2. The discussion in chapter 2, however, emerges from reconsidering a psychologistic diagnosis of a classic problem in normative ethics, and revolves around the difficulty of understanding what it is to *know* and to *deliberate from* situationally-relevant moral facts, while the later iterations of social alienation begin with the question of what it is for metaethics to *explain* moral facts. What all these discussions have in common is the underlying conviction that morality is centrally a matter of how we find ourselves with others, and that this is something philosophical ethics has yet to fully account for.

2.0   An Individual Reality, Separate from Oneself

2.1   Introduction

Though we do not always use the term 'alienation,' a common complaint against moral theory alleges that it is often disconnected from everyday life, or perhaps that it leaves us standing outside of the world, reflecting abstractly on facts about duties or rights or aggregate utility. The underlying impulse is to challenge moral theory's relevance to lived ethical experience. If this or that moral theory were correct, we sometimes worry, some feature of ordinary ethical life would be rendered mysterious, or in Iris Murdoch's phrase, "theorized away." Critics of alienation object that moral theory has traded in its groundedness in or relevance to the ethical phenomenology of everyday life for the appearance of rigor, universalism, systematicity, simplicity, or some other virtue typically attributed to scientific enterprises. The hallmark of alienation is that it sunders us from something of value: a part of ourselves, a part of our lives, or a part of our world. The goal of alienation critique is to reunite ourselves with whatever it is, to restore a sense of ourselves as unified ethical subjects for whom the things that matter can be intelligible.

The concept of alienation receives its canonical expression within 20th-century Anglophone moral philosophy in Williams's *Critique of Utilitarianism*. There and in (Williams 1976*b*) he insists that the reader attend to the lifeworld of the agent, to the commitments, projects, and relationships that are especially *hers.* They belong to her in a way that contributes to the realization of her agency as more than a mere chooser, more than what Murdoch called "the quick flash of the choosing will".[1] They are part of what it is to be an agent.[2]

Williams sees in moral theory a threat only to the internal integrity of the agent; the feature of human life from which moral theory threatens to alienate us is theorized as a

---

[1](Murdoch 1971, 52)

[2]Or at any rate they are part of what it is to be a *human* agent, to be the kind of embodied, embedded agents that we in fact are. I am not attributing to Williams a claim about agency as such (as theorized by, for example, metaethical contructivists), nor do I intend to make any such claims myself. In what follows, "agency" should be understood to have this narrower meaning.

part of ourselves. Moral theory paints a picture of us as disconnected from feelings, desires, passions; from our 'affective selves'. Perhaps this criterion is an important one. If the kinds of persons, agents, or moral subjects that a given moral theory envisions are inadequate as a philosophical account of what we are like—if we can't be like that because we would be essentially fractured, restless, at war with ourselves—then we ought to reject any view that understands us as such. But the other persons to whom we stand in the social relationships that partly constitute the realization of our agency drop out of Williams's story.

In his response to Williams, Railton (1984) suggests that a complete response to the problem of alienation will involve making room not only for the integrity of the agent but for a kind of *social* integrity as well. It is not enough to reassure us that a harmony of the rational and affective parts of soul is possible without at the same time accounting for the way that we are essentially among others. Williams does not appear to be concerned with social alienation, or even to have it in view (though it seems at least suggested by the drowning spouse example). Railton's treatment illuminates the puzzle of social alienation, but he does little to elaborate what a solution would look like beyond hinting that Williams's original objection implicitly contains it.

It is this problem of social alienation that I want to get a better grip on, and the corresponding demand for an account of social integrity. As a placeholder, what I want to suggest is sundered or at least obscured by moral theory is not (or at least not only) the psychic unity of an individual but what I will call the ethical *contact* that we make with one another. The metaphor of 'ethical contact' is helpful in illuminating the parallel between the worry I want to uncover and a long-standing worry in the epistemology of perception: that a theory leaves us trapped behind a 'veil of ideas', never really making *perceptual contact* with the world.[3]

---

[3]See e.g. (Sturgeon 2008, 112): "Visual experience is remarkable for two reasons. It seems to involve conscious portrayal of the world; and it seems to involve perceptual contact with the world;" the term is also used by other contributors throughout the same volume (Haddock & MacPherson 2008) and throughout the literature on perception. As Sturgeon notes there is little agreement about what perceptual contact amounts to, but near consensus that whatever it is, a theory of perception must make good sense of it.

Another advantage to the term 'ethical contact' is that, having no existing technical use, it has accreted none of the associations or competing precisifications that might incline a skeptical reader to reject the letter of my proposal without first at least considering the spirit. This is also a disadvantage, however, insofar as it may be entirely unclear just what it is (aside from psychic harmony) that I think moral theory risks theorizing away. I'll try to say more to characterize it positively later, once I've put in place the analogical

In doing so I hope to diagnose a philosophical picture that holds us captive and gets in the way of real understanding—both in the sense that it gets in the way of proper philosophical understanding and in the sense that it gets in the way of proper practical understanding of how we exist in the world among others. I'll argue that moral philosophy's focus on the individual and her mental states can obscure or alienate us from our fundamental sociality and its irreducible place in moral life.

When we route the normative significance of others and our relationships with them through normative and psychological statuses that attach solely to individual agents, we lose our grip on the lived social reality that gave rise to anxieties about alienation in the first place. Thus, moral theory leaves us alienated from the being-among-others that constitutes the substance of ethical life and leaves any agent who understands herself in these terms alienated from the concrete others with whom she shares a world.

2.2 Alienation in General

Williams (1973) objects that consequentialism misconceives agency, providing a picture of the agent as a mere point of intervention, a 'causal nexus'. The agent weighs the possible outcomes of different actions, taking no account of whether they will have been brought about as a result of her choice, by mere happenstance, or by the agency of another. Insofar as she can permissibly take on projects and commitments that serve the greater good, she cannot fully *commit* to them: she must be prepared to abandon them the moment the deliberative calculus shifts in favor of doing so.

Even worse, we must take equal account of the projects and commitments of others, weighing their contribution to the overall good in the same way and to the same extent as we weigh our own. Our own projects are not truly *ours* if they don't mean any more to us than those of others. We can neither genuinely commit to anything nor identify with what little commitment is left to us.

---

basis in perceptual epistemology—I hope that at this stage there is some appeal to the vague idea that there is some sense in which moral theory risks alienating us from others.

This picture of agency is a threat to our integrity. We cannot understand ourselves in the way consequentialism requires, at the risk of losing our grip on ourselves. What exactly that amounts to is left somewhat vague. However, Williams raises a similar objection in (Williams 1976b), wherein the diagnosis of what has gone wrong is a bit clearer. What emerges is that for Williams a virtuous agent acts out of an immediate recognition of what the situation calls for, while it is the mark of a merely *enkratic* (continent, self-controlled) agent that she needs to will herself to do what she believes is right.[4]

Virtue renders appealing to an explicit deontic order unnecessary. Talk of 'reasons' and 'oughts' is epiphenomenal to a well-lived life. The kind of unity Williams is interested in— that of a well-ordered soul—manifests in responding appropriately to the salient features of one's environment. The 'one thought too many' mistake is one of intellectualism, of thinking that a judgment about duty is necessary when the virtue of, say, fidelity will orient an agent toward the right action without the intervening apprehension of (or any computation based on) a rule. The paradigm case of alienation for Williams is of an *enkratic* agent who is motivationally estranged from her moral judgments; integrity, here understood as virtue, obtains when one is immediately motivated to act in the right way.

The conception that Williams is here arguing against—a kind of Nietzschean police passion that moves us to act on our better moral judgment and against our natural inclinations—

---

[4]For the most part Williams does not use this Aristotelian vocabulary (aside from invoking the notion of character), and facially his topic is, again, what is wrong with moral theory. However, what interests me here is the implicit picture of non-alienated agency that underlies the objection, and on my reading the language of virtue and *enkrateia* captures it nicely. The same point can be easily made without it: impartial, abstract moral theory is a threat to agency because it allegedly requires us to exercise an explicit calculative faculty, rather than to rely on the ways in which simply being a person with values, commitments, projects, and so on can directly guide us to act in ways that we should not, on behalf of moral theory, try to disallow. In the context of the 'one thought too many' case in particular, Williams's point is that *even if* moral theory can arrive at the right result we shouldn't need it when an agent with integrity will just know what is to be done:

> The consideration that it was his wife is certainly, for instance, an explanation which should silence comment. But something more ambitious than this is usually intended, essentially involving the idea that moral principle can legitimate his preference, yielding the conclusion that in situations of this kind it is at least all right (morally permissible) to save one's wife... But this construction provides the agent with one thought too many: it might have been hoped by some (for instance, by his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife. (Williams 1976b, 18)

leaves in place the internal conflict. To do without the police passion is to live an internally unified life, but in arriving at such a recommendation the other has dropped out of the picture. Suppose I manage to cultivate the right virtues of character—what does that have to do with my spouse, my friend, or any other ethically significant person with whom I come into contact?[5] If I do manage to regard my commitments and projects as uniquely my own I may do so without being integrated into a social context in the right way—integrity in this sense is partly a matter of maintaining *separateness* from others and their projects.[6] The demand for integrity reflects a concern for what one is or ought to be like psychologically, and not primarily for how one must stand and see oneself in relation to others.

The now-standard diagnosis of alienation as a threat to intra-psychic harmony is especially clear in Railton's discussion: with those who regard their spouses as sources of utility or opportunities to fulfill general duties, he notes, "there would seem to be an estrangement between [agents'] 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).[7] With an objection like Williams's in mind, he is concerned in particular with whether a consequentialist approach to moral theory encourages us to be the kinds of agents who cannot meaningfully commit to friendship or love. His solution is a sophisticated form of consequentialism that allows—or even requires—an agent to commit to practices with certain internal ends and to cultivate traits of character well-served to meet them.

Railton argues that morally good agents may commit to relationships that require the

---

[5]A proper understanding of virtue may not be available independently of an understanding of social integrity of the kind I'm trying to uncover here, at least for social virtues like fidelity or justice. Williams may have the resources to respond along these lines, though there is reason to worry that his implicit notion of virtue as correspondence of practical judgments and motivations rules it out, particularly given the Humean moral psychology he adopts elsewhere (Williams 1979). However, even if this reply is available to Williams it is striking and sufficient for my purposes that he seems uninterested in asking the question to which it is an answer.

[6]It is at least a stylistic if not a substantive theme in Williams's work that his thought experiments center individuals—always men—who *resist* being constrained by the putative demands of morality. Perhaps the clearest example is Gauguin in (Williams 1976a): Williams holds it against moral theory that it might attempt to block the painter from abandoning his family and moving to Tahiti to be an artist. Personal integrity, for Williams, seems to have little to do with accounting for the ways in which we are socially situated.

[7]Compare Michael Stocker, whose central case (1976) involves someone explaining their choice to visit a friend in the hospital by appealing to the duties of camaraderie. For Stocker, '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.

devotion of resources that might sometimes do more good elsewhere, provided it's better overall that such relationships exist. Sustaining such relationships will require character traits that ground psychological harmony. This is consistent with the agent meeting his moral demands if, Railton concludes, "while he ordinarily does not do what he does simply for the sake of doing what's right, he would seek to lead a different sort of life if he did not think his were morally defensible" (Railton 1984, 151). In other words, the psychologically unified agent does not reason on an act-consequentialist basis while ignoring their partial motivations but has a background belief that consequentialism supports being the kind of person who has such motivations.

While the cases that motivate Williams and Railton involve a perverse kind of erasure of morally significant others from the sorts of considerations that show up in a moral theory, this criterion of integrity is a curiously psychologistic one.[8] So understood the problem of alienation is an internal concern about the psychology of a single agent. An attempt to resolve it will have to show that psychological unity is possible, and in particular that the correct normative theory (whatever that turns out to be) is consistent with that unity.

The interiorizing move is one way of getting a grip on the intuitive worry but in characterizing the cause of that worry as internal we arrive at a theory that has no place for the other. Throughout his article Railton gestures toward this further concern, noting early on

---

[8]There is a similar problem with philosophical reflections on love that treat it as a bundle of beliefs, affective states, dispositions, and reasons an agent has, rather than fundamentally as a relation two (or more) agents stand in together. For instance Setiya (2014, 270–1) explains the justifiability of saving one's spouse over three drowning strangers in terms of facts about one's mental states: "the rationality of your action does not depend on aspects of love that go beyond the disposition you manifest in saving M, a disposition to give priority to her needs." In contrast, Buber characterizes love as irreducibly relational:

> Feelings dwell in man; but man dwells in his love... Love does not cling to the I in such a way as to have the Thou only for its "content," its object; but love is between I and Thou. The man who does not know this, with his very being know this, does not know love; even though he ascribes to it the feelings he lives through, experiences, enjoys, and expresses. (Buber 1923, 19)

Though Kolodny doesn't directly address the question of reducibility with respect to love-relationships, in treating loving relationships as the source of at least some reasons—and thus the relation as prior to the having-a-reason property of an individual—Kolodny (2003) is in a sense closer to Buber here. He also speaks of 'alienation' from other persons in a way similar to the one I propose (157,161), though it is nowhere near his central concern. Another characteristically internalist formulation of a related problem (under the label "estrangement") is Moran's (2001), though as a theory of self-knowledge it is only indirectly related to the kind of moral alienation at issue here.

that there is a risk of alienation not only from from one's personal commitments and one's feelings or sentiments but from other people as well. While his main focus throughout is personal alienation and personal integrity he points out that these are bound up together: "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).

Williams gets us no further than the thought that things are going wrong if we conceive morality as the business of some isolable, rational part of the soul, and it would be better if we integrated our propensity for feelings with our reason. This criticism doesn't grasp the significance of misconceiving human relations. Recommending an integrated conception of practical reason (with passions not conceived as separate from reason) leaves the worry about social alienation untouched.

Railton adds to this the condition of seeing ourselves as socially situated. This provides the frame for an account, but the question immediately arises: *what does it take to conceive ourselves as situated?* We will not get that right if all that we find missing in the case of the direct-utilitarian husband is a proper connection to his own feelings. It's not nothing to say that we must be allowed to have sentimental relations to one another that we're not

estranged from; but this is not enough to return us to a non-alienated standpoint.

Consider a generic case of someone whom Williams would recognize as in some sense socially alienated: if we understand his alienation from his world in terms of a lack of other-directed affections, we will think it can be solved by enriching his psychology. Once he has the feelings associated with friendship and love, and once those are properly integrated with his values, the problem will have been solved and integrity achieved. But this is still only an account of *personal* integrity. It says nothing about the social relations he stands in themselves.

The object-level dispute between Railton and Williams (and between Williams and Smart, in the volume containing Williams's *Critique*) is moral theory: an account of the good and the right and a corresponding account of deliberation done well. But the worry about alienation itself is a worry about agency as conceived by moral theory; about what, given that theory, we must be like. The problem of social alienation is that moral theory alienates us not just from ourselves but from our relationships to others as well when it analyzes social relations in psychological terms and thereby attempts to translate an external dyadic relationship into an internal relationship or property. That other persons figure in our ethical worldview only as they figure in the contents of our mental states means that we lose, so to speak, the otherness of the other (in a way that I hope to make more precise as I go on). The task is to provide a philosophical reunification that keeps the focus on relationships, rather than deflecting to psychology.

I don't mean to suggest that intra-psychic disharmony is unproblematic. It is arguably an important criterion of adequacy for moral theory that it be consistent with unified a moral psychology.[9] But this is not the only sense, and to my mind not even the most salient sense, in which Williams and Railton's cases risk producing a shattered picture of moral agency.

---

[9] I suspect, however, that the importance of intra-psychic harmony is sometimes exaggerated. Not only, as I will shortly propose, might personal integrity depend on social integrity, but in some social circumstances the demands of social integrity may even *require* accepting personal alienation. Sufficiently non-ideal political environments, for example, may require that some agents distance themselves from their projects and values, and even learn to distrust their own inclinations.

The same may be true of social alienation, however—some political contexts may require not only distance from parts of oneself, but distance from problematic social relations in which one is involved. Railton notes, similarly, that "the alienation of some individuals or groups from their milieu may at times be necessary for fundamental social criticism or cultural innovation" (Railton 1984, 148).

The psychological concern is that we could not be the way moral theory proposes to understand us because then we would be essentially fractured, restless, and at war with ourselves. It seems to me, however, that we can't be the way moral theory proposes to understand us because then we would be merely incidentally in contact, occasionally bouncing into each other—windowless moral monads, each containing the complete story of our ethical lives from our own perspectives. That my ethical representations of you and your ethical representations of me correspond at all would be a matter of grace or luck, but there would be no real existence to ethical relations, just the correspondence of monadic properties. This is a bad way to understand ourselves, one that denies the possibility of ever standing in normative relations with each other.

The demand of social integrity is to have others show up in the right way in the account of what we are, in their separateness and concreteness, and not merely as reflected in our affective states. We must be able to recognize them as "a reality separate from ourselves" (Murdoch 1971, 42), and ourselves as standing in relation to them.

To get a better grip on the kind of concern I have in mind I want to take a quick detour through the epistemology of perception.

2.3   Perceptual Alienation

A common criticism of traditional epistemological theories is that they leave us trapped behind a 'veil of ideas', never perceiving more than the contents of our own minds. Though not typically characterized as a worry about 'alienation,' it may as well be since it concerns a disconnect between subjects and the objects of their experience—or more generally between subjects and the external world.[10]

In response, epistemologists attempt to explain how the denizens of the external world can genuinely feature as that to which we are related in perceptual experience, as what we

---

[10]The objection is probably first expressed by Sextus Empiricus—see the introduction to (Brewer 1999) for a review of the literature framed around this question. A rare instance where the term 'alienation' is used in this context is the title of Mike Martin's "On Being Alienated" (2006), though he nowhere uses the word within the paper. In addition to those cited throughout, Charles Travis has done a lot to push in this direction, and much of the discussion here draws indirectly on (Travis 2004).

perceive rather than as what we merely represent to ourselves in an inner theater. If there is room for a notion of representation in an account of perceptual experience at all, we should not be fooled into thinking that these representations are themselves the only things that we can be in perceptual contact with: rather, we *see* external objects *through* or *by* representing them. Better, however, to think of perception not in the first instance as representing things at all but as *relating* us to those things ((perhaps though representations), or to understand ourselves as *directly aware* of things when we perceive them. The first criterion of perceptual integrity, so to speak, is that a theory can make sense of genuine perceptual contact with an external world.

However, perceptual alienation is about more than 'perceptual contact', as I've glossed it so far; there is a further concern about how we get *concrete particularity* into a picture where all of the resources available to us are general.[11] That is, it may not be enough to explain how perception can get us out from behind the veil of ideas without explaining how it can have a particular object, and not just a set of general attributes the object has, as itself that to which perception puts us in relation. The question of how generality can relate us to a particular—known sometimes as a worry about 'demonstrative' or '*de re*' content—is less commonly thought to establish a criterion of adequacy.[12] Nevertheless, the practical version—that others must be in view in their concrete particularity—may have more intuitive attraction. The second criterion of perceptual integrity is that an epistemological theory can make sense of the concrete particularity of worldly objects, and not just their external-worldliness.[13]

Given the prominence that the first worry enjoys in the epistemology of perception, it is striking that when we look to how other persons are involved in our ethical self-understanding we are tempted to insist on a crude kind of psychologism, according to which our normative standing vis-à-vis others is mediated by their ability to produce motivational states in us or to instantiate universal rules of conduct. We seem content to understand ourselves as normatively cut off from one another, trapped behind a veil of motivations or intentions, of

---

[11]See e.g. Soteriou (2000).

