NORMATIVE PRIMITIVISM AND THE POSSIBILITY OF PRACTICAL THOUGHT

SAMUEL DAVID STEADMAN

A DISSERTATION SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PHILOSOPHY

YORK UNIVERSITY

TORONTO, ONTARIO

SEPTEMBER 2022

ABSTRACT

Reasons are essentially addressed to agents. Many contemporary efforts to illuminate this feature of reasons effectively reduce them to features of agents, e.g., to rationally-pruned desires, plans, or roles. Such reductive accounts neglect a second feature of reasons, namely, their capacity to transcend agential nature. They also neglect a feature of agents, namely, their orientation to normative entities as entities that transcend—and thus, that can guide and give shape to—agential nature. This dissertation offers a conception of the relation running from reasons to agents that captures both the transcendent character of reasons and the transcended character of agents.

I synthesize two strains of thought about reasons. The first captures their formal dependence on agency, which is manifested in each reason's being essentially a reason for some agent to do or think something. The second captures their substantive independence from agency, which is manifested in the fact that reasons needn't answer to what agents are like. These two strains of thought can be united in a single conception, but only if the elaboration of the formal features of reasons isn't taken to license the reduction of reasons to features of agents. In fact, unifying the two in a single conception requires that the relevant agential features be themselves depicted as formally dependent on features of reasons, so that the explanatory landscape for the philosophy of reasons and agents is properly represented in terms of the symmetric relations of a circle, rather than the asymmetric relations of reduction.

This refusal to reduce is best framed by primitivism about reasons, i.e., the view that characterizes the idea of *reason* as primitive. But such a primitivism must nevertheless supply the materials for an account of the practical thought by which agents can receive reasons as addressed to them. I seek to demonstrate how an idea can be primitive while at the same time supplying those materials, and thereby explaining the possibility of practical thought.

For my Aunt Joanne

TABLE OF CONTENTS

This dissertation applies a familiar philosophizing strategy: reveal something to be Janus-faced, and then undertake to reconcile its two faces. My Janus-faced something is the idea of *normative reason*, that is, the idea of *something that prescribes that some agent have a certain attitude or act a certain way*. Normative reasons are essentially addressed to rational animals *qua* agents, so that the idea of *normative reason* must turn a face to those animals, one that meets them where they are. But being essentially authoritative with respect to rational thought and conduct, normative reasons must command from a space that lies beyond the attitudes and actions of individual agents, and so the idea must have a second face, one that turns away from rational animals. This provides a basis for a threat to the intelligibility of the idea of *normative reason*—after all, nothing can turn to and away from the same thing, not at the same time. But my aim is only to threaten the idea so as to have the opportunity to rescue it. In composing this image of a vacillating idea, of a simultaneous approaching to and withdrawing from agency, I underscore the need for an account of the degree to which normativity in fact accommodates agency, and of the degree to which agency in fact accommodates normativity. And, in working through the exercise of reconciling the two faces, I develop resources for the illumination of both the normative domain and of our own nature as creatures capable of thinking and acting from aspects of that domain.

I characterize the two faces of the idea of *normative reason* in terms of the well-worn immanence/transcendence binary. Normative reasons bear on the thought and behaviour of thoroughly terrestrial creatures, like us, and in this way, they share in the immanence, the material worldliness, of those creatures. But no terrestrial thinking or doing can intervene in the authority with which reasons prescribe thought and behaviour. The notion of such an intervention is confused—if we could manipulate our reasons, that manipulation would itself fall within the jurisdiction of further reasons that direct us in making certain changes and in refraining from others.

1

We can endeavour to manipulate again at the second-order, but if we do, third-order reasons would arise to direct us there, and so it must continue *ad infinitum*. Reasons necessarily outrun our capacity to engage with them. Or to say the same thing, they transcend that capacity.