[12]Tyler Burge (2010), for example, argues on empiricist grounds that we ought not to expect a metaphysical difference of this kind.

[13]Perhaps these are two independent criteria, perhaps they are different ways of specifying the same one—this question is not important for me here.

pro-attitudes or maxims. While we are disturbed by the idea that all we are ever in perceptual contact with are the contents of our own minds, in moral theory we seem untroubled by the idea that our ethical contact with others takes place at the level of beliefs, desires, maxims, etc; that others are not a part of what it is for us to be moral agents. The problem of perceptual alienation and corresponding criteria for perceptual integrity can be re-purposed to help us get clearer on what it would mean for the other to appear in the right way in an account of what we are: social integrity requires making sense of how others can be significant for us as external and as particular. Borrowing the term from Murdoch, I'll call this the *individual reality* condition on social integrity.

At a general level, one way of capturing the demand for perceptual integrity is through 'disjunctivist' theories of perception. Disjunctivists resist identifying experience itself with what is (or seems to be) common between genuine (i.e. knowledge-affording) perception and merely apparent perception (appearances, seemings, sense-data, etc.), where the genuine case involves the same content plus some extra fact about its accuracy. Instead, a genuine case is understood by the disjunctivist as one in which the experiencing subject is in contact with a concrete, particular object in the external world, and a degenerate case is understood as one in which the subject merely seems to be in this kind of contact. What is common between good and bad cases is re-interpreted as a disjunction: the appearance, seeming, sense-datum etc. is not common to both kinds of cases. Rather, one is experiencing either a *mere* seeming (etc.), and thus not really perceiving, or one is genuinely perceiving the object. While the two cases are subjectively indistinguishable, the theory vindicates that at least when things are going right perceptual experience meets the equivalent of the individual reality constraint. Whether or not we are perceiving something is not a matter of what's going on inside of us, plus the cooperation or luck of nature; it is a feature of the relation between someone and the object of their experience.[14]

---

[14]In addition to the Brewer and Travis cited above this view is probably closest to McDowell's—see (McDowell 1983*b*; 1984; 1994). See Pritchard (2012) for an overview of epistemological disjunctivist approaches to perceptual experience.

2.4   Openness to the Other

The individual reality condition gives us the beginning of an understanding of social integrity, but it cannot be enough on its own. In the perceptual case the problem is generated by the obscurity of having an object in view in the right way. Conceiving of our agency as bound up with others, however, brings along the additional demand that they show up in the story *as subjects*.

If the gold standard for an integrated theory of perception is that it makes sense of our epistemic "openness to the world", what an integrated moral theory requires is that it make sense of the agent's 'practical openness to the other.'[15] The openness required in this case isn't merely to a worldly object but to another subject—another like oneself. My practical openness to another isn't separable from their 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.

Compare a variation on John Perry's famous case of the shopper, who, both leaving a trail of sugar behind himself and searching for the person spilling sugar all over the store, eventually sees himself in a mirror and takes himself to be looking at "the messy shopper going up the aisle on the other side" (Perry 1979, 12). Perry's shopper is alienated from himself in this moment in a peculiar way, in that he is both the object and the subject of his awareness but not as a case of self-awareness. Similarly, I might spy one of my friends across Schenley Park, apparently looking not at me but into a restaurant. Meanwhile, she notices me in the reflection of the window, but because of the reflection's angle does not realize that I'm looking at her. In this situation we both recognize each other but neither of us is aware of being recognized. Our symmetrical recognition, in this case, does not constitute the shared experience of one another that characterizes properly *mutual* recognition. This is the possibility that moral theory must leave open, and which is closed by a psychologistic analysis of alienation.

The kind of recognition at issue here is mutual not just because both parties involved have matching recognitional attitudes towards one another but because what it is to recognize

---

[15]The phrase 'openness to the world' appears throughout (McDowell 1994).

another is to recognize them as recognizing oneself. The implicit conceptual or propositional content is not just that another is a human, a person, an equal, but a fact that places her with respect to oneself. Each of us does not only recognize what the other is like but also how they stand to us and thereby we to them. They are thus not (only) the content of our awareness, as they could be while the independent reality condition was met, but sharer with us a recognitional nexus. This is not to say that every experience of another has this structure—what I'm trying to get at here is a structure that a satisfying picture of agency as a socially-embedded phenomenon must have room for, because it is a central part of how we understand ourselves that we are among others.[16]

What I'm trying to sketch here might be understood as a kind of interwoven, mutually constitutive practical openness to one another, such that one's practical relation to another and the other's practical relation to one are not independent but two ways of specifying the same shared self-consciousness. Consider the shared experience of a sunset.[17] Two friends are looking at a sunset together. What's happening is more than that one is looking at the sunset and the other is also looking at the sunset. It is part of the experience for each that it is shared; it is an experience of looking at a sunset with a friend. But this in turn is not just a matter of one having an experience that includes the other's having an experience in its content in addition to the other having an experience that includes the first one's experience in its content. The experience is of looking at a sunset with another who is also having the experience of looking at a sunset together. Velleman refers to this kind of phenomenon as 'joint attention'.

This kind of structure is sometimes analyzed reductively in terms of an infinite series of nested mental states that by definition constitute "common knowledge". The alternative that I'm recommending here (as is Velleman, in different terms) involves a recognitive element in each person's experience that is not merely symmetrical but somehow combined.[18]

---

[16]Even though it is not the structure of every experience of another, it is perhaps the structure of paradigmatic moral relationships: the most fully realized form of self-other relation. At the least it is one important form relating to others can take. Setiya (MS) and Korsgaard (2018) argue for different reasons that important moral relationships are often asymmetrically recognitive, Setiya in the case of love and Korsgaard in the case of non-human animals. What I say here is intended to be consistent with both.

[17]This discussion draws on (Velleman 2013).

[18]See (Lederman 2018a) for an review of the literature on common knowledge, and Lederman (2018c;b) for additional criticism of the notion's viability. My objections to theorizing joint attention in terms of

A toy disjunctivist theory of joint attention will serve to at least see how an alternative to the common knowledge analysis is coherent.[19] The thought is that a true case of joint attention involves a relation that two parties stand in to one another, vis-à-vis a sunset (or whatever), rather than being factorable into what is happening in one and what is happening in the other, plus the luck or cooperation of nature that they correspond in the right way. (Recall the worry about monads, corresponding as a matter of grace.)

The degenerate case, one where I think I'm sharing such a moment with a friend but she is in fact thinking about how little she gets out of our friendship and ignoring the sunset entirely while she plots to sever ties, may be subjectively indistinguishable, for me, from a case of genuine joint attention (though I would like to think that if I were perceptive enough I could spot the difference). The same can be said for a case where, unbeknownst to me, my friend has been replaced by a cleverly-disguised robot, and thus there is no one else for me to share the experience with. Still, where genuine joint attention differs from ersatz joint intention is not that I am in the same state plus something else is true: in a case of genuine joint attention I really am in the state that I seem to be in, whereas in a case of ersatz joint attention I am in fact having an experience of something that is not what it seems (itself a troubling kind of alienation).

Recognition is thus not the apprehension of a fact about another but a fact about *us*, the apprehension of which partially constitutes its object. When one experiences a mutually recognitive interaction with another there are not two symmetrical recognitional states but a single, shared experience of recognition, in which one's recognition of another is a part of the other's recognition of them, and vice versa.[20] Agency always already involves, in A. J. Julius's phrase, "interpenetrated capacities for living interpenetrated lives" (Julius 2016,

---

common knowledge are less technical, though Lederman's points are well-taken. I don't have the space here to discuss why I find common knowledge accounts of phenomena like these unsatisfying, though I am broadly sympathetic to Michael Thompson's (Thompson 2012). See also Richard Moran (2018) on what he, following Reid, calls 'social acts of mind'.

[19]Disjunctivist theories have their own difficulties, and in the end we may find that disjunctivism is the way to theorize shared self-consiousness or just a convenient stopping place along the way.

[20]There is an echo here of what is sometimes called "Hume's circle", wherein one cannot make a promise without using the concept of promising, and thus the term is partially constituted by the practice to which it refers, and the practice presupposes the use of the term (see Anscombe 1978). The common feature of the two cases is a state of affairs (a promissory obligation between two parties, a recognitional nexus) that is partially constituted by the fact that those involved are aware of it.

208).

This might be called the *shared self-consciousness* condition on social integration. However, it subsumes the individual reality condition—shared self-consciousness is the overcoming of separateness between two persons who are worldly and concrete and yet at the same time not only objects to each other—and so the demand for shared self-consciousness *is* the demand for social integration.

Returning to the moral case we can now say that robustly vindicating the moral significance of others (in the way that worries about social alienation demand) involves making sense of others entering our deliberation as more than mere content: they appear not only insofar as they satisfy a general representation of another, but as another voice that joins in. Others are not just objects in the external world that we can aid or thwart, harm or protect, but persons with claims on what we choose, and the capacity to recognize us reciprocally. Their authority over us is structurally like our own, if not always as strong.[21] This authority can be limited or circumscribed, or it may be merely *prima facie*. But if we are to vindicate this basic thought what it cannot be is somehow mediated by our own. That I'm the ultimate practical authority over my actions—that any authority others have is no more than I grant them (even if the reason I grant it to them can be cashed out in constitutivist terms such that I have no choice but to grant it to them)—is a paradigmatically alienated conception of practical authority.

2.5 Conclusion

The concern underlying Williams's alienation objection to consequentialism is that it all but eliminates any recognizable notion of agency. It reduces the agent to a calculator of

---

[21]The ideal of structurally similar authority can be heard in a way that risks reinforcing dominating social relations, especially those of men over women, wherein women are socialized or conditioned to treat the authority of such men as superseding their own in even the most private and personal matters. The point I want to make here, however, is fairly weak: insofar as we regard others as legitimate sources of moral authority, this authority is at least some of the time interwoven with our own agency, but this should not be taken to validate overcorrection in the other direction. Further, there may be good non-ideal reasons to reassert the superiority of individual authority under social conditions where some people's agency has been compromised or subordinated.

outcomes, to a moment of choice, with no commitments, projects, or relations to identify with.

Murdoch makes a similar criticism aimed at the image of an isolated Kantian will, and made vivid in her example of a common but mistaken conception of morality as a 'visit to a shop':

> I enter the shop in a condition of totally responsible freedom, I objectively estimate the features of the goods, and I choose. The greater my objectivity and discrimination the larger the number of products from which I can select. (Murdoch 1971, 8)

The mistake she wants to diagnose involves looking only at the 'needle-thin' moment of choice—understood in terms of the public action that issues from it—as the source of value in the world. According to the simplistic picture, agency is identified as the capacity for choice: an isolable faculty or power residing in a human creature, a "burrowing pinpoint of consciousness, inside, or beside, a lump of being" (Murdoch 1971, 47). Nothing else about this creature is of any ethical significance beyond the will it supports.

Murdoch's point, however, is not just that the life of an agent is more than a discrete series of choices. She understands agency in the first instance as a constant churning process of inner struggle—the diachronic cultivation and transformation of one's unique standpoint on the world. The process aims not just to perfect the self but to prepare oneself to be properly receptive to the reality of persons and things outside of oneself: what I have called the individual reality constraint. Williams offers us a gloss on personal integration, and Railton identifies social integration as an interrelated problem. Murdoch urges the beginning of a vision of social integration, wherein the other shows up in moral theory in a special sort of way, not merely as a thought but as a concrete being. I take this to be part of the story of how we must understand ourselves among others, as Railton insists we must, for personal integration is impossible without social integration. The idea of intra-psychic harmony is not available in detachment from the idea that we are a kind of living being who stands in recognitive relations with others.

The demand to get an individual reality properly into view sets up the problem of social alienation in that it clarifies the nature of what must be overcome. Overcoming it, however, requires making sense of how an individual reality can be not only a concrete object external

to oneself but another subject, to whom one is joined in a mutually recognitive nexus. Getting outside of ourselves and back in again is the work of constituting what Hegel calls "the I that is We and the We that is I". And this is what I think a fuller picture of social integration must involve.[22]

---

[22]I do not mean to suggest that Murdoch conceives of the individual reality of others in objectifying terms, but she has little to say about return-to-self by way of mutual recognition that I have argued is a requirement for making sense of shared self-consciousness. These Hegelian remarks are, however, consistent with what I take her to have in mind.

3.0 Alienation and the Metaphysics of Normativity

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 Sovereigny of Good*[1]

3.1 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

---

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

[2]In 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).

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 §3.2 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 explains. Here I draw on a few familiar literatures and argue that they express different flavors of the same underlying anxiety. In §3.3 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. 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 subject of moral demands, but the resulting *subject-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

---

[3]A 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.

paradoxically seem to need to center both the agent and the subject. 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.

3.2   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*.[4]

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.[5] One thing that they do have in common, however, is that they underlie many of the familiar

---

[4]This 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.

[5]Shamik 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.

challenges to traditional forms of normative realism, and are offered in support of various agent-centered alternatives.[6] 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 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.

3.2.1   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.

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

---

[6] "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 2011*a*, Enoch 2011*b*, 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).

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 mysterious 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.[7]

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)

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).

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.

---

[7]Bagnoli (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 reason*ing*.

It is like a brute fact, except that it is at the same time an a priori and necessary fact. (Korsgaard MS, 2)

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 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.

3.2.2 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,[8] 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.[9] 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

---

[8]Cf. (Schafer 2016)
[9]See Finlay (2007; 2009).

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.

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.[10] 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

---

[10]See (Williams 1973; 1976*a*;*b*).

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.

3.2.3   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 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.[11]

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.[12] Given that realists find it especially difficult to do so, Schroeder notes that "the main divide among realists between

---

[11]See (Schafer 2015a) for a discussion of this point.
[12]Compare (Harman 1977).

reductivists and non-reductivists used to be characterized as the dispute about whether intuitionism is true."[13] In other words, the fate of non-reductive realism depends on realists' ability to defend their rejection of 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 reductive realists and anti-realists tend to see things.

For those who find it mysterious or even occult that we should see normative facts as essentially knowable just because they are, 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.

3.2.4   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

---

[13]Schroeder places himself and other reductive realists, along with "irrealists", in what I'm calling the agent-centered camp, as against traditional realists, or "intuitionists". Assuming that Schroeder intends to include only nonnaturalist realists among the intuitionists, it is true that the worry seems sharpest for them. Non-reductive naturalist realists, however, seem to occupy a fuzzy middle-ground.

The Cornell realists in particular developed their distinctive form of non-reductive naturalism by analogy with the special sciences, holding that moral facts are discoverable in more or less the same way as biological facts. They thus do not have a ready answer to the question of how moral facts are knowable for agents *qua agents*, but nevertheless have more to say about how they are knowable than to just insist that they are. Those who find the epistemic guise of normative alienation especially powerful will likely be inclined to insist that knowledge of normative facts, insofar as it is a form of practical knowledge, must have a more intimate connection to *practical* reason in particular, and not simply be available for easy discovery, but the charge is likely more potent against the non-naturalists.

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.

3.3   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.[14]   As a first pass, the challenge is to ground the essential

---

[14]The 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*b*), Michael Thompson (2004; 2012), Kenneth Walden (2012; 2017; 2018*a*; 2020), and Kieran Setiya (MS). 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.

sociality of morality.[15] 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.

3.3.1 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 1976*a*, 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.

---

[15]Social 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.

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, '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.[16] 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.[17] We 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.'[18] One's practical openness to another isn't separable from their being practically open to one. 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

---

[16]See chapter 2.

[17] 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 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*b*).

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).

[18]To paraphrase John McDowell's (1994) slogan that avoiding what I called perceptual alienation requires epistemic 'openness to the world'.

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 calls "the extremely difficult realisation that something other than oneself is real."

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 I'll offer an example of how it manifests in a set of issues in normative ethics, so as to give a sense of the stakes: the phenomenon of 'directedness'. Recognizing another as the subject of a directed obligation, I will suggest, is a case of practical openness to another, and one a proper understanding of which is threatened by agent-centered metaethics.

3.3.2 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 am important sense the *victim*, not merely the *occasion* of wrongdoing.[19]

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 respon-sive 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

---

[19]This way to refer to the distinction is due to (Thompson 2004, 340).

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.

Whether or not Tarasenko-Struc is ultimately correct his argument offers a vivid example of how things can go wrong with accounting for the sociality of morality—how metaethics can lend itself to a form of social alienation. Importantly, this is a matter of what follows from the Kantian constructivist theory of normativity. The explanation the Kantian provides for the truth-aptness of normative facts, according to Tarasenko-Struc, 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 difficult to see how to recover the status of the subject as the one who stands to be wronged.

One way to put the point is that the *reason* relation that forms the basis of normativity has argument places for the fact that is a reason, the agent for whom it is a reason, the action it is a reason to do, and perhaps a context, but not for a subject. The subject 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.[20] But to account for directedness in terms of merely corresponding reasons

---

[20]Darwall (2012; 2006) uses reactive attitudes and the standing to hold them to explain directedness, but

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.[21]

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 §3.3.4), rather than as a normative relatum. This is just a sketch of an argument and I do not mean to portray it as fatal to any particular metaethical theory if successful, only as an example of a prima facie challenge any metaethical theory will have to address: 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 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

---

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.

[21]E.g. Gilbert (2004) and Wallace (2019).

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.

3.3.3 Social Alienation and Agent-Centered Metaethics

The traditional realist has no trouble accounting for the sociality of morality, and thus avoiding social alienation: whatever the facts are that turn out to play an important role in morality, the realist can say, yes, those are the facts that there simply are. The agent-centered metaethical theories that we saw provide the resources to answer the challenge of normative alienation, however, face extra difficulties. 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. Note that 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. (Here Korsgaard echoes the constitutive role of the hail—*Aufforderung*—in Fichte's understanding of self-consciousness.)

However, there are reasons to doubt whether Korsgaard can accommodate a robust

understanding of the sociality of morality.[22] I don't have the space here to push this objection in any detail, and I don't appeal to it here on the basis of presuming that it's decisive. The point is that 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. My read on things is that 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.

3.3.4 The Solution: Subject-Centered Metaethics

The demand to appreciate the significance of others as external and particular, and as subjects themselves is, I think, 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.

In 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, my

---

[22]Brielfly: first, because 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. Second, the sense in which agency is social for Korsgaard is, so to speak, inside-out. In other words, what it is to be an agent is such as to potentially stand in recognitive relations with others, if any others show up: 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.

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, and though I quibble with the assimilation of this difference to one of grammatical person I am inclined to endorse something like this line.[23]

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 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.

Realists about normative facts can, for example, simply insist that the facts that are most important to social morality are *social* facts, or facts about relations we stand in to one another. Realists about value or the good can insist that these properties play a mediating role in disclosing others to us, that they are part of our experience of others. Wallace, for example, adopts more or less Scanlon's quietist form of nonaturalist realism about the normative, but emphasizes not the reasons an agent has but the moral nexus that joins agents and subjects. The facts about this relation are simply true, and the explanation of their truth does not and perhaps cannot begin with a story about the agent, her desires, her values, or her capacity to reflectively self-determine.

---

[23]Most famously probably Darwall (2006), but see especially 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)

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 *subject-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.

3.4   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 subject-centered metaethics (typically nonnaturalist realism).[24] It is difficult for a theory of normativity to avoid both normative alienation *and* social alienation, but not impossible.

---

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

Supposing that a satisfyingly non-alienated theory of normativity must be in some sense agent-centered *and* subject-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 subject 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.

I've used Kantian constructivism as a stalking horse throughout this chapter, 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 is. 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 subject-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 elsewhere[25] 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 subject-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

_____

[25]See chapter 5.

development of a distinctive practical standpoint on the world is something that we continuously and actively cultivate and revise, and 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.

4.0   Hegelian Constructivism

The human mind is self-conscious in the sense that it is essentially reflective.

<div align="right">

Korsgaard, *Sources of Normativity*
</div>

A human being becomes human only among others.

<div align="right">

Fichte, *Foundations of Natural Right*[1]
</div>

4.1   Introduction

In this paper I want to consider the possibility that agency is essentially social, in that it involves standing in mutually recognitive relations with other agents. The idea is not new— it goes back at least to post-Kantians like Fichte and Hegel—but it has received surprisingly little attention from those working on agency in analytic metaethics. My main focus here will be on the metaethical work that this conception of agency can do, in particular for metaethical constructivism.

I will begin by arguing in §4.2 that Kantian constructivism has been largely misunderstood. The conventional wisdom is that in promising to derive substantive moral requirements from the standpoint of practical reason itself, the Kantian aims to vindicate the objectivity of normative talk without relying on suspiciously non-naturalistic metaphysics. By this standard it's generally regarded as a failure: from any theory of agency thin enough to be uncontroversial, one cannot derive requirements thick enough to approximate morality. This reading is an error. The chief problem for which Kantian constructivism is intended as a solution is not how to get universality without metaphysical baggage.

Instead, as I discuss in §4.2.1, Korsgaard's insight is not just that we should want a way to vindicate the objectivity of normative talk without recourse to mysterious properties, it is that unless we do so our metaethics will leave us alienated from the normativity that

---

[1](Korsgaard 1996, 92), (Fichte 1797, 37, trans. modified)

supposedly binds us. Moral facts are supposed to command us and get a grip on us in a way that facts about numbers, possibilities, and finger snaps do not. They do this by addressing us in the practical standpoint: they matter to us *as agents*.

The central question to which constructivism in fact provides an answer is thus *how does morality get a grip on us as agents?*, and the constructivist answer is that morality arises from agency. The Kantian constructivist answer is that it arises from a Kantian picture of agency as self-legislation.

In §4.2.2 I argue that relying on a *Kantian* theory of agency as the basis for our constructivism leaves us open to another form of alienation: not between an agent and their reasons, but between one agent and another. A workable metaethics must not lose sight of the fact that *you and I*—not moral facts themselves—are able to get a grip on each other.

Metaethics must be able to address both forms of alienation. However, the demands of one trade off against the demands of the other: if the problem in the first sort of alienation is a normativity too distant from agency, the solution is to bring them closer. But in driving normativity inward, we drive agents apart. Constructivism's challenge is to maintain the focus on agency as the source of normativity *and* conceive of agency so that it relates us to others in the right way. The social, post-Kantian view of agency offers a way to do so.

Having motivated the post-Kantian theory of agency, in §4.3 I sketch a picture of agency as constituted by relationships of mutual recognition with other agents. In §4.4 I consider a few different ways this picture could ground a constructivist metaethical theory. One could simply replace the individualistic Kantian theory of agency with this social one and then run a similar kind of constructivism. This would mean leaving in place the Kantian constructivist strategy of *deriving* substantive moral requirements, which, I claim, exposes the theory to many of the same problems. The most promising path instead involves thinking of mutual recognition as *realized* in or *developing* into concrete forms of social life.

4.2 Two Kinds of Alienation

In 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). Constructivism splits the difference, rendering moral facts genuinely objective, truth-apt, and subject to genuine disagreement while naturalistically respectable.[2] Enoch (2011*a*, 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.

Constructivism's claim to strike this balance is based on the fact that it understands normative facts as determined by the outcome of a specified procedure, for example systematic scrutiny of one's values or practical deliberation from the first-person point of view. More generally, it holds that normative facts are determined by what follows from occupying a certain (practical) standpoint.[3] Thus, insofar as there is a fact of the matter about what follows, we can be wrong about our reasons, and they are in that sense objective. However, the standard for determining what they are—what does and what does not follow—involves no appeal to normativity beyond that which is implicitly involved in occupying the practical standpoint, something that we all recognize that we do.

---

[2]In addition to Enoch, see (Shafer-Landau 2003), (Tiffany 2006), (Lenman & Shemmer 2012), and arguably (Smith 1999) and (Gibbard 1999) for this understanding of what motivates Kantian constructivism. 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 as I will argue shortly it's not the central problem for which her constructivism is supposed to be a solution.

[3]For the former, see e.g. (Darwall *et al.* 1992) and (Enoch 2011*a*), who calls this version "canonical". For the latter see (Street 2012).

The trouble with this way of motivating constructivism is that its naturalistic credentials appear to rely on characterizing the practical standpoint in such minimal terms that our occupying it is uncontroversial, but it is difficult to see how such a minimal characterization of the practical standpoint can provide the resources to meaningfully constrain the results. Though some Kantians remain optimistic, something like a consensus has emerged that from a theory of agency thin enough to be uncontroversial, nothing substantive can follow via a procedure of rational scrutiny.[4]

This bleak assessment of the prospects for metaethical constructivism, however, relies on a misunderstanding of the question for which it's an answer. Returning to *The Sources of Normativity* with this picture in mind, one is likely to be struck by Korsgaard's lack of evident concern for "queerness," or anything else so metaphysical. In fact the motivating concern is that reasons for action are reasons *for agents*, and cannot be understood as entities like any other.[5]

4.2.1 Constructivism and Normative Alienation

In the canonical statement of Kantian constructivism, Korsgaard (1996) argues that more traditional forms of realism leave an explanatory gap between the normative facts and the agents for whom they are reasons. They allegedly hold that "we have normative concepts because we've spotted some normative entities, as it were wafting by" (44). The task for metaethical constructivism is to understand reasons for action and our capacity to act for reasons in such a way that their harmony is non-accidental. Constructivism proposes to explain normative facts in a way that accounts for how they address *us* as agents and

---

[4]The dialectic that leads to this consensus begins with Enoch's (2006) attack on the possibility of constitutivism. Ferrero (2009) defends the possibility of such constraint, generated by an understanding of a domain closed under its own characteristic activity. We arrive at a dilemma: constitutive constraint is possible only where it's too minimal to ground substantive morality (Tiffany 2012; Schafer 2015*a*; cf. Baiasu 2016).

[5]Street's (2006) Humean constructivism is arguably motivated by something like the theoretical desiderata laid out above. Though her way into constructivism is via the epistemology of normative facts she does appear to have something like these metaphysical concerns in mind. In my view Korsgaard offers a more compelling way into constructivism, and in what follows I will treat her arguments as characteristic. Insofar as Street's epistemological concerns full under the epistemic guise of what I call normative alienation below, I suspect that she is ultimately closer to Korsgaard on this point than it sometimes appears.

feature centrally in the exercise of practical reason.[6] 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)

Simply ascribing to certain facts a 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)[7]

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.[8] One way to motivate the worry a bit more is to place it in the context of a more general concern one might have about theories of normativity: that part of the task for such a theory is to explain not only how it could be that any normative facts were true, but *what they have to do with us.* Korsgaard's 'normative problem' is an instance of a more general anxiety felt by many philosophers, which I call a fear of *normative*

---

[6]Bagnoli (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 reason*ing*.

[7]Cf. (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.

[8]See Scanlon (2014, 44): "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."

alienation. The need to satisfy some version of this desideratum underlies many of the the familiar challenges to traditional forms of normative realism.[9]

The threat of normative alienation appears in different guises: that normative facts could fail to be motivating, that they could fail to be rationally authoritative, and that they could fail to be identifiable. The first of these is probably the most familiar, and associated with Williams's (1979) 'Humean' challenge to motivational externalism about reasons. The thought is that it is a part of the very idea of something's being a reason for one that it has the ability to motivate her.

The second and third guises of normative alienation are associated with metaethical constructivism. Constructivism explains the reasons we have in terms of the very capacity to recognize ourselves as having reasons, and to act on that basis. Further, constructivism explains why normative facts are knowable for agents in virtue of being agents: because it is the exercise of practical reason that determines what we have reason to do.[10]

In each case we see a similar structure: an anxiety 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 expressions of normative alienation. Morality cannot be alien to us.[11][12]

---

[9]Dasgupta (2017) identifies the first and second guises of normative alienation (as I characterize them in the next paragraph) as versions of the same desideratum, though he does not include the epistemic challenge or characterize them as a threat to normativity's relation to agents.

[10]See (Schafer 2015*a*) for a discussion of the role that these last two desiderata play in motivating constructivism. On the epistemic challenge specifically, see as well chapter 11 of (Schroeder 2007).

[11]Contrast Blanchard (2020): following up on Parfit's famous claim that if non-naturalism is false then nothing matters (and he and his colleagues have wasted their lives), Blanchard tries to uncover the anxiety that makes realism seem important and hard to give up. In my view it is answering to the threat of normative alienation that is more important.

[12]But what of the standard by which what I've called the 'received view' proposed to judge constructivism? In the decades since Korsgaard's *Tanner Lectures* the metaethical terrain has shifted. Non-naturalist forms of realism have become increasingly quietist, or 'non-ontological', and their proponents today tend not to think that their views come with the kind of metaphysical consequences that threaten the modern scientific worldview. Meanwhile, increasingly sophisticated versions of expressivism claim to have recovered enough objectivity to say everything we want to say about normative facts. To the extent that expressivism and realism remain competitors the stakes are less commonly defined in terms of balancing naturalism and objectivity. This is not to suggest that there exists a consensus that these issues are no longer live, but I am happy to ally myself with those who claim to have moved past them, as should all constructivists.

Prominent quietistic non-naturalists include Scanlon, Dworkin, and Parfit; expressivists who claim to

4.2.2 Social Alienation

There is another sort of alienation that should worry the metaethicist, which I call *social alienation*. Roughly, the idea is that much of morality involves getting a grip on each other, agent to agent. But a theory such as the Kantian constructivist's, which pulls the requirements of morality out of the practical standpoint of the self, runs the risk of erasing this distinctive grip, giving us at best an indirect way of recognizing other people. At least a part of morality is irreducibly social in that it is fundamentally constituted by relations that concrete persons stand in to one another, and in conceiving of agency in individualistic terms constructivism risks being unable to explain this sociality.

The threat of social alienation in metaethics has gone largely unrecognized, and I do not have the space here to develop the objection in any detail. However, we can see some of its intuitive force by beginning with a more familiar version of the worry: the fear that on a perfectionist reading virtue ethics turns morality into an exercise in self-improvement. The thought is that even in the realm of what we owe each other (justice, or *dikaiosune*), all of the normative 'oomph' derives ultimately from the need to perfect oneself—to be excellent. As Wallace (2019, 46) puts it, "at the level at which normative requirements are explained, the interests of other people enter as occasions for the realization of virtue, rather than direct sources of requirements on the virtuous agent." If this form of perfectionism is right it amounts to denying that we really do owe anything to one another in the first place.

Korsgaard herself glosses a similar objection Gauthier's 'Hobbesian' attempt to construct

---

have entitled themselves to treat normative claims as objective include Blackburn, Gibbard, and Price. See (Kremm & Schafer 2017) for a survey of some of these views. Golub (2017) and Parfit (2016) argue that expressivism and realism need not be seen as conflicting, Further, the claims made by metaethical constructivism are ultimately orthogonal to these debates—as Hussain & Shah (2006; 2013) argue they will be compatible with any plausible semantics or metasemantics of normative terms. (I disagree that constructivism is therefore not a metaethical theory at all, but I think they are correct that it is compatible with a variety of realist and expressivist views.) Korsgaard (2003, 188) herself suggests a similar kind of compatibility:

> If constructivism is true, then normative concepts may after all be taken to refer to certain complex facts about the solutions to practical problems faced by self-conscious rational beings. Of course it is only viewed from the perspective of those who actually face those problems in question that these truths will appear normative. Viewed from outside of that perspective, those who utter these truths will appear to be simply expressing their values. Realism and expressivism are both true in their way.

moral reasons, which are essentially public, out of reasons of self-interest, essentially private. She notes that:

> If reasons *were* essentially private, consistency would not force me to take your reasons into account. And even if it did, it would do it in the wrong way. It should show that I have an obligation *to myself* to treat you in ways that respect the value which I place on you. It would show that I have duties with respect to you, about you, but not that there are things I owe *to* you. (1996, 134)

In my view the real force of the objection comes from the implicit move from the normativity of first-person practical reason to the normativity of social relations. Korsgaard's emphasis on the private/public distinction can obscure this, but the last sentence quoted above captures what is really at stake. There is a difference between having a duty that features another as part of its content, and having a duty owed to another.

The technical term for this distinction is that some duties are *directed*. If one agent is under a duty directed toward another, the other enjoys a special standing as the holder of a right to the first's performance. When an agent violates a directed duty she does not merely do something wrong but *wrongs someone* in particular: the one to whom the duty is owed. As Thompson (2004, 340) puts it, the one who is wronged is the *victim* and not merely the *occasion* of wrongdoing.

Directed duties are, if not coextensive with morality, at least an important part of it, and a key instance of the kind of sociality that I'm interested in. They reflect what Wallace (2019) calls the 'moral nexus' that joins concrete persons, equally real: a fundamentally social relation that cannot be explained in terms of arrangements of individual normativity.

A particularly dramatic way for a theory of metaethics to fail, then, would be for it to render our moral requirements in terms that ultimately collapse back in on the self. Tarasenko-Struc argues that Korsgaard is committed to precisely this: "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). The explanation the Kantian provides for the truth-aptness of normative facts, according to Tarasenko-Struc, entails that those facts are ultimately facts about how we stand in relation to ourselves, and not about how we stand with respect to others. Just as with the simplistic virtue theorist, all of the normative 'oomph' of directed

obligations derives ultimately from the individual requirements of self-constitution.

What's important here is not whether any one of these objections is successful as I've reconstructed them, but that we can see them pointing toward a distinctive way of going wrong in metaethics: the possibility of what I have called *social alienation*.

Kantians like Korsgaard hold that morality involves relations of mutual accountability with other people. One thought might be that the Kantian has therefore dodged the threat. But the Kantian thinks these relationships of accountability can be derived from the idea of the agent as self-legislator, rather than by treating them as *sui generis*. There is widespread skepticism that this derivation can be carried off, but even if it can, in conceiving of relations of mutual accountability as mediated by more basic self-relations, the Kantian risks social alienation. Since Kantian constructivism was in the first instance an answer to how agents recognize moral reasons, it leaves us with only indirect ways of recognizing one another.

My suspicion—and for now it will have to remain merely a suspicion—is that something like this worry underlies much of the dissatisfaction with Kantian constructivism. As I noted above, the worry has not been explicitly addressed in these terms but I think it offers a plausible diagnosis of some of the skepticism faced by Kantian constructivism. Even if the Kantian can establish that agency involves the lawlike form of the will's exercise, some object, it simply cannot explain how anything recognizable as morality can follow from this alone. My diagnosis is that agency, so understood, is a reflective self-relation, and morality involves relations with others. No amount of reflection can bring them into view in the right way if they are not there already.[13]

---

[13]Korsgaard has explanatory aims aligned with what I am suggesting, and indeed she holds that there is a role for sociality in, if not agency as such, at least its characteristic exercise: reflecting on essentially public reasons, responding to the call of another (here echoing the constitutive role of the hail—*Aufforderung*—in Fichte's understanding of self-consciousness).

However, there are two reasons to doubt whether Korsgaard can accommodate a robust understanding of the sociality of morality: first, because it's not clear that her theory of agency is up to the task of grounding the kind of sociality she appeals to in lecture 4.2 of *Sources*. Second, the sense in which agency is social for Korsgaard is, so to speak, inside-out: what it is to be an agent is to be such as to potentially stand in recognitive relations with others, if any others show up because 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. But for Fichte or Hegel agency is outside-in: being an agent requires already standing in recognitive relations with actual, concrete others. For Korsgaard the *Aufforderung*, and the recognition it presupposes, is constitutive of various normative facts but not of agency itself. For Fichte, it is through the *Aufforderung* that we become agents (see especially (Fichte 1797, 35) and see (McNulty 2016) for more discussion).

To the extent that Korsgaard does turn out to have the explanatory resources to take sociality seriously then what follows can be seen as an elaboration on a social reading of Korsgaardian agency. There is a

4.2.3 The Dilemma

In order to explain the grip that reasons have on us constructivists center the individual agent in their account of normativity. The dilemma is that metaethical accounts that center the individual agent typically 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).

This dilemma is not specific to constructivism. Those moved by the worry about normative alienation tend to gravitate toward more *agent-centered* approaches to metaethics (constructivism, subjectivism, Humeanism). If the problem is that normativity could be too distant from agency, the solution is to bring it closer. But doing so risks obscuring the social relations that constitute morality's core.

The problem of social alienation suggests taking a more *subject-centered* approach to metaethics, one that emphasizes, in its explanation of normativity, those with whom morality is principally concerned. The problem of social alienation is relatively easy to address for a realist: facts about relations between agents and subjects, or facts about subjects themselves, are the normative facts that there really are.

To navigate the demands imposed by worries about both kinds of alienation, constructivism will need to *both* maintain the focus on agency as the source of normativity *and* conceive of agency in such a way that it relates us to others in the right way. This gives us reason to look for a different theory of agency.

A theory of agency that builds sociality in from the beginning has a chance of making sense of how the irreducible sociality of ethical life is possible. Inspired by Fichte and especially Hegel, who are animated by a similar concern with Kant's moral philosophy, it's this suggestion I now want to develop. Rather than appealing to law as the constitutive form of agency, my proposal appeals to *mutual recognition* of agents with one another, insofar as they are agents. I will call this the *post-Kantian* theory of agency. In the next section I

---

historical analogue in the relationship between Hegel and Kant: some interpreters take Hegel to be going beyond Kant while others take him to be working out in greater detail and explicitness a sociality that is already there in Kant, albeit less obviously (see e.g. *KrV*, A739/B767: "The very existence of reason depends upon this freedom, which has no dictatorial authority, but whose claim is never anything more than the agreement of free citizens, each of whom must be able to express his reservations, indeed even his veto, without holding back.")

elaborate this theory of agency, which is better suited than Korsgaard's to address the social problem of alienation, and in §4.4 I return to the broader question of how to explain morality based on a theory of agency alone, to see what a constructivist can do with post-Kantian agency.[14]

4.3   Mutual Recognition

Like the Kantian, the post-Kantian takes agency to be the activity of acting reflectively on the basis of considerations that one endorses as reasons. But unlike the Kantian, the post-Kantian holds that a creature cannot engage in the activity of agency alone. Being an agent requires standing in mutually recognitive relations with other agents: to be an agent is to be and to understand oneself as one among many.