A fair bit of religious scholarship has been expended wrestling with the immanence/ transcendence binary, and the result has been, I think it's fair to say, vertiginous in its complexity. It's reasonable, then, to be a little squeamish about the prospect of tackling this same binary in the context of metanormative theory. But then, it's hard to see how it could be avoided. We might opt to exercise the pluralist's prerogative—another popular philosophizing strategy—and sever our Janus in two, so as to treat each face independently. But to do so would be to miss the extent to which the two faces are fastened together. Reasons are intelligible as prescriptive specifically because there are agents who are intelligible as potentially responding to prescription, and vice versa. More dramatically, reasons are reasons because they can be given flesh, so to speak, by agents, and agents are agents because they can fall short of, be humbled by, their reasons.

Thus, *reason* and *agent* are mutually dependent notions. The dependence relation running from *reason* to *agent* is embodied in the fact that reasons exert normative force specifically in relation to agents. An agentless world couldn't sustain any reasons; that our world comprises reasons is a consequence of its comprising agents. And because agents are corporeal, extended in time and space, made of mud and carbon, it follows that reasons, being tethered to agents, are tethered to these materials. We can see this reflected in the principle that there can be no *ought* without *can*. Though reasons transcend our actual thinking and doing, they are nevertheless bound by our possible thinking and doing. Perhaps normativity *simpliciter* isn't beholden to agency in this way—it may be that axiatic entities could inhabit an agentless world—but normative reasons certainly are.

Meanwhile, agents exercise their agency in a world that is, by their lights, incomplete. Normative reasons are the standards with reference to which this incompleteness comes into view;

this is what establishes the dependence relation running from *agent* to *reason*. If agents found the world exactly as it ought to be, they would have no reason to manifest their agency in changing it. And once denied a canvas upon which to let loose their creative power, their status as agents is withdrawn. So, the agent's world is necessarily a recalcitrant world, one that resists reason to some degree, not so much as to frustrate agency at every turn, but enough to provide the requisite friction, the kind that makes thinking and acting a project. And agents themselves occupy this same world, and so here too, they find deficiencies, the likes of which their reasons direct them to improve by turning their agency inward.

These considerations establish a kind of symmetry between the two ideas: normative reasons rely on agents specifically as immanent things, and agents rely on reasons specifically as transcendent things. This symmetry is inscribed into the idea of *normative reason*. It fastens the two faces of that idea tightly together, so tightly as to preclude pluralism. Our only avenue, then, is to unify the faces. We do this by manoeuvring our conception of *normative reason* so as to meet a pair of constraints, an immanence constraint and a transcendence constraint. The former constraint requires of metanormative theory that it establish the possibility of practical thought, and the latter constraint requires that metanormative theory establish the possibility of the normative assessment of the whole of our agential nature.

The immanence and transcendence constraints provide the contours for the sort of reconciliation that's wanted. We mustn't undertake to reconcile via assimilation, that is, by dragging the transcendent down to the immanent or dragging the immanent up to the transcendent. The symmetrical dependence relations between *reason* and *agent* require that each be preserved in its role. And anyway, if we were to make agency otherworldly, we'd make a mystery of the possibility of practical thought, so that we'd fail by the immanence constraint. And if, instead, we made normativity banal, we'd lose the perspective from which we can evaluate our nature, and fail by the

transcendence constraint. We walk something of a tightrope here; the unity of immanent and transcendent that we must find for the idea of *normative reason* is a delicate thing.

Regrettably, there's a disposition among metanormative theorists to leap to either extreme—to the too-immanent picture, which badly misunderstands its subject matter, or to the too-transcendent picture, which is, at best, bafflingly austere, and at worst, altogether occult. The remedy for the disposition to opt for the immanent extreme is non-reductionism, i.e., the repudiation of the view that our normative vocabulary can be translated into some non-normative vocabulary, like the vocabularies of decision theory, of reproductive advantage, of dialectical history, or whatever it may be. The remedy for the disposition to opt for the transcendent extreme is the recognition that metanormative theory can be constructive, i.e., it can advance substantive, positive theses about normative phenomena without betraying the *sui generis