Mutual recognition has historically been understood in a number of different ways. We do not need to settle on a single conception of mutual recognition to see the general direction of post-Kantian view, and to appreciate its potential virtues as an amendment to Kantian constructivism. But it will be helpful to consider how a mutual recognition theorist might answer what I take to the be three key questions facing any account of mutual recognition: what it is to recognize another; what, if anything, is the 'content' of recognition (i.e. what do we recognize another as); and how should we understand the mutuality of mutual recognition? I address each of these questions in turn, in the process illustrating what a mutual recognition-based theory of agency can look like.[15]

---

[14]Perhaps metaethical constructivists need not even work with a theory of agency in such abstract terms that it is suited to characterize Martians, robots, or artificial intelligences—agency *as such*. What is at stake is our understanding of ourselves, and what the constructivist needs is a theory of agency that plausibly captures what it is that we are like. Cf. Walden (2018*b*) who argues similarly that constructivism need not begin with practical reason as such, when what matters is practical reason as we exercise it. In the rest of this paper, however, I will follow the post-Kantian philosophers whose proposals I am attempting to appropriate in treating theirs as a theory of agency as such, and not just human agency.

[15]Here I will be preoccupied with offering a sense of what post-Kantian agency might look like, on the grounds that it has an important theoretical role to play as discussed in §4.2, and I will not have the space to argue for it on the merits. For various conceptions of mutual recognition associated with Fichte and Hegel, and arguments to the effect that it is an essential feature of self-consciousness and human agency, see (Fichte 1797, Hegel 1807, Neuhouser 1986, Wood 1990, Franks 1996, Pinkard 1994, Honneth 1996, Brandom 2007, Pippin 2008, Schmidt am Busch & Zurn 2009, Clarke 2009, McNulty 2016, Brandom 2019). Darwall (1977) and Scanlon (1998) use a different, more morally substantive sense of 'mutual recognition'.

One note of clarification: the picture I am about to reconstruct is originally of what is involved in *self-consciousness*, and not specifically agency. In the German idealist tradition agency is understood as the capacity for action, which is differentiated from mere behavior by the presence of self-consciousness. Inclinations become potential reasons when we become reflectively self-aware, just as appearances become potential judgements. In other words, mere animal locomotion and mere animal perception are transformed, respectively, into *action* and *experience* through the awakening to oneself that makes a creature self-conscious.[16] As Rödl (2007, 105) puts it, "It is the principle thought of German Idealism that self-consciousness, freedom, and reason are one."

4.3.1 What Is it to Recognize Another?

Recognition, in the relevant sense, is a normative attitude. It is an *attitude* in the sense of a position, posture, or stance: an orientation with respect to something. In other words, recognition is not just a special kind of belief or desire, but a more general way of being-toward its object.[17] As Wittgenstein (1953, iv.22) famously says (of a friend), "My attitude towards him is an attitude towards a soul. I am not of the *opinion* that he has a soul."

It is a *normative* attitude in that it is a distinctively practical kind of orientation, one that manifests in acting toward its object as having a certain kind of practical significance. In characterizing normative attitudes this way I do not mean to deny that they are cognitive. Though they are in the first instance practical, insofar as a normative attitude at least implicitly classifies its object as significant in a particular way it has an implicit conceptual structure.[18]

---

[16]See (Boyle 2012)]. One can find this view as well in a reading of Aristotle's *De Anima*.

[17]This terminological choice serves several purposes. First, the term is vague enough that in defining it the way I am about to I will not be competing with long-standing associations. Second, it has a few intended secondary meanings that support the use I plan to make of it: the attitude of an airplane in the air—i.e. its orientation in space—and the attitude that one displays to others, by a look, a gesture, an expression. The attitude of recognition involves both orienting oneself in normative space (that is, orienting oneself with respect to the reasons there are; the normatively significant features of one's world; the practices, institutions, and social formations that compose a distinctive form of ethical life), and one's outward-looking, social face, that which we show to others.

[18]Compare Dennett's (1987) notion of the 'intentional stance' or Langton's (1997) discussion of 'treating someone as a person'. An important sense in which the attitude of recognition is like a Dennettian stance is that stances are, for Dennett, crucially embodied. As Kukla (2018a) emphasizes, the intentional stance is in many respects like the stance of a boxer: a posture, a readiness to act, a physical awareness of the presence

The key features of recognition are that it:

1. features in or structures practical deliberation;

2. manifests in patterns of behavior and other forms of embodiment;

3. attributes a status to its object; and

4. can be made explicit in judgment, but is not necessarily so

In general it is natural to use "recognition" in this sense when we speak of recognizing the authority of a person or institution, or recognizing a right or claim as binding or legitimate. What we recognize is not a fact, a property, or an object, though of course, we do in some sense recognize an object under an aspect, or as-of-a-kind.[19] But even insofar as what we recognize is an object under an aspect, the aspect under which we recognize it is not a descriptive property, but a status, a normative significance. Thus, accordingly, to recognize something as authoritative is centrally be disposed to treat it in a certain way, to practically orient oneself with respect to it in light of this normative significance, to deliberate accordingly, and not necessarily to explicitly judge *that* it is authoritative.

Consider the example of recognizing something as holy. To recognize an object as holy might be to use it in certain rituals, to handle it with greater care, or to stay away entirely. Recognition of holiness is also embodied in ways beyond treating something as holy in deliberating and acting with respect to it: recognition of holiness is embodied in experiencing anxiety when the holy object is mishandled or at risk of damage, in feeling calm or awe in its presence (as appropriate), in having the bodily skill required to treat it as holy, in holding oneself in a certain posture in its presence, and in being habituated to a set of ritual practices.

Such recognition is typically accompanied by the belief *that* the object is holy, but the recognition of its holiness is neither equivalent to nor exhausted by having the corresponding belief: such recognition is in fact consistent with never making explicit to oneself the belief that the object is holy—one could even lack competence with the concept of holiness, or, in peculiar cases, explicitly deny that the object is holy while nonetheless consistently treating

---

of another, and a way of holding oneself that is essentially bound up in an activity that one is involved in with another.

[19]As when, according to some theories of perception, we see a table-as-table, though perhaps can only judge *that* there is a table, or that the table *is* a table. See Sellars (1968, ch. I) and McDowell (1998*a*).

it as holy. [20] Recognizing something as holy is, thus, an overall practical orientation toward it as something holy, and not identical to the mere belief in the object's holiness.

4.3.2   What do we Recognize the Other As?

Insofar as recognizing another classifies them a certain way, granting them at least an implicit, practical significance, it has a particular content. In a case where recognition never rises to the level of explicit judgment, there is nonetheless some concept in the offing that gives unity to the behavioral manifestations of recognition, a concept that would feature in a judgment making the recognition explicit. When we recognize another, in the sense relevant to post-Kantian agency, we recognize them *as* something.

Mutual recognition, however, does not classify its object as having a merely normative significance, like the significance of holiness, which demands a certain discrete pattern of treatment. Rather, mutual recognition classifies its object, the other mutual recognizer, as a recognizer, thereby granting it a *metanormative* significance. It classifies the other as something that can attribute normative significance. In other words, to recognize another as a recognizer is to recognize that its relationship with its environment is significance-granting.

If I recognize you as a recognizer of the holy, and I see you handling an unfamiliar object gingerly, using it to perform an unfamiliar ritual, I will recognize what you are doing as treating it as holy, and I will think 'ah, there appears to be a holy thing that I missed', and then go on to treat it as holy myself. Of course, I may come to disagree—I didn't miss the object's holiness, it is you who are confused. But in recognizing you as a recognizer of the holy I am at least defeasibly disposed to treat as holy that which I see you treat as holy. I treat you as an authority in matters of holiness, though perhaps an imperfect one.

This helps to explain why I have not opted for the more familiar suggestion that the ur-ethical relation is the (mutual, perhaps) recognition of another's humanity. The idea of

---

[20]Such cases might go under the heading of 'alief' (Gendler 2008) or 'in-between belief' (Schwitzgebel 2001). Given the ways that belief is theorized in moral psychology, I am reluctant to characterize recognition as a belief, though on a certain understanding of belief (e.g. Schwitzgebel 2002) I might just as well have. However, stipulating a strict distinction between cognitive and conative states, where the former are motivationally inert, recognition is not a cognitive state in that sense, though does have a kind of cognitive content, as recognizing something as-of-a-kind at least implicitly attributes to it a particular significance (e.g. the significance of legitimacy).

recognizing the humanity in another has a long history of being drafted into doing serious moral work, and it might have seemed like a natural one to reach for here: part of what I'm suggesting is that the kind of sociality that shows up later in the Kantian story[21]— often as the formula of humanity—ought to be promoted to a more foundational role in our understanding of agency. Thus one might simply propose that recognition of the humanity of others is itself constitutive of agency, rather than something that follows from and depends on being an agent, the constitutivist account of which makes no reference to the humanity of the other.

But if we understand by "humanity" something like holiness—a normative significance that demands a pattern of special treatment—it will require at once too much and too little of the ur-agents recognizing it. It requires too much because on the most natural sense of the term, and the one that tends to be presumed when it is relied on to help answer first-order moral questions, it builds too much into the constitutive constraints on agency.[22] It requires too little because it does not deliver the mutuality we have been chasing, or if it does, it does so at best coincidentally. My recognition of another's humanity need not have anything to do with their recognition of mine.

Now, it's possible that the sense of 'humanity' in use in Kant's *Groundwork* may be better understood as a metanormative significance like that of a recognizer—in respecting the humanity in others, Kant commands us to act in light of the recognition that others are reflective, self-determining agents like ourselves, capable of changing the normative significance of the objects of their own wills by taking them to have different normative significance. This would bring the Kantian story closer in line with the recognition I have in mind.

However, even stipulating that to recognize another as human involves something like recognizing them as a recognizer, that won't be enough to render the mutuality of mutual recognition more than coincidental. For mutual recognition to be a relation that plausibly plays a role in getting agency up and running, in explaining our own reflective relations with ourselves, it must be a relation through which we are awakened to ourselves as individuals, a

---

[21]See n.13.

[22]Conversely, it's not clear that when the concept of humanity is recruited to do substantive moral work it is up to the task. Historically it is not uncommon to find those who, like early American Jesuits baptizing those they enslaved, recognized others as human in the context of great injustice. Cf. (Manne 2016). This is not a kind of work I expect mutual recognition to do (see §4.4).

relation that is bound up with our capacity to recognize ourselves. If you appear to recognize something as holy but it doesn't seem holy to me, that makes you a defective recognizer of holiness, nothing more (necessarily) than an instrument that delivers occasional false positives. That won't be enough to force open a space between my immediate judgments of holiness and my reflective role as an arbiter of such judgments. What generates our capacity to achieve reflective distance from ourselves, according to the post-Kantian view, is the distinctive kind of pushback we get from another we recognize as a recognizer of recognizers, and as recognizing ourselves.

This gets to the fundamental disagreement between the Kantian and post-Kantian conceptions of reflective self-consciousness: the Kantian thinks that one can first achieve reflective distance from one's own inclinations, and then take oneself to be a recognizer of recognizers, thereby attributing that same power to others, without first coming to recognize oneself as recognized by a recognizer like oneself.[23] The post-Kantian insists that reflective distance in the first place presupposes the kind of distinctively normative pushback one gets from others with whom one stands in mutually recognitive relations, not just the kind one gets from a malfunctioning recognition machine. On the post-Kantian view we are awakened to ourselves, and thus achieve reflective distance from our own inclinations, by recognizing that we are being recognized, and thus recognizing the other reciprocally.

4.3.3   On *Sui Generis* Mutuality

Thus far we have considered the 'recognition' component of mutual recognition, and asked what we recognize each other as. In what is sense must this recognition be mutual? To be awakened to oneself through recognizing another as recognizing oneself, doing so cannot simply be a matter of both recognizing each other symmetrically; there must be mutual recognition of the fact that we have recognized each other.

Consider a variation on John Perry's famous case of the shopper, who, both leaving a trail of sugar behind himself and searching for the person spilling sugar all over the store, eventually sees himself in a mirror and takes himself to be looking at "the messy shopper

---

[23]In Korsgaard this appears under the guise of the 'second-person within'—see n.13.

going up the aisle on the other side," unaware that it is himself that he sees (Perry 1979, 12). Similarly, I might spy, across Schenley Park, one of my friends, apparently looking not at me but into a restaurant. Meanwhile, she notices me in the reflection of the window, but because of the reflection's angle does not realize that I'm looking at her. In this situation we both recognize each other (in a more ordinary sense of "recognize" than the technical one at work here), but neither of us is aware of being recognized. Our symmetrical recognition, in this case, does not constitute the kind of shared experience of one another that characterizes properly *mutual* recognition. Properly mutual recognition, as opposed to merely symmetrical recognition is shared, not paired.

Properly mutual recognition, of the kind involved in the development of agency, shares this structure. It's *sui generis*; it cannot be constructed by stacking individual recognitive states one on top of the other, because it's essentially self-conscious of its own mutuality. The kind of recognition we are after is not just a matter of both parties having matched recognitional attitudes towards one another, but a matter of standing in a mutually recognitive relation.

The implicit conceptual or propositional content is not just that another is a human, a person, an equal—mutual recognition places her with respect to oneself. Each of us does not only recognize what the other is like but also how they stand to us and thereby we to them. They are thus not (only) the content of our awareness, as they could be while the independent reality condition were met, but sharer with us a recognitional nexus. This is not to say that every experience of another has this structure, only that it's a structure that a satisfying picture of agency as a socially-embedded phenomenon must have room for.[24] A central part of how we understand ourselves is as among others.

One's practical relation to another and the other's practical relation to one are not independent but two ways of specifying the same self-consciousness relation-to-other. Consider the shared experience of a sunset.[25] Two friends are looking at a sunset together. What's

---

[24]Even though it is not the structure of every experience of another, it is perhaps the structure of paradigmatic moral relationships: the most fully realized form of self-other relation. At the least it is one important form relating to others can take. Setiya (MS) and Korsgaard (2018) argue for different reasons that important moral relationships are often asymmetrically recognitive, Setiya in the case of love and Korsgaard in the case of non-human animals. What I say here is intended to be consistent with both.

[25]This discussion draws on (Velleman 2013).

happening is more than that one is looking at the sunset and the other is also looking at the sunset. It is part of the experience for each that it is shared; it is an experience of looking at a sunset *with a friend*. But this in turn is not just a matter of one having an experience that includes in its content the fact that another person is also having that experience. The experience is of looking at a sunset with another who is also having the experience of looking at a sunset together.

Recognition is thus not the apprehension of a fact about another but a fact about *us*, the apprehension of which partially constitutes its object. When one experiences a mutually recognitive interaction with another there are not two symmetrical recognitional states but a single, shared experience of recognition, in which one's recognition of another is a part of the other's recognition of them, and vice versa.[26]

4.3.4 Mutual Recognition and Alienation

We are now in a position to understand the foregoing not as the logical construction of mutual recognition, but as its reconstruction. In other words, mutual recognition is not the product of adding layer on top of layer of normative attitudes, but it can nevertheless be explicated partly by decomposing it into these elements. In doing so, we do not see the ingredients that, when combined, yield a mutually recognitive relation, but the key aspects of a relation that is by its nature *sui generis*.

I considered this picture of mutual recognition as a plausible candidate for a theory of agency, on the hypothesis that it might give the constructivist a way to chart a course between our two forms of alienation. But if mutual recognition is our picture of agency, don't we encounter a puzzle: to *be* an agent must we be mutually recognized as one?

This apparent paradox—that we must recognize one another as agents in order to be agents—is meant to be virtuously circular. The idea is not that we assemble these relations piece by piece, but rather that by the time we are capable of reflecting on our own reflective

---

[26]There is an echo here of what is sometimes called "Hume's circle" wherein one cannot make a promise without using the concept of promising, and thus the term is partially constituted by the practice to which it refers, and the practice presupposes the use of the term (see Anscombe 1978). The common feature of the two cases is a state of affairs (a promissory obligation between two parties, a recognitional nexus) that is partially constituted by the fact that those involved are aware of it.

capacities we find that we already implicitly recognized one another. However it is that the process happens, once we achieve full-blown agency—the capacity to achieve reflective distance from our inclinations, to see them as (potential) reasons for action—we find ourselves as each one agent among many.[27]

To recognize others as standing in this relation to one involves granting them the authority to interpret, grant significance to, and determine one's actions, and, conversely, recognizing their behavior as action like our own, which depends on our recognition in the same way that our action depends on theirs.[28] Agents, as such, recognize other agents as co-constituteors of a shared space of reasons, as co-determiners of the norms that structure a shared practical reality (what Hegel calls 'ethical life'—*Sittlichkeit*). We orient ourselves in normative space with respect to one another as poles of authority. Mutual recognition thus structures *all* of our practical possibilities, and not just our deliberation and action with respect to a certain object.

Conceiving of agency in terms of mutual recognition puts us in a position to resolve the tension between the demands of normative and social alienation. It addresses social alienation by placing sociality at the ground floor of agency, so that self-conscious relations with others are not something that a constructivist theory will need to go on and explain. As I will argue in the next section, when appealed to in the first phase of a constructivist metaethical theory it will provide the resources to address normative alienation by explaining how the normative order produced by the constructivist theory is properly connected to the capacity we have as agents to respond to it.

4.4   Hegelian Constructivism

Constructivists hold that the best way to explain the tight connection between morality and the agents it binds is to think of morality itself as arising from agency. Thus a

---

[27]Perhaps, as McDowell (1994) suggests, our eyes are 'opened to the space of reasons' by our upbringing, or, per Wittgenstein (1969, ¶141), light dawns gradually over the whole...

[28]This authority-granting could be understood in terms of 'deontic scorekeeping'; see Brandom (1994) and Lewis (1979).

constructivist theory will need to include both of the following components:

1. An account of the practical *standpoint* (agency)
2. An account of the determining/grounding/explanatory relation between (1) and a *normative order* (viz. morality)

Kantian constructivism conceives of the practical standpoint in terms of the lawlike form of the will, and its relation to morality as entailment: a particular moral order is determinately generated by explicating what follows from occupying the practical standpoint.

I argued in §4.2.2 that drawing morality closer to individual agency trades off against vindicating the fundamental moral significance of social relations, and in §4.3 I elaborated a social theory of agency that provides the resources to explain the significance of social relations to agency itself. A natural thought might be to try to tell the same kind of story about how agency so conceived entails a moral order with social relations at its core.

4.4.1 Deriving Morality from Mutual Recognition?

If being an agent already involves implicitly recognizing others as having some basic authority to co-constitute a shared form of life it plausibly follows that we owe them some basic form of respect, from which a recognizably moral order could perhaps be derived. This is the general shape of a view like Darwall's (2006), and for the sake of convenience I will refer to any theory that appeals to post-Kantian agency and what follows from it as *Fichtean constructivism*.

I do not take the following concerns to be decisive, and for all I've said the core argument of this paper could be construed as offering support for Darwall's project, with which I have considerable sympathy. However, I am not persuaded that Fichtean constructivism is the way forward. While adopting the post-Kantian theory of agency addresses the dilemma I discussed in §4.2, in relying on the same strategy as the Kantian for grounding morality in agency it risks inheriting some of the Kantian's other problems.

As I suggested above much of the skepticism philosophers have regarding the Kantian constructivist project may reflect an implicit concern with what I'm calling social alienation: that morality involves others in a way that agency as self-legislation cannot explain. How-

ever, Kantian constructivism faces a number of powerful objections that target its broader explanatory strategy, rather than the details of its account of agency. Broadly speaking, the worry is that in placing on a theory of agency the explanatory burden of accounting for the determinate content of a normative order it stretches its explanatory resources too thin. The worry I want to deal with here is that this objection will apply to any constructivism, no matter its theory of agency.

This skepticism is anticipated by Hegel, in his infamous 'empty formalism' objection to Kant. Street (2012) echoes Hegel in arguing that, setting aside the various moves in Korsgaard's 'regress argument', pure practical reason is just not well-suited to deliver anything so concrete as a substantive morality. If the worry about social alienation targets the Kantian's account of agency (1 in the schema above), the empty formalism objection and its more recent descendants target the Kantian account of how agency is related to morality (2): that the latter is entailed by the former. Call this a *deductive* account of the agency-morality relation.

4.4.2 Developmental Constructivism

Deductive constructivist theories, like the Kantian and Fichtean, hold that determinate normative facts can be derived from an account of agency. A more modest way to conceive of the agency-normativity relation is in *developmental* terms, that is, that agency develops into or is actualized in ways of being an agent that are not entailed by the formal features of abstract agency, but which nevertheless count as ways of being an agent only insofar as they realize the abstract form. An example of a developmental relation of this kind might be found in the explanation a capacity offers for its actualization: my speaking English is (at least partly) explained by my capacity for language-use.

The suggestion in this context is that agency develops into concrete realizations in something like the same way. Street's (2010, 2008) Humean constructivism may serve as a useful example: she holds that the content of the practical standpoint is supplied from without—in the first instance, by an agent's normative attitudes, but these in turn have a causal-historical basis—and that what it is to be an agent is to be a particular, concrete agent, essentially

65

embedded in an evolutionarily, socially, historically conditioned context. To be an agent is to be a realization of abstract agency, which develops into different concrete forms in different human creatures as their evaluative attitudes are systematized.

Given the choices between individualistic and social accounts of agency and between deductive and developmental accounts of the agency-morality relation, we arrive at the following possibilities:

	individualistic agency	social agency
deductive	Kantian (Korsgaard)	Fichtean (Darwall)
developmental	Humean (Street)	Hegelian (me)

Table 1: Four kinds of constructivism

To do more than just identify an opening in logical space, I need to offer a Hegelian account of the developmental agency-morality relation—an account of what mutual recognition develops into, and how. I don't have the space here to do so in any detail, but I want to at least gesture toward what I think is a promising approach to metaethical constructivism, one that addresses the dilemma evinced in §4.2 while, by scaling back its explanatory ambitions, it avoids the pitfalls of Kantian constructivism's account of the agency-morality relation.

The Hegelian proposal is that mutual recognition develops into different concrete forms of social life, which in turn provide the actual content required to fix the normative facts. The abstract form of constitutively social normativity suffuses and animates the social formations, practices, habits, relationships, institutions, and all the rest of the 'ethical substance' that composes a form of life.

To be an agent is to stand in relations of mutual recognition with other agents, and what is normative for agents is not simply a matter of how they do or how they must reason as individuals. To be an agent is always to be a particular kind of socially and historically

conditioned agent, one whose actual agency depends on the social conditions in which one's agency is realized.

If the sense of 'construction' at work in deductive approaches to constructivism involves an ideal construction in thought, the term takes on a more literal meaning in the developmental context: for the Hegelian, we determine what it means to recognize one another, what recognition consists in, by actually constructing forms of life for ourselves in historical time.

4.5   Conclusion

Hegelian constructivism avoids normative alienation by proposing to understand normativity as arising out of agency itself. It avoids social alienation by emphasizing the way that social relations are fundamental to agency. It avoids the emptiness problem faced by Kantian constructivism by scaling back its ambitions: we may not be able to derive a substantive moral order from the very idea of agency, but we can nevertheless understand the order that provides the normative background against which our agency is developed and exercised as itself a realization of mutual recognition in institutions, practices, norms, and so on.

This suggestion, however, raises the spectre of a new kind of alienation: that between an individual and her social world. Importantly, this kind of alienation is not a feature of the metaethical theory, but a troubling feature of life for which Hegelian constructivism fails to provide a way out. In other words, those who find it comfortable to identify with the contingent form of social life in which they find themselves embedded—which does not require that they endorse every norm, practice, or institution, only that they are at home in the framework they compose and optimistic that its limitations can be reformed through internal criticism—will not be alienated from it. But those who find the contingent order inhospitable, oppressive, or otherwise impossible to identify with will not be comfortable with the apparent conservatism of Hegelian constructivism.

This final form of alienation, what might be called *real alienation*—alienation from the realization of mutually recognitive agency that grounds one's own—is a real problem for us,

not a philosophical problem to theorize away. Moral philosophy struggles with the perennial temptation of a transcendent, ahistorical standpoint from which to criticize the status quo wholesale. The need to make sense of radical, and not merely internal, criticism is real, but the promise of a socially and historically unencumbered critical standpoint is illusory. It's a strength of the Hegelian constructivist account that it makes this temptation recognizable as a problem. The solution, I fear, is not more metaethics.

5.0   Thin as a Needle, Quick as a Flash

5.1   Introduction

The Sovereignty of Good—especially the first essay, "The Idea of Perfection"—is often associated with a critique of a certain picture of agency and its proper place in ethical thought. There is implicit in this critique, however, an alternative, much richer one. Though her immediate target is the Kantianism of Stuart Hampshire, which no longer enjoys the centrality in moral philosophy that it did in Murdoch's time, a broadly Murdochian conception of agency provides just as compelling an alternative to both the Humean and Kantian varieties predominant in metaethics today.

The worry is that Humean and Kantian conceptions of agency are too thin, too isolated from everything else that makes for a life. Where the Humean emphasizes the complex personal context from which choice and action emerge, she conceives the agent as for the most part passive in creating it; and where the Kantian emphasizes the activity of the agent she focuses on the moment of choice at the expense of the development, maintenance, and improvement of the background ethical awareness through which the world is disclosed to the agent as normatively saturated.[1] Neither approach accounts for the ongoing process of setting up the choice space, which by Murdoch's lights is to miss out on what is most important. Making sense of what we have reason to choose is not a task that can be understood independently of making sense of how we arrive at the normative circumstances

---

[1]These characterizations of 'the Humean' and 'the Kantian' are simplified for the sake of drawing the contrast, and I don't intend them as portrayals of any particular philosopher. I will discuss below the extent to which they do and do not capture important features of leading Humeans and Kantians, especially Sharon Street and Christine Korsgaard, and what resources Street and Korsgaard have available to respond to the Murdochian critique. See especially n.11. To the extent that the toy Kantian and Humean views I invoke are simplified to the point of caricature I hope they are nevertheless valuable as devices for drawing out important details of the view I attribute to Murdoch.

Murdoch herself has in mind Hare and Sartre in particular. The contours of these debates have of course shifted, but they have remarkably much in common. The lineage from Hare to today's Humean non-cognitivists (and their similarities and differences from cognitivist Humeans; constructivist Humeans like Street or reductive realist Humeans like Mark Schroeder (2007)) are well-known, and Murdoch's concerns about Sartre's grounding of ethics in rationally unconstrained acts of choice appear in criticisms of Korsgaard, sometimes with Sartre invoked for comparison. See e.g. (Chang 2009).

that structure our practical possibilities to begin with. Murdoch underlines that agency is not a power that is dormant in us most of our lives, springing into act at moments of choice—it is the active aspect of our constant, inner, reflective engagement with the world. What emerges from her critique is a more sophisticated picture of agency, the ongoing, active work of which turns out to be implicated in ethical vision after all.

Another under-appreciated feature of *Sovereignty* is that Murdoch's account of moral progress involves an implicit account of alienation as the inadequacy of one's practical standpoint to ethically relate oneself to concrete others. [2] Murdoch is often read as a humanist for whom moral progress begins with simple egoism and ends with loving care directed at the individual other, understood as a kind of immediate, ahistorical respect for human dignity. There are elements of that sort of view in the text but there is also the material for a story about how to get people in view in their concrete, particular reality, in social and historical context, and in relation to oneself.[3] Thereby we can understand moral progress less in terms

---

[2]Neither this feature of Murdoch's view nor the account of agency play a role in (Setiya 2013), Hopwood (2018; 2017), or most of the essays in (Broackes 2012); though Carla Bagnoli's entry (Bagnoli 2012) does discuss the positive role of agency in Murdoch's work, the features that she emphasizes are not those discussed here. Though most of the essays in the Broackes volume do not deal with the aspects of Murdochian agency that I am concerned with, Bridget Clarke's contribution (Clarke 2012) deals with the relationship between moral and political knowledge in a way that overlaps with the discussion in the second half of the paper, and Lawrence Blum's (Blum 2012) touches on some of the same ideas. I discuss both below—see especially n.16.

[3]The centrality, for Murdoch, of recognizing concrete particularity is discussed in illuminating detail by Lawrence Blum (1986), and some of his examples in (Blum 1991) help to draw out the importance of social context. The case of Theresa and Julio is exemplary in this respect:

> Theresa is the administrator of a department. One of her subordinates, Julio, has been stricken with a debilitating condition in his leg causing him frequent pain. He approaches Theresa to help work out a plan by which the company and in particular she and his division can accommodate his disability. Theresa is unable to appreciate Julio's disability and the impact it is having on his work. While in principle Theresa accepts the company's legal obligation requirement to accommodate to Julio's disability, in fact she continually offers Julio less than he needs and is entitled to.
> More generally, Theresa makes Julio feel uncomfortable in approaching her and gives him the impression that she thinks he is perhaps too self-pitying and should just "pull himself together." It is not that Theresa fails entirely to see Julio as "disabled" and as "in pain," but she does fail fully to grasp what this means for him and fails fully to take in or acknowledge that pain. The level of his pain and its impact on his mental state is insufficiently salient for Theresa. (704–5)

While Blum does not explicitly address the role of ableism in preventing Theresa from seeing Julio clearly, he discusses the case with great sensitivity toward precisely this aspect.

of achieving clarity with respect to timeless truths about the human qua human, and instead as a process of getting clear about the social and moral world, often in ways that can be unflattering.

In §5.2 I will try to draw Murdoch's positive conception of agency out of her critique of the then-predominant alternative, and in §5.3 I will try to fill in this picture by seeing how it works in her conception of moral progress. My aims here are not purely exegetical. I want to provide a reading of the text that I find compelling and in doing so use Murdoch to pose a challenge to the orthodox presentation of logical space. I am convinced that by reflecting on and somewhat updating Murdoch's intervention in her contemporary debates we can see new and promising possibilities.

So far from being intended as a contribution to Murdoch scholarship, some of what I say here will at least apparently conflict with what Murdoch says in other work (see e.g. n.24). I want to emphasize that the intended audience for this paper is a philosopher working in ethics or metaethics who has had some contact with Murdoch's work and not found much in it to engage with. When I refer to aspects of Murdoch's work as "under-appreciated", or interpretations as "misleading" I mean only to suggest that someone encountering her work may reasonably come up with misleading interpretations and fail to appreciate what I think is important, and not that Murdoch scholars have systematically misread the text (I do not think that they have, though as noted above they have tended not to emphasize the same themes that I emphasize here—see n.2). What follows is a reading of *Sovereignty* in particular: one that highlights ideas that I think are worth exploring for moral philosophers and philosophers of agency.

5.2   Agency

5.2.1   Murdoch's Critique

That Murdoch offers a way to think about agency and its role in the life of a person may seem a puzzling suggestion given that she emphasizes her criticism of an agency-centric

moral philosophy, promising to focus on moral contemplation and vision instead. On a first glance Murdoch appears to want to make agency all but unimportant, taking a backseat to a contemplative picture of morality on which the primary concern is vision.

Nevertheless, her target is not agency in general, nor the idea that agency is important to ethics, but an especially narrow and simplistic conception of agency, promoted to the place of exclusive fundamentality in ethical thought. Murdoch's real target is the image of an isolated will, made vivid in her example of a certain common but mistaken conception of morality as a 'visit to a shop':

> I enter the shop in a condition of totally responsible freedom, I objectively estimate the features of the goods, and I choose. The greater my objectivity and discrimination the larger the number of products from which I can select.[4] (8)

The kind of mistake she wants to diagnose involves looking only at the 'needle-thin' moment of choice—understood in terms of the public action that issues from it—as the source of value in the world. According to the simplistic picture agency is identified as the capacity for choice, an isolable faculty or power residing in a human creature, a "burrowing pinpoint of consciousness, inside, or beside, a lump of being" (47). Nothing else about this creature is of any ethical significance beyond the will it supports.

Because the moment of choice itself is understood as the locus of all value, the capacity must act against a background of value-neutral options: "If the will is to be totally free, the world it moves in must be devoid of normative characteristics, so that morality can reside entirely in the pointer of pure choice" (40). It is not just the rest of the human creature that is relevant only insofar as it bears the will but the whole world of mere things that is relevant only because the will operates upon it.

Finally, agency so conceived is exercised only in the isolated moment of decision, and otherwise inactive: "the machinery is relentless, but until the moment of choice the agent is outside the machinery" (15). The only thing of moral significance is the "point of action," (15) at which "the agent, thin as a needle, appears in the quick flash of the choosing will" (52). What happens in between moments of choice is of no more ethical significance than

---

[4]Her target here is clearly not just a character in philosophical work on agency, but encompasses as well the image of the rational agent associated with a popular understanding of economics. Citations without references are to (Murdoch 1971).

the body as the will's vessel or the world upon which choice is exercised.

5.2.2 Inner Struggle

Murdoch rejects a picture of agency as a capacity that is dormant until the moment of choice, is significant only insofar as it issues in reaching out to pick something, and acts against a background of mere things that are simply present to us. Her more sophisticated picture is one in which having options before one is already an achievement of agency, and the moment of reaching out to pick something is almost an epiphenomenon to the whole affair of bringing the value-laden options to awareness—by then the choice has for the most part already been made.

If the metaphor of morality as a visit to a shop illustrates what is wrong in a certain way of thinking about agency and its role in ethical life, Murdoch's famous example of M and D captures what has gone missing. The broad strokes of the example are that M is a mother whose son has married D, and while M initially forms harsh judgments of D, over time she comes to see D in a more favorable light, as a result of an inner effort. Certain details of the case are important (sometimes in ways that I think have been under-appreciated) and I will return to them in due course. But the most obvious point the example makes is that morality is not just concerned with what is public—public reasons, public actions, public concepts—but in addition and in some ways more importantly what is inner: inner acts, attention, moral knowledge, and the ongoing processes obscured by a mistaken emphasis on the moment of choice and action.

The inner acts Murdoch ascribes to M include reflection, introspection, observation, consideration, and an ongoing process of inner struggle. They flow from M's conviction that D is worthy of such reconsideration. M is characterized as "well-intentioned" and "capable of self-criticism", and the process begins when she says to herself, "let me look again" (17). Thus they cannot be understood as the mere operation of the 'machinery' of M's passive psyche—Murdoch notes that it needn't turn out this way, that M could instead "[settle] down with a hardened sense of grievance" (17). That M commits herself to the inner struggle is what makes these inner *acts* and not simply the alterations that her outlook undergoes.

A kind of inner activity upon which Murdoch places special emphasis is attention, a concept she borrows from Simone Weil: the "just and loving gaze directed upon an individual reality" (33). Murdoch is not entirely clear about this but she appears to understand attention not as a specific type of inner act but instead as a sort of functional kind: our reflections, observations, introspections, and other inner acts—including inner acts of omission that result from ceasing to think in certain ways—pattern together as forms of attention when they are directed with loving care upon a particular person.

Not all inner activity is attention in this sense. We can introspect (perhaps only in bad faith) in ways oriented toward rationalizing our own behavior; reflect on things we despise, reveling in our disgust rather than challenging it; or carefully observe our enemies while we plot revenge. Just as M could have remained un-reflective about D out of a lack of self-criticism, she could just as well have seethed out of spite or jealousy. This as well would be inner activity, and presumably of (negative) moral significance for Murdoch. But it would not be attention.[5]

What individuates inner activity as attention is its object (an "individual reality") and its loving character. It aims at a certain kind of morally crucial recognition, at love understood as "knowledge of the individual," (27) which is the "central concept of morality... thought of in light of the command 'be ye therefore perfect'" (29).

That attention aims at *knowledge* introduces a further purpose of the example, and the other of the two likely most remarked-upon: that inner moral activity results in an epistemic improvement, a state of clear vision. Moral knowledge is of the real as it is, not the idle

---

[5]There are also forms of inner activity that appear to meet the criteria for attention, but which are ethically or politically problematic. The phenomenon Kate Manne (2017*a*) has dubbed 'himpathy' ("the excessive or inappropriate sympathy extended to a male agent or wrongdoer over his female victim"), for example, involves directing excessive attention to the wrong individual reality. Perhaps we should say that attention is not always a good thing, or perhaps we should say himpathy is not attention after all: that the social conditions that make someone an inappropriate object of sympathy for the same reason make a loving gaze directed upon him *un*just; a lot depends on how much work justice does in the definition of attention. (Compare Miranda Fricker's (2007) concept of a credibility surplus; see especially (Medina 2011).) To consider an object of attention in abstraction from his social and political context would be to center love at the expense of justice.

An ethics that takes as its subject-matter de-contextualized individuals, relying on a standard set by their humanity as such, for example, would be humanistic in the way that I worry Murdoch might be. But if we take seriously the idea that justice is a constitutive feature of attention, then taking into account the social and political context in which an agent's relationship with another is situated will be required to establish what clear vision amounts to, and Murdoch's view certainly provides the resources to make good on this approach. I will return to this theme in the conclusion.

background of free choice, but normatively saturated. Moral concepts, Murdoch explains, don't move about in a non-normative world but "set up... a different world" (27). In this context the term 'reality' itself "appears as a normative word" (36).[6]

Important in this process is the role of moral concepts. For Murdoch they are learned from involvement in public language but refined and developed internally over a lifetime. We derive them initially from our surroundings, but "take them away into [our] privacy" (25). Refinement of our moral concepts is something like the *mechanism* by which attention reforms our moral vision. M comes to have a new understanding of D in which concepts like vulgarity, indignity, and juvenility have been replaced by the newly-refined concepts of refreshing simplicity, spontaneity, and youthfulness. Presumably M would earlier have been incapable of seeing D through these positively-valenced concepts, and being able to do so is the achievement of inner struggle. Though doing so amounts to seeing D clearly, moral vision is always mediated by concepts.[7]

5.2.3 Struggle as Agency

Taken altogether this story of inner activity, moral vision, and an always already normative reality might appear to be an alternative to an ethics of agency. But what is crucial here is that it is a story about ethical *activity*. Indeed, attention, for Murdoch, is "the characteristic and proper mark of the *active moral agent*" (33, emphasis added). Hers is not a two-stage account of what is morally relevant in the life of a person—first inner contemplation, and then only after that the act of choice—but a picture of a continuous, active process: "what we really are seems much more like an obscure system of energy out of which choices and visible acts of will emerge at intervals in ways which are often unclear and often dependent on the conditions of the system in between the moments of choice" (53). She occasionally refers to this as a picture of human agency but more often as a picture of human freedom, which comes to the same thing.

---

[6]Cora Diamond (1988, 266ff) discusses the converse phenomenon, viz. the *in*ability to occupy a 'life-with-a-concept' once the background conditions of the concept's full applicability have been lost.

[7]From a certain perspective this kind of concept-dependence, and in particular dependence on concepts that one struggles to grasp, appears to trade off against any kind of realism. This is a much bigger issue than I have the space to address here, indeed it is one way of seeing the central challenge taken up by McDowell (1994; 1979), whose approach is broadly Murdochian in spirit.

If inner struggle serves, for Murdoch, as a conception of agency rather than an alternative to it, what sort of conception of agency is it? In some ways it is more an expansion of the concept than a rival: rather than a capacity that is mostly dormant, active only in the moment of choice, agency is an ongoing process from which choices emerge, when they do, as a kind of externalization or 'outward movement' of the inner process (36). Rather than acting against a background of inert options (products as Givens, present to us whenever we glance at the shop shelf) with choice as the source of value, agency aims at achieving clear perception of an already normative reality: "freedom is not the sudden jumping of the isolated will in and out of an impersonal logical complex, it is a function of the progressive attempt to see a particular object clearly" (23).[8] And rather than living within the creature— itself a mere 'lump of being'—the activity of agency is integrated in the appetitive, emotional, and locomotive aspects of our embodiment. One of its most important effects is cultivation of proper desire, for we are "unified being[s] who see, and who desire in accordance with what [we] see" (39).

Compared to a standard conception of agency Murdoch shifts the emphasis away from the moment of choice and toward the ongoing process of generating and maintaining a context within which choice takes place. As she notes, very little of what we do involves stopping to consider the moral pros and cons of the available options, the reasons for and against, and making a choice that expresses our freedom. Most of what we do, even when we do it consciously, involves asking and straightaway answering, "shall I go? Oh yes, I promised to" or simply paying the check at a restaurant when we see it set before us (35). Given that we arrive at those rare moments of choice after a lifetime-so-far of achieving a moral perspective on a normatively structured world, "at crucial moments of choice most of the business of choosing is already over (36). Still, these outward actions are no less the expression of agency: they are expressions of agency precisely because it is our agency that is at work all along.

---

[8]Cf. 36: "I can only choose within the world I can *see*, in the moral sense of 'see' which implies that clear vision is the result of the moral imagination and moral effort."

Maintaining that it can take effort to put oneself in a position to see value, without treating value as any less real, is a central theme of both Callard (2018) and Ebels-Duggan (2017). The idea of clear perception here also works something like 'practical understanding' in (Schafer 2017).

5.2.4 Rival Conceptions of Agency

Murdoch's picture of agency emerges from her critique of Hampshire's, and half a century later his is all but forgotten. Today's Kantians have largely abandoned the behaviorism that, according to Murdoch at least, informed mid-20th-century moral theory, as well as the flat-footed Wittgensteinianism that appears to begin and end with a behavioristic reading of the 'private language argument'. [9] Nevertheless a broadly Murdochian conception of agency provides just as compelling an alternative to both the Humean and Kantian varieties predominant in metaethics today.

Humeans tend not to emphasize agency as such, instead focusing on the subjective context in which choices are made. Given an agents' values, preferences, attitudes, desires, or ends, choosing is a straightforward affair of maximizing, applying a bit of instrumental reason perhaps, but not contributing much itself to the normative circumstances. What the Humean gets right, for Murdoch, is the way in which arriving at a choice with a rich, complicated, contingent, and personal point of view leaves little room for the capacity for choice itself—the rational will—to generate or anoint value.

Kantians, on the other hand, emphasize the active role that agency plays in determining choice and action, bearing ultimate responsibility for value in virtue of having the capacity to reflect on possible actions. Kantian agency takes in a world that is practically significant only insofar as it causes us to form desires and inclinations, and normative only insofar as our ability to step back from desires and inclinations and choose whether to endorse them is what makes them reasons. While it is not ultimately up to each of us what reasons we have, even the categorical reasons that we necessarily share are explained by our agency itself, and not by anything about the context in which our agency is exercised.[10]

Neither of these pictures is complete by itself, by Murdoch's lights: choice does not take place in a vacuum, generating its own normativity against a normatively inert backdrop, nor does the normative context in which we act simply appear to us as given. The particular

---

[9]Murdoch's own picture is, I think, Wittgensteinian in its own way, and readers of the *Philosophical Investigations* are no longer as likely to find in it a flat out rejection of the idea that concepts could be 'taken into privacy' and developed in the way Murdoch proposes.

[10]Diamond (2010) makes a similar observation about the relationship between Murdoch's anti-Kantianism and Korsgaard's particular brand of Kantianism.

form of awareness we bring to bear on the world does most of the work most of the time in determining what we do, but to cast us as passive in its creation is to miss the crucial role that agency plays.[11]

For Murdoch agency is implicated in how we form a moral understanding, in how we refine our moral concepts, and in how we cultivate desires. The act of choice is the 'outward movement' of this inner process but it is only agency becoming explicit and concrete, and not the principal work of agency itself. The decision has most of the time already been made by the formation of an awareness, in which agency is always already involved. Agency does not move the person to act in a mechanistic world of mere objects but mediates our involvement with a normatively saturated world of persons and things of value.[12]

5.2.5 Practical Standpoints

At this point I want to depart from the text and make what I hope is a friendly suggestion for how to think about this alternative picture of agency: it is one on which the characteristic

---

[11]Both Street and Korsgaard themselves—as well as other Humeans and Kantians like Annette Baier and Barbara Herman, who do not thematize agency in the same way but offer richer visions of moral psychology—have more sophisticated pictures than this discussion suggests. Korsgaard at least may be working with a conception of agency that goes some way toward the one I'm proposing, even if the 'received view' of Korsgaard (see e.g. Bratman 1998, Gibbard 1999, Ridge 2005) is considerably more simplistic.

Whatever resemblance they bear to particular thinkers, the Humean and Kantian views I address here are, I hope, recognizable in the collective imagination of metaethics. Beyond those articulating what I refer to above as the 'received view' of Korsgaard (which does, I think, map quite closely onto the toy Kantianism I contrast with Murdochian agency in the body text), the ideas of agency as discrete choice and as instrumental rationality with respect to fixed aims and desires exert considerable force on metaethical debates. My argumentative strategy here somewhat mirrors Murdoch's own, in that she juxtaposes her view against an especially flat-footed kind of Kantianism for the sake of drawing a more vivid contrast. Readers of Murdoch have been quick to note that there is more compatibility between her and a more sophisticated Kantian on the role of love and attention in ethics (see e.g. Bagnoli 2003 and Merritt 2017). Insofar as the Murdochian critique of toy-Kantian agency is effective, it sets a standard that any more sophisticated Humeans and Kantians will have to meet.

See also (Regan 2002) for a similar criticism of the arbitrariness of the Kantian will, e.g. "what do this 'reflecting' and 'choosing' amount to in a world where there are no standards of value outside ourselves?" (280).

[12]Insofar as there are any accounts available of how agency is involved in self-constitution (e.g. Korsgaard (2009; 1996) and Chang (2009; 2013)) they rely on too thin a picture of what agency *does*, and the role of agency not only in making a choice but in constructing a context for choice and action. For Korsgaard, for example, "your identity is in a quite literal way *constituted* by your choices and actions" (Korsgaard 2009, 19), and reflection aims only at arriving at a description under which you value yourself (Korsgaard 1996, 100–1). One can see in Chang's attempt to look beyond choice alone for a source of normativity a kind of tacit admission that setting up the choice matters—that's what 'given reasons' do, until they run out—but treating them as given misses all of the action.

exercise of agency is not choice but the development, construction, and transformation of the *practical standpoint* from which choice emerges.

The practical standpoint is typically understood as that occupied by an agent, a valuer, a creature that acts for reasons as such.[13] An individual's distinctive practical standpoint, then, is a set of evaluative attitudes and judgments, a practical identity through which possibilities are made reason-giving. Here again we can see Murdoch not as presenting a rival conception of a practical standpoint so much as an expansion: what this generic characterization leaves out is the way that the development and revision of a practical standpoint is integrated into in an agent's life.

Further, a practical standpoint is not *merely* a set of attitudes or judgments—it constitutes a distinctive outlook on the world, a way in which the world itself is present to one. This is just to say both that there is some degree of unity or holism to the content of a practical standpoint, and that while it is personal it is not therefore subjective in the strong sense that would exclude there being something outside of ourselves with which our evaluative attitudes and judgments put us into contact. Our practical standpoints relate us to the world, and to the others with whom we share it.

Finally, there is a structuring element played by the distinctive moral concepts that we rely on as resources in determining our attitudes and judgments. Different sets of moral concepts provide for different ways of disclosing the world.

So, a Murdochian conception of a practical standpoint is as a form of awareness of the world—a kind of self-world relationship—articulated through concepts derived from public language but refined in subtle ways for each agent, through which a reality outside of ourselves can be seen as normatively saturated, and from which choices and actions emerge as expressions or 'outward movements'. And practical standpoints, in turn, are constructed, maintained, and transformed through an ongoing process of inner activity: the constant work of agency. This picture incorporates the features of the Humean and Kantian pictures but provides for a more complete account of how a practical standpoint—including the work that goes into forming and reforming it, and the choices and actions that it produces—can

---

[13]For general discussions of the idea of a practical standpoint along these lines, see Street (2008; 2010), Enoch (2011*a*), Schafer (2015*a*).

be the locus of ethical self-understanding and evaluation.

This picture additionally helps to make good on another central feature of the role Murdoch ascribes to inner reflection in the life of a moral agent: she warns against a conception of morality limited to discrete concerns like keeping promises and paying debts, and understands it instead as a constant pursuit of self-improvement, which suffuses every part of life. With this conception of agency in mind we can see that lifelong ethical task as one of tending to and improving our practical standpoints.[14]

Even though her main concern is to dislodge a misleading picture of agency we can find in Murdoch an alternative one, as the ongoing process of refining and transforming one's practical standpoint. It is, in my view, an attractive picture, one that can make sense of the seamlessness with which our habitual and skillful actions emerge from our simply being in a world of significance, and of agency as a feature of human life rather than an isolable activity: the embodiment of our capacity to determine ourselves.

This picture stands on its own as an attractive alternative to those that dominate moral philosophy. It is, however, a picture that I am extracting from a particular way of thinking about moral progress, with a particular conception of its aim and what stands in its way. Given that so far I've only offered a somewhat schematic characterization of the idea of a practical standpoint, to get a bit clearer it will help to work through Murdoch's own conception of moral self-development, to get a sense of the sort of practical standpoint that she takes us to begin from and the one that underwrites an ethical ideal.

5.3   Moral Progress

In this section I want to apply Murdoch's conception of agency to a picture of moral progress. There is, as it turns out, more structure to her account than just the idea of a practical standpoint, to be filled in with different commitments, judgments, or sets of values—some good, some better, perhaps one ideal. For Murdoch moral progress involves

---

[14]Contrast again Korsgaard: "there is work and effort—a kind of struggle—involved in moral life... the ongoing struggle for integrity..." (Korsgaard 2009, 7), which is to say the struggle to achieve coherence among one's commitments.

moving from an alienated standpoint to a self-consciously social one, where alienation and self-conscious sociality are not standpoints exhaustively characterized by their contents: they each represent formally different ways of relating to the world outside of oneself. Hers is a picture of moral self-improvement on which the movement between standpoints is not just a matter of trading in one for another, structurally identical but substantively different, but a movement that proceeds by improving the "quality of our relations with the world" (95).

For Murdoch, moral progress is a matter of transcending egoism in order to achieve clear vision of a normatively-saturated reality. My stalking horse in this section will be an overly simplistic interpretation of this picture; while it is a misreading of Murdoch, I think it would be a natural way of developing the same general idea of moral progress using the resources provided by the sort of Kantian and Humean approaches to agency against which I've tried to position hers.

The simplistic idea is this: that the egoism with which moral progress begins is the natural state of a creature whose power of reason is instrumental and whose purposes are narrowly self-interested, that what makes the egoistic practical standpoint ignoble is its lack of complementary altruistic purposes, and that moral progress aims to generate them by carefully attending to the reality of other persons, because to appreciate the humanity in another brings about a psychological change toward greater altruism.

Unlike the image of morality as a visit to a shop, which Murdoch herself displays as a contrast for the M and D example, she does not explicitly consider this picture of moral progress. The essays on moral progress, unlike "The Idea of Perfection", are less focused on rebutting a dominant approach than directly reflecting on themes she finds compelling. Still, it is helpful to have in view for the same contrastive purpose, as it differs from the one I want to attribute to Murdoch along each axis that interests me. For Murdoch egoistic alienation is not a matter of natural self-concern, nor is the ethical standpoint achievable independently of reflecting on social context and history.

5.3.1 Alienation

Murdoch characterizes the enemy of moral progress as the 'fat relentless ego', and elsewhere as 'fantasy' and 'self-focus' (51, 57). It may be tempting to find here an implicit theory of human nature as essentially selfish and in need of being overcome, a version of the familiar Hobbesian state-of-nature-dweller, *homo economicus*, the Calliclean in the breast of us all.

One reason to hope, however, that this isn't what Murdoch is up to is that she cautions against precisely this kind of move in Hampshire, whom she accuses of "imposing upon us a particular value judgment in the guise of a theory of human nature" (2). And indeed we find in her description of M not an obsession with maximizing self-interest but an intersecting set of moralized judgments, and in particular class bias. D initially appears to M as "unpolished and lacking in dignity and refinement... silly and vulgar" (16–7); M doesn't like how D dresses, or her accent—she thinks her son married beneath him; M's polite outward behavior is ascribed to correctness; and as she becomes reflective about her motives she discovers snobbery and narrow-mindedness.

True, she is also jealous and protective of her son. But the complex of attitudes and judgments with which she begins reflect, for one thing, a concern that could only be self-interested in a broad sense of 'self-interest,' given that D is no threat to her pursuit of material gain.[15] If M is driven by self-interest at all it is by the drive for honor, status, or esteem, and not even her own but her son's. And it is not at all clear that she is even driven by this socially-articulated form of self-interest: Murdoch nowhere describes her as concerned with the effects D will have on her or her son's life, nor properly concerned with particular consequences at all. Rather, M's harsh appraisal of D follows directly from the standpoint of a certain social class, defined by custom and manner. M is nothing like the image of a Calliclean egoist—her moral failings are the result of inhabiting a historically conditioned class position. Her moral failings have a political character.[16]

---

[15] All the more so given that D is, by hypothesis, out of the picture once M's reflections begin.

[16] Lawrence Blum (2012) identifies the political reading of the M and D case but argues that Murdoch did not intend it, and that this reflects a more general shortcoming of her work, viz. her focus on personal fantasy as a source of distorted vision at the expense of investigating the social and political sources of the same. He laments that "Murdoch's moral philosophy shows very little appreciation of the social and culture forms of the distorting images that block an appreciation of the humanity and the individual reality

If M is exemplary of the kind of self-focused, self-aggrandizing, clouded vision that Murdoch identifies as the enemy of moral progress it is not because she exhibits any kind of supposedly natural selfishness, but because she cannot see D through her class-based social conditioning. She suffers not from innate drive to out-compete her conspecifics but from a socialized drive to evaluate others on the basis of class-marked behaviors and manners. The point generalizes: the ethical standpoint from which moral progress departs is not a natural one but a social one. We are all products of time and place and circumstance, we encounter the world from a particular social position, from a particular form of *subjectivity*.[17]

---

of other human beings" (317). Blum may very well be right about Murdoch herself, but the reading of *The Sovereignty of Good* that I'm offering here is one that centers precisely the social and political dimension of moral perception that Blum takes Murdoch not to appreciate. As I noted from the outset my concern is to draw a compelling picture of moral agency and moral progress out of the text, and not to contribute to the project of historical scholarship that concerns itself primarily with what Murdoch actually meant to claim.

Another reader of Murdoch who has called attention to the political aspect of the M and D case is Bridget Clarke. She does so in the course of developing, on Murdoch's behalf, a response to the concern that placing virtuous habits at the center of moral theory risks making a sufficiently critical perspective on existing social practices inaccessible. She argues that part of moral perception, for Murdoch, is seeing how others are different from oneself, often in ways conditioned by social practices:

> For an agent to attend to another person in the Murdochian sense, the agent must grasp both the relevant similarities and the relevant differences which obtain between himself and the individual to which he attends. In these terms, to attend is to walk the fine line between overestimating and underestimating the continuities between oneself and others... The appreciations of similarity crystallize in the recognition that he is as real as oneself. The appreciations of difference crystallize in the recognition that he is separate from oneself. And this, I want to suggest, is another way to understand what it means to perceive another justly and lovingly, in all his particularity and complexity, i.e. in the light of the Good as Murdoch conceives it. (Clarke 2012, 238)

She argues that "true (Murdochian) perception of individuals involves understanding the past and present position of those individuals within the larger social structure," (251) which is only possible for agents whose practical standpoints are informed by and situated within communal, political practices of critique and consciousness-raising.

While I have some minor disagreements with Clarke I think (and hope) that the view I defend here is complementary to, rather than in conflict with, hers. In particular I have some reservations with the importance Clarke places on seeing similarity, and as I argue just below what we see when we see clearly is more than a morally significant subset of the properties others have, but Clarke's suggestion that the individual moral standpoint is conditioned and enabled by social-political resources is congenial to the picture of the relationship between morality and politics that I sketch in the conclusion.

[17] Judith Butler (2015) offers a similar discussion of the ways that agency is conditioned by subjectivity, noting that "I am already affected before I can say 'I' and that I have to be affected to say 'I' at all" (2). There are further resonances between my reading of Murdoch and Butler's positive vision, including with the way that the context of action is itself of ethical significance: "the ethical does not primarily describe conduct or disposition, but characterizes a way of understanding the relational framework within which sense, action, and speech become possible" (12).

The problem with a practical standpoint marked by selfishness, then, is not that it is wanting in complementary altruistic impulses but that it obscures "a reality separate from ourselves" (46). Her concern is not with the traditional difference between egoistic and altrustic motive but with the ways that self-focus can trap us behind a cloud of fantasy, the "tissue of self-aggrandizing and consoling wishes and dreams which prevents one from seeing what is there outside one" (57), which in turn bears the traces of social and political context.[18]

Moral improvement—the achievement of moral knowledge—issues from escaping fantasy and self-focus, from "detachment" or "suppression of the self," and from "clearing our minds of selfish care" (64, 82[19]). Thus, moral improvement is a process of getting us outside of ourselves and into genuine contact with others. If this is what an improved practical standpoint consists in, and agency is the ongoing process of improving one's practical standpoint, we can see its activity not just in the production of discrete actions but in the ongoing transcendence of mediation by the self and its interests.

Generalizing from the example of M alienation from others is not, for Murdoch, a matter of unquestioned, natural egoism, but of *unreflectively* occupying a socially and historically conditioned practical standpoint from which others are ethically obscured.[20] Perhaps, for

---

[18]This is one of the places where Murdoch most clearly diverges from Korsgaard, for whom moral failing is explained by making oneself into the wrong kind of agent by endorsing the wrong kinds of practical principles—those of self-love, say—rather than having a mistaken understanding of one's world. For the Kantian, defect is a matter of acting on the wrong motives and on the wrong maxims, rather than seeing through a distorted conceptual framework. (See e.g. (Korsgaard 2009, 162ff).)

[19]There are interesting echoes here of Schopenhauer (1958)—who argues that the ethical task is to transcend the illusory *principium individuationis* and recognize our unity with all beings—and of course with the Buddhist and Vedic philosophical traditions he draws from. See for example (253): "[the will] sees through the form of the phenomenon, the *principum individuationis*; the egoism resting on this expires with it." Indeed Schopenhauer also employs, sometimes on behalf of classical Indian philosophy, the metaphor of clouded vision (e.g. 352). On the relation between Murdoch and Schopenhauer, see (Crisp 2011, 283–4)

[20]This is a different sense of alienation than the one diagnosed as by Bernard Williams (1973) as a defect in moral theory, interpreted in terms of a kind of internal, psychic disunity (compare Stocker (1976) on 'moral schizophrenia'), and the historical sense discussed by, for example, Schacht (1970) and Jaeggi (2014), drawing on Rousseau, Hegel, Marx, etc., as a defect in social formations, interpreted in more metaphysical terms, involving estrangement from one's *essence.*

Elsewhere Murdoch (1958) addresses the latter under its Marxian guise, arguing that mid 20th-century socialism lacks a compelling vision of the nature and costs of being alienated from one's labor power, which a renewed commitment to theorizing in terms of the development and refinement of concepts can help to address. As far as I can tell, Murdoch's discussion of alienation there, as a defect in economic relations, is only very indirectly connected to the kind of alienation I am concerned with here.

example, we are alienated by pride rather than narrow, material self-interest, and thus unable to see ourselves as complicit in social practices with consequences we refuse to acknowledge. Coming to see others clearly, then, might involve overcoming prideful self-consciousness, coming to see oneself in social relation to others in ways that are shameful (as a snob, say, for M, or as the beneficiary of oppressive power relations). It might also or instead involve transitioning from a standpoint characterized by the false appearance of having no particular standpoint (Reason as the universal white male, etc.) to a standpoint in which one self-consciously relates oneself to others through politics and power (sometimes imaginatively casting oneself as a villain—the apologist for or beneficiary of injustice, the colonizer, the one on the wrong side of history).

5.3.2   Self-Conscious Sociality

It is easy to see how social contingency conditions an alienated standpoint—we are all socially and historically positioned in different ways, ways that come along with distinct ways of having one's vision clouded—but it is less clear how social contingency could condition moral progress, the overcoming of alienation. Plausibly, each alienated standpoint is alienated in its own way but clear vision is always the same.

The first thing to note, however, is that there is more to achieving a standpoint of clear vision of others than simply having another person in view. To occupy an ethical standpoint involves seeing not just the other but oneself in relation to the other. This kind of recognition is not just a matter of knowing that another is there or what she is like, but of knowing how one stands in relation to her—of knowing her in relation to oneself and oneself in relation to her. Ethical knowledge is not the mere discovery of an attribute something has (personhood, say) but a self-conscious placing of oneself in a world of persons.

If this is right it suggests that we can have a kind of ersatz knowledge of others even through a cloud of fantasy and self-focus—nothing is stopping M from *believing* that D is, for example, a person. In the sense in which the concept of personhood is public it paradigmatically applies to creatures like D, and if M didn't take D to be in the concept's extension we might doubt whether she had even minimal competence with it. But as M reworks her

moral concepts she arrives at a kind of recognition of D cum ethical self-knowledge. Or rather, to recognize D clearly and to understand herself in relation to D are and the same state of ethical vision.[21]

This is the first bit of structure I want to add to the idea of moral progress as overcoming alienation: it is a movement from a kind of alienated knowledge of others that amounts to merely endorsing a proposition that predicates something of another to a kind of self-conscious knowledge of others that places oneself with respect to them. This is what achieving clear vision amounts to, and through the process we come into contact with a normatively-saturated reality outside of ourselves.

5.3.3 Ideology

There are a number of ways a standpoint can be alienated but one common factor in obscuring our ethical vision of others is ideology: roughly, an implicit system of values that disguise relations of power.

Exactly how to think about overcoming alienation, then, is a delicate matter. The most straightforward story to tell is one on which, in virtue of occupying a particular social position and therefore starting with a socially-conditioned form of subjectivity, moral progress amounts to transforming our practical standpoints from those infused with ideological biases to standpoints free of any. At first we see others only through an ideological haze of stereotypes and assumptions, unable to appreciate their individuality. Moral progress is achieved by carefully reflecting on the ideology we find in ourselves in order to rid ourselves of it, to see others clearly.

This story is familiar from a genre of *Ideologiekritik* that has recently been embraced by some political philosophers in the so-called analytic tradition,[22] and whatever there is to be said for it I am reluctant either to endorse it or to ascribe it to Murdoch. For one

---

[21]The status of personhood is of course not what is at issue in the M and D case; the unsuitable concepts through which M initially sees D pick out specific, judgment-laced ethical features, and the concepts through which M comes to see D clearly equally so. The point I'm making here is broader: failures of recognition even in the most general sense are not failures to have predicative beliefs with ethical content, but failures of self-conscious relatedness *through* ethical concepts. Given that recognition of personhood is the kind of thing that is often thought to do real moral work the example helps to show the broader application of my point.

[22]Tommie Shelby (2003, 174), for example:

86

thing, at least on a particularly naive version, when combined with Murdoch's account of the *mechanism* of moral progress—inner activity—we are left with the unsatisfying suggestion that the solution to political problems is individual moral improvement. If we just reflect thoroughly enough we can rid ourselves of ideological illusion, revealing the world as it really is, free of ideology.

There are a number of reasons to push against this picture. The idea of individual, moral solutions to political problems involves a distastefully naive kind of humanism—ideology becomes all too easy to unlearn, and the self-other relations that come to self-consciousness have no content of their own. (Empirical work on cognitive bias suggests that introspection really does not get us very far.) And for my own part I'll just confess that I'm much more attracted to an Althusserian approach to Ideologiekritik, captured by the slogan "ideology has no outside," which strikes me as appropriately modest and realistic.[23]

My own commitments aside, however, there is some reason to doubt that Murdoch would accept the equation of clarity with escaping ideology. For one thing she opens "On 'God' and 'Good'" with the plea for "a moral philosophy which can speak significantly of Freud and Marx, and out of which aesthetic and political views can be generated" (45). She explores aesthetics throughout *Sovereignty*, but nowhere explicitly returns to politics.[24] One hopes,

---

Ideologies perform their social operations by way of illusion and misrepresentation. What this means practically is that were the cognitive failings of an ideology to become widely recognized and acknowledged, the relations of domination and exploitation that it serves to reinforce would, other things being equal, subsequently become less stable and perhaps even amenable to reform.

See also (Shelby 2002). Jason Stanley (2015) defends a similar view, on which ideology 'occludes' widespread injustice and 'misleads us about the structure of reality' (207). Sally Haslanger (2012)—likely the other most prominent analytic theorist of ideology—has a more complex conception of ideology, but insofar as she uses 'ideology' as a pejorative term her view appears to be one on which the ideal would be to escape it.

Of course neither Shelby nor Stanley think overcoming ideology is easy to achieve through internal reflection alone; indeed these 'restricted' conceptions of ideology are subtle and sophisticated in their own way, and my criticism in the main text is not directed at them. Only combined with the Murdochian conception of moral progress discussed here would they yield such implausible results. I do, nevertheless, think that the Althusserian alternative discussed below is more promising. See (Swanson 2017) for a criticism of Stanley along the same lines.

[23]See (Althusser 1971), and compare (Kukla 2018*b*).

[24]Nowhere, that is, within the three essays that constitute *Sovereignty*; Murdoch does discuss politics and its relationship to ethics elsewhere, especially in *Metaphysics as a Guide to Morals*. Indeed much of what I say in this section and in what follows at least superficially conflicts with Murdoch's attitude toward politics in other work. For example, she notes that "we do not 'live' the world of politics in the way we 'live' our private lives" (Murdoch 1993, 744), which suggests that she sees the interpersonal dimension of morality and

however, that she wants moral theory to generate political views with more sophistication than the proviso *achieve enlightenment by escaping ideology through personal reflection*. A virtue of the reading I'm sketching here (or at any rate, a feature that I see as a virtue) is that it makes room for a more reciprocal relationship between moral and political thought than the more common, foundationalist picture of politics as applied ethics.

5.3.4   The Theory-Ladenness of Clear Vision

Finally, this picture leaves out the role Murdoch ascribes with great emphasis to the development and refinement of moral concepts. Appreciating the role of moral concepts may be key both to declining to attribute to Murdoch the picture of moral progress as tracing an arc from selfishness to a-historical altruism, and to exploring an alternative. Recall that the process of transforming our practical standpoint consists *inter alia* of taking once-public concepts and working them into a more nuanced apparatus for disclosing the world to us. For Murdoch the inner effort of agency has the ultimate aim of making itself obsolete: we work to get our concepts in order so that they become invisible in disclosing the world to us.[25]

The task for agency, then, is to overcome the distinction between the inner and the real. The real is saturated by normativity but it takes a constant inner effort to see it through the

---

involvement in social movements, institutions, and other forms of engagement with the process of reforming the social world as playing different roles in the lives of individual agents. My aim here is to propose and apply a reading of Murdochian agency as elaborated in *Sovereignty*, and a detailed discussion of Murdoch's stance toward politics more generally would be beyond the scope of this paper. See (Clarke 2012; 2006) for discussion of the political dimension of individual morality in Murdoch's work more broadly.

[25]A perennial temptation among philosophers urges that we can clearly distinguish the descriptive content of perception, which is theory-independent, from the evaluative content, which supervenes on the descriptive (or the natural, the material, or the non-evaluative). The kind of theory-ladenness I attribute to Murdoch here, however, is well-captured in Nicholas Sturgeon's arguments that even perception of empirical phenomena is theory-laden, and in precisely the same way as that in which evaluative concepts saturate our perceptual relationship with the world (Sturgeon 1988). Indeed his discussion of the case of Mary and Jane (Sturgeon 1986, 74) echoes Murdoch's description of the moment at which M decides to 'look again' and thereby (eventually) comes to see D differently. See also (Sturgeon 2009) for related doubts about supervenience.

Murdoch is more explicit about endorsing something like this in (1956), where she claims that different concepts may "[decide] the relevance of the facts and may, indeed, render them observable," that "if the concept is withdrawn we are not left with the same situation or the same facts," and that "if moral concepts are regarded as deep moral configurations of the world, rather than as lines drawn around separable factual areas, then there will be no facts 'behind them'" (54–5).

haze of self-interest and fantasy. But the aim of that inner struggle, of the agential ordering of attention and development of adequate moral concepts, has the aim of making itself invisible. If correct perception occasions right action then to achieve this kind of perceptual contact with a normatively saturated world of others, invisibly mediated by the outcome of a constant active effort, is to make moral action natural, to make one a part of the world—a different world, 'set up' by moral concepts, as we saw above—in which one is embedded and with which one interacts.

Nevertheless, achieving a moral standpoint amounts to developing adequate concepts. It is the achievement of a process of conceptual development. Thus, even as our moral concepts fall away at the point of clear vision there is still a sense in which they are operative, in which clear vision is a matter of conceptually-mediated awareness of others that makes our social relations into a part of how we see ourselves. This looks like a case of 'theory-laden' awareness, where clarity and conceptual mediation are not exclusive but complementary. The moral standpoint involves a better theory, one that that gets us outside of ourselves and into reflective contact with a world, but no less conceptually articulated. This kind of theory-ladenness is a close cousin of ideology, and if this is right moral progress propels us toward a standpoint that is no less political than the one with which we began.

To overcome alienation, then, is often to divest oneself of socially-based moral illusion— the illusion that one deserves what one has, say, or that more moral progress has taken place than really has. (As I write this, recent events revealed, to some of those whose vision had previously been clouded by ideology, the experience of being a person of color in interacting with police, or the amount of explicit racism and white supremacism that still exists. Importantly, on the Murdochian picture this does not happen automatically, and some people will remain alienated from this bit of social reality. Being presented with evidence is never enough without reflection.) The kind of clear vision that one achieves is not just of others qua human but as others situated with respect to oneself in historically-contingent ways. We see others precisely in our shared social contexts rather than escaping those contexts to a realm of pure dignity, respect, or abstract rightfulness.

5.4   Conclusion

I've proposed a Murdochian conception of agency as the capacity to occupy and transform a practical standpoint, and a schematic Murdochian account of moral progress in terms of overcoming an alienated standpoint to achieve a standpoint of clear vision. I've further suggested that alienated standpoints obscure our vision because of the particular way in which they are ideological, and that seeing oneself in relation to others is still theory-laden insofar as it relies on having developed adequate moral concepts. Thus the distinction between cloudy and clear vision—between fantasy and reality—is not a matter of whether or not one's relationship with the world is mediated by a socially-contingent conceptual framework, but the extent to which that framework makes our social relations visible to us and allows us to become reflectively self-aware in relation to others.

The result of these kinds of transformations involves coming to understand oneself in relation to others not in ahistorical, decontextualized, purely human ways, but precisely in ways that are socially and politically realistic, that bring the social, historical, and political circumstances in which these relations are embedded into explicit self-awareness: what is often called *consciousness raising*. From a standpoint of ethical clarity these relations are visible and normatively-valenced—seeing things as they are, socially and politically, is not to see them 'neutrally', morally speaking, but to see them as calling for response.

This picture of moral knowledge as socially-situated self-knowledge raises a handful of closely connected questions, questions about the *content* of the moral standpoint—that is, about those concepts that are most central to the kind of conceptually-mediated clarity of vision at which moral progress aims, about the relationship between the internal struggle to reach moral clarity and political struggle to improve the social relations made self-conscious in it, and about just how much moral progress is possible in the span of one lifetime. I don't have the space here to offer anything like a complete accounting of these issues, either on Murdoch's behalf or my own. Still, I would like to conclude with some reflections on the direction in which such an accounting would likely go.

A fairly conventional moral ideal would be to transcend circumstance and social position, struggling toward an a-historical, morally absolute standpoint: a moral view from nowhere.

As we have seen, however, for Murdoch making social and political relations explicit in our understanding of ourselves among others leaves us still trapped within them, morally speaking. Through inner activity alone the best we can hope to achieve is a realistic view from where we are, with all of its contingency.[26] Working to change the context that constitutes this standpoint is a matter of politics, and not of moral progress alone. The world must be set aright before any of us can truly achieve clear moral vision.

The specter of a certain kind of pessimism looms: the idea that for some of us, born in the wrong time or place to have ethical life available, there might just be no point in moral struggle—or, with a 'pessimism of the intellect, optimism of the will' twist, that without hope of ever truly achieving a moral standpoint one must still commit to the task.[27] More hopefully, however, this picture may recommend appreciating the limit of moral self-improvement alone, without the proper accompaniment of political struggle. The idea would be that the inner struggle to achieve morally clear vision and the outer struggle to achieve a shared, ethical form of life, are complements. Moral progress for the individual and political progress advance hand in hand.

Insofar as inner struggle aims, among other things, to put one in a position to contribute to political progress, it might involve cultivating solidarity, becoming an ally or a comrade, learning to recognize hidden forms of injustice that are made explicit and given names as required by circumstance, being liberated from a false sense of obligation, discovering that one is entitled to what has historically been sanctioned. "Morally correct action," as Lukács (1919, 9) puts it, "is related fundamentally to the correct perception of the given historico-philosophical situation," and often aims to overthrow it.

For a more concrete example, consider the concept of womanhood, which many radical feminists thought depended on a class system that had to be overturned, while simultaneously being tactically necessary for feminist organizing (e.g. Wittig 1982). More generally this can help to explain why concepts corresponding to socially constructed and historically marginalized groups can be crucial for political struggle, even if what they denote is an

---

[26]On the idea that we can allow some contingency into our evaluative standpoints without losing our grip on an independent reality, see McDowell (1983*a*).

[27]See (Gramsci 1929), Letter from Prison (19 December 1929). He uses the phrase in the context of political struggle, not a moral one, but as I note just below I think the two combine nicely.

essentially unjust social formation. The often-used metaphor of 'colorblind racism' is nicely illustrative: even while the struggle for racial justice plausibly aims to create a society in which indifference to race is an aspect of public life, one's ability to see others in relation to a shared social context requires race-concepts, without which one might be trapped in a fantasy of justice that has not yet arrived. (That temporary allegiances structured by political concepts are crucial for political struggle is most explicitly thematized by Haraway (1984), from whom I take important inspiration.)

All of this will be responsive to the actual social circumstances, and making the right concepts available for inner struggle will often be itself a political achievement. [28] Moral progress is achieved through inner struggle but may nonetheless require social, environmental, and institutional scaffolding. And likewise political progress may be impossible for those trapped behind the veil of fantasy, who watch the dramas of their own lives play out with themselves at the center, unwilling to accept a supporting role in a political narrative—or who see themselves through the eyes of their social 'superiors' as mere instruments without moral or political agency of their own.

A lot remains to be done to make this picture complete, let alone plausible, but here are the seeds of what I think is a promising way to think about agency, moral progress, and its relationship with politics. The account of agency as the capacity to cultivate and transform one's practical standpoint, and of moral progress as the movement from an alienated standpoint to a socially self-conscious one through which we can clearly see others in relation to ourselves, are Murdoch's, and in speculating about the relationship between moral struggle and political struggle the basic resources I've used are hers. Thus, I hope to have made good on my claim that looking again to Murdoch as a neglected figure can help us to see some things that we may have been missing.

---

[28]One fault line I have so far largely avoided, though I have hinted at it, is that between those who take the concept of the human in particular to do serious moral work, and those, like me, who think it is almost always either too broad or too narrow. Diamond (1991) offers a defense of a kind of moral humanism in a broadly Murdochian spirit.

Bibliography

Althusser, Louis (1971), "Ideology and Ideological State Apparatuses." In *Lenin and Philosophy and Other Essays*, New York, NY: Monthly Review Press.

Anscombe, G. E. M. (1978), "Rules, Rights, and Promises." *Midwest Studies in Philosophy*, vol. 3, no. 1:pp. 318–323.

Bagnoli, Carla (2003), "Respect and Loving Attention." *Canadian Journal of Philosophy*, vol. 33, no. 4:pp. 483–516.

— (2012), "The Exploration of Moral Life." In (Broackes 2012).

— (2016), "Kantian Constructivism and the Moral Problem." *Philosophia*, vol. 44, no. 4:pp. 1229–1246.

Baiasu, Sorin (2016), "Constitutivism and Transcendental Practical Philosophy: How to Pull the Rabbit Out of the Hat." *Philosophia*, vol. 44, no. 4:pp. 1185–1208.

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.

Blum, Lawrence (1986), "Iris Murdoch and the Domain of the Moral." *Philosophical Studies*, vol. 50, no. 3:pp. 343–367.

— (1991), "Moral Perception and Particularity." *Ethics*, vol. 101, no. 4:pp. 701–725.

— (2012), "Visual Metaphors in Murdoch's Moral Philosophy." In (Broackes 2012).

Boyd, Richard (1988), "How to Be a Moral Realist." In *Essays on Moral Realism*, Geoffrey Sayre-McCord, ed., pp. 181–228, Cornell University Press.

Boyle, Matt (2012), "Essentially Rational Animals." In *Rethinking Epistemology: Volume II*, Günter Abel & James Conant, eds., de Gruyter.

Brandom, Robert B. (1994), *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

— (2007), "The Structure of Desire and Recognition: Self-Consciousness and Self-Constitution." *Philosophy and Social Criticism*, vol. 33, no. 1:pp. 127–150.

— (2019), *A Spirit of Trust*. Harvard University Press.

Bratman, Michael (1998), "Review of Korsgaard's *The Sources of Normativity.*" *Philosophy and Phenomenological Research*, vol. 58, no. 3:pp. 699–709.

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.

Broackes, Justin (2012), *Iris Murdoch, Philosopher.* Oxford University Press UK.

Buber, Martin (1923), *Ich und Du.* Reprinted in 2004 as *I and Thou*, trans. Ronald Gregor Smith; A&C Black.

Burge, Tyler (2010), *Origins of Objectivity.* Oxford University Press.

Butler, Judith (2015), *Senses of the Subject.* Fordham University Press.

Callard, Agnes (2018), *Aspiration: The Agency of Becoming.* Oxford University Press.

Chang, Ruth (2009), "Voluntarist Reasons and the Sources of Normativity." In *Reasons for Action*, David Sobel & Steven Wall, eds., pp. 243–71, Cambridge University Press.

— (2013), "Grounding Practical Normativity: Going Hybrid." *Philosophical Studies*, vol. 164, no. 1:pp. 163–187.

Clarke, Bridget (2006), "Imagination and Politics in Iris Murdoch's Moral Philosophy." *Philosophical Papers*, vol. 35, no. 3:pp. 387–411.

— (2012), "Iris Murdoch and the Prospects for Critical Moral Perception." In (Broackes 2012).

Clarke, James Alexander (2009), "Fichte and Hegel on Recognition." *British Journal for the History of Philosophy*, vol. 17, no. 2:pp. 365–385.

Crisp, Roger (2011), "Iris Murdoch on Nobility and Moral Value." In (Broackes 2012).

Darwall, Stephen (1977), "Two Kinds of Respect." *Ethics*, vol. 88, no. 1:pp. 36–49.

— (2006), *The Second-Person Standpoint: Morality, Respect, and Accountability.* Harvard University Press.

— (2012), "Bipolar Obligation." *Oxford Studies in Metaethics vol. VII*, pp. 333–.

Darwall, Stephen, Gibbard, Allan, & Railton, Peter (1992), "Toward Fin de Siècle Ethics: Some Trends." *Philosophical Review*, vol. 101, no. 1:pp. 115–189.

Dasgupta, Shamik (2017), "Normative Non-Naturalism and the Problem of Authority." *Proceedings of the Aristotelian Society*, vol. 117, no. 3:pp. 297–319.

Dennett, Daniel C. (1987), *The Intentional Stance.* MIT Press.

Diamond, Cora (1988), "Losing Your Concepts." *Ethics*, vol. 98, no. 2:pp. 255–277.

— (1991), "The Importance of Being Human." *Royal Institute of Philosophy Supplement*, vol. 29:pp. 35–62.

— (2010), "Murdoch the Explorer." *Philosophical Topics*, vol. 38, no. 1:pp. 51–8.

Ebels-Duggan, Kyla (2017), "Beyond Words: Inarticulable Reasons and Reasonable Commitments." *Philosophy and Phenomenological Research.*

Enoch, David (2006), "Agency, Shmagency: Why Normativity Won't Come From What is Constitutive of Action." *Philosophical Review*, vol. 115, no. 2:pp. 169–198.

— (2011*a*), "Can There Be a Global, Interesting, Coherent Constructivism About Practical Reason?" *Philosophical Explorations*, vol. 12, no. 3:pp. 319–339.

— (2011*b*), *Taking Morality Seriously: A Defense of Robust Realism.* Oxford University Press.

Ferrero, Luca (2009), "Constitutivism and the Inescapability of Agency." *Oxford Studies in Metaethics vol. IV*, pp. 303–333.

Fichte, Johann Gottlieb (1797), *Grundlagen des Naturrechts nach Prinzipien der Wissenschaftslehre.* Reprinted as *Foundations of Natural Right* in 2000, M. Baur, trans., Cambridge 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.

Franks, Paul (1996), "The Discovery of the Other: Cavell, Fichte, and Skepticism." *Common Knowledge*, vol. 5:pp. 72–105.

Fricker, Miranda (2007), *Epistemic Injustice: Power and the Ethics of Knowing.* Oxford University Press.

Gendler, Tamar Szabó (2008), "Alief and Belief." *Journal of Philosophy*, vol. 105, no. 10:pp. 634–663.

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.

Golub, Camil (2017), "Expressivism and Realist Explanations." *Philosophical Studies*, vol. 174, no. 6:pp. 1385–1409.

Gramsci, Antonio (1929), *Letters from Prison.*

Haase, Matthias (2014), "For Oneself and Toward Another: The Puzzle About Recognition." *Philosophical Topics*, vol. 42, no. 1:pp. 113–152.

Haddock, Adrian & MacPherson, Fiona, eds. (2008), *Disjunctivism: Perception, Action, Knowledge* . Oxford: Oxford University Press.

Haraway, Donna (1984), *A Cyborg Manifesto.*

Harman, Gilbert (1977), *The Nature of Morality: An Introduction to Ethics.* Oxford University Press.

Haslanger, Sally (2012), *Resisting Reality: Social Construction and Social Critique.* 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.

Honneth, Axel (1996), *The Struggle for Recognition: The Moral Grammar of Social Conflicts.* MIT Press.

Hopwood, Mark (2017), "Murdoch, Moral Concepts, and the Universalizability of Moral Reasons." *Philosophical Papers*, vol. 46, no. 2:pp. 245–271.

— (2018), "The Extremely Difficult Realization That Something Other Than Oneself Is Real: Iris Murdoch on Love and Moral Agency." *European Journal of Philosophy*, vol. 26, no. 1:pp. 477–501.

Hussain, Nadeem & Shah, Nishi (2006), "Misunderstanding Metaethics: Korsgaard's Rejection of Realism." In *Oxford Studies in Metaethics vol. I*, Russ Shafer-Landau, ed., pp. 265–94, Clarendon Press.

— (2013), "Metaethics and Its Discontents: A Case Study of Korsgaard." In *Constructivism in Ethics*, Carla Bagnoli, ed., Cambridge University Press.

Jaeggi, Rahel (2014), *Alienation.* Columbia University Press.

Julius, A. J. (2016), "Mutual Recognition." *Jurisprudence*, vol. 7, no. 2:pp. 193–209. Based on the Jurisprudence Annual Lecture, 2016.

Kant, Immanuel (1785), *Grundlegung zur Metaphysik der Sitten.*

— (1788), *Kritik der praktischen Vernunft.*

Kolodny, Niko (2003), "Love as Valuing a Relationship." *Philosophical Review*, vol. 112, no. 2:pp. 135–189.

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.

— (2018), *Fellow Creatures: Our Obligations to the Other Animals.* 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/~korsgaar/Korsgaard.on.Parfit.pdf` on 4/29/2019.

Kremm, Douglas & Schafer, Karl (2017), "Metaethical Quietism." In *The Routledge Handbook of Metaethics*, Tristram McPherson & David Plunkett, eds., pp. 643–658, Routledge.

Kukla, Rebecca (2018*a*), "Embodied Stances: Realism Without Literalism." In *The Philosophy of Daniel Dennett*, Bryce Huebner, ed., Oxford University Press.

— (2018*b*), "Slurs, Interpellation, and Ideology." *The Southern Journal of Philosophy*, vol. 56:pp. 1–26.

Langton, Rae (1997), "Love and Solipsism." In *Love Analyzed*, Roger E. Lamb, ed., pp. 123–52, Westview Press.

Lederman, Harvey (2018*a*), "Common Knowledge." In *The Routledge Handbook of Collective Intentionality*, Marija Jankovic & Kirk Ludwig, eds., pp. 181–195.

— (2018*b*), "Two Paradoxes of Common Knowledge: Coordinated Attack and Electronic Mail." *Noûs*, pp. 921–945.

— (2018*c*), "Uncommon Knowledge." *Mind*, vol. 127, no. 508:pp. 1069–1105.

Lenman, James & Shemmer, Yonatan (2012), "Introduction." In *Constructivism in Practical Philosophy*, Jimmy Lenman & Yonatan Shemmer, eds., Oxford University Press.

Lewis, David (1979), "Scorekeeping in a Language Game." *Journal of Philosophical Logic*, vol. 8, no. 1:pp. 339–359.

Lukács, Georg (1919), "Tactics and Ethics." Reprinted by 2014, trans. Michael McColgan; Verso, Brooklyn: NY.

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.

— (2016), "Humanism." *Social Theory and Practice*, vol. 42, no. 2:pp. 389–415.

— (2017*a*), *Down Girl: The Logic of Misogyny*. Oxford University Press.

— (2017*b*), "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 (1979), "Virtue and Reason." *The Monist*, vol. 62, no. 3:pp. 331–350.

— (1983*a*), "Aesthetic Value, Objectivity, and the Fabric of the World." In *Pleasure, Preference, and Value*, Eva Schaper, ed. Reprinted in (McDowell 1998*b*).

— (1983*b*), "Criteria, Defeasibility, and Knowledge." In *Proceedings of the British Academy, Volume 68: 1982*, pp. 455–79, Oxford University Press.

— (1984), "De Re Senses." *Philosophical Quarterly*, vol. 34, no. 136:pp. 283–294.

— (1994), *Mind and World*. Cambridge, MA: Harvard University Press.

— (1998*a*), "Lecture II: The Logical Form of an Intuition." *Journal of Philosophy*, vol. 95, no. 9:pp. 451–470.

— (1998*b*), *Mind, Value, and Reality*. Harvard University Press.

McNulty, Jacob (2016), "Transcendental Philosophy and Intersubjectivity: Mutual Recognition as a Condition for the Possibility of Self-Consciousness in Sections 1–3 of Fichte's Foundations of Natural Right." *European Journal of Philosophy*, vol. 24, no. 4:pp. 788–810.

Medina, José (2011), "The Relevance of Credibility Excess in a Proportional View of Epistemic Injustice: Differential Epistemic Authority and the Social Imaginary." *Social Epistemology*, vol. 25, no. 1:pp. 15–35.

Merritt, Melissa McBay (2017), "Love, Respect, and Individuals: Murdoch as a Guide to Kantian Ethics." *European Journal of Philosophy*, vol. 25, no. 4:pp. 1844–1863.

Moore, G. E. (1903), *Principia Ethica*. Dover Publications.

Moran, Richard (2001), *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press.

— (2018), *The Exchange of Words: Speech, Testimony, and Intersubjectivity*. Oup Usa.

Murdoch, Iris (1956), "Symposium: Vision and Choice in Morality." *Aristotelian Society Supplementary Volume*, vol. 30, no. 1:pp. 32–58.

— (1958), "A House of Theory." Reprinted in (Murdoch 1998).

— (1971), *The Sovereignty of Good*. Routledge.

— (1993), *Metaphysics as a Guide to Morals*. London, England: Penguin Press.

— (1998), *Existentialists and Mystics: Writings on Philosophy and Literature*. Allen Lane/the Penguin Press.

Nagel, Thomas (1986), *The View from Nowhere*. Oxford University Press.

Neuhouser, Frederick (1986), "Deducing Desire and Recognition in the Phenomenology of Spirit." *Journal of the History of Philosophy*, vol. 24, no. 2:pp. 243–262.

Parfit, Derek (2011*a*), *On What Matters Volume I*. Oxford University Press.

— (2011*b*), *On What Matters Volume II*. Oxford University Press.

— (2016), *On What Matters: Volume Three*. Oxford University Press UK.

Perry, John (1979), "The Problem of the Essential Indexical." *Noûs*, vol. 13, no. December:pp. 3–21.

Pinkard, Terry P. (1994), *Hegel's Phenomenology: The Sociality of Reason*. Cambridge University Press.

Pippin, Robert B. (2008), *Hegel's Practical Philosophy: Rational Agency as Ethical Life*. Cambridge University Press.

Pritchard, Duncan (2012), *Epistemological Disjunctivism*. Oxford: 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 (1980), "Kantian Constructivism in Moral Theory." *Journal of Philosophy*, vol. 77, no. 9:pp. 515–572.

Regan, Donald H. (2002), "The Value of Rational Nature." *Ethics*, vol. 112, no. 2:pp. 267–291.

Ridge, Michael (2005), "Why Must We Treat Humanity with Respect? Evaluating the Regress Argument." *European Journal of Analytic Philosophy*, vol. 1, no. 1:pp. 57–73.

Rödl, Sebastian (2007), *Self-Consciousness*. Harvard University Press.

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 (2015*a*), "Realism and Constructivism in Kantian Metaethics : Realism and Constructivism in a Kantian Context." *Philosophy Compass*, vol. 10, no. 10:pp. 690–701.

— (2015*b*), "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.

— (2017), "Rationality as the Capacity for Understanding." *Noûs*.

Schmidt am Busch, Hans-Christoph & Zurn, Christopher F. (2009), *The Philosophy of Recognition: Historical and Contemporary Perspectives*. Lexington Books.

Schopenhauer, Arthur (1958), *The World as Will and Representation*. New York: Dover Publications.

Schroeder, Mark (2007), *Slaves to the Passions*. Oxford University Press.

Schwitzgebel, Eric (2001), "In-Between Believing." *Philosophical Quarterly*, vol. 51, no. 202:pp. 76–82.

— (2002), "A Phenomenal, Dispositional Account of Belief." *Noûs*, vol. 36, no. 2:pp. 249–275.

Sellars, Wilfrid (1968), *Science and Metaphysics: Variations on Kantian Themes*. New York: Humanities P.

Setiya, Kieran (2013), "Murdoch on the Sovereignty of Good." *Philosophers' Imprint*, vol. 13.

— (2014), "Love and the Value of a Life." *Philosophical Review*, vol. 123, no. 3:pp. 251–280.

— (MS), "Other People." Unpublished manuscrtipt retrieved from `http://www.ksetiya.net/uploads/2/4/5/2/24528408/other_people.pdf` on 6.6.2019.

Shafer-Landau, Russ (2003), *Moral Realism: A Defence.* Oxford University Press.

Shelby, Tommie (2002), "Is Racism in the "Heart"?" *Journal of Social Philosophy*, vol. 33, no. 3:pp. 411–420.

— (2003), "Ideology, Racism, and Critical Social Theory." *Philosophical Forum*, vol. 34, no. 2:pp. 153–188.

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.

Soteriou, Matthew (2000), "The Particularity of Visual Perception." *European Journal of Philosophy*, vol. 8, no. 2:pp. 173–189.

Stanley, Jason (2015), *How Propaganda Works.* Princeton University Press.

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.

— (2010), "What is Constructivism in Ethics and Metaethics?" *Philosophy Compass*, vol. 5, no. 5:pp. 363–384.

— (2012), "Coming to Terms with Contingency : Humean Constructivism About Practical Reason." In *Constructivism in Practical Philosophy*, Jimmy Lenman & Yonatan Shemmer, eds., Oxford University Press.

Sturgeon, Nicholas (1986), "Harman on Moral Explanations of Natural Facts." *Southern Journal of Philosophy*, vol. 24, no. S1:pp. 69–78.

— (1988), "Moral Explanation." In *Essays on Moral Realism*, Geoffrey Sayre-McCord, ed., Cornell University Press.

— (2009), "Doubts About the Supervenience Of The Ethical." In *Oxford Studies in Metaethics vol. IV*, Russ Shafer-Landau, ed., Oxford University Press.

Sturgeon, Scott (2008), "Disjunctivism about Visual Experience." In *Disjunctivism: Perception, Action, Knowledge* , Adrian Haddock & Fiona MacPherson, eds., Oxford University Press.

Swanson, Eric (2017), "Critical Notice of Jason Stanley *How Propaganda Works.*" *Mind*, vol. 126, no. 503:pp. 937–947.

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.

— (2012), "Why Be an Agent?" *Australasian Journal of Philosophy*, vol. 90, no. 2:pp. 223–233.

Travis, Charles (2004), "The Silence of the Senses." *Mind*, vol. 113, no. 449:pp. 57–94.

Velleman, J. David (2013), "Sociality and Solitude." *Philosophical Explorations*, vol. 16, no. 3:pp. 324–335.

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*a*), "Morality, Agency, and Other People." *Ergo*, vol. 5.

— (2018*b*), "Practical Reason Not as Such." *Journal of Ethics and Social Philosophy*, vol. 13, no. 2.

— (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.

Williams, Bernard (1973), "A Critique of Utilitarianism." In (Smart & Williams 1973).

— (1976*a*), "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.

— (1976*b*), "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.

Wittgenstein, Ludwig (1953), *Philosophical Investigations*. Wiley-Blackwell. Reprinted in 2009, G. E. M. Anscombe et al. trans.

— (1969), *On Certainty*. Basil Blackwell.

Wittig, Monique (1982), "The category of sex." *Feminist Issues*, vol. 2, no. 2:pp. 63–68.

Wood, Allen W. (1990), *Hegel's Ethical Thought*. Cambridge University Press.

Zylberman, Ariel (2014), "The Very Thought of You." *Philosophical Topics*, vol. 42, no. 1:pp. 153–175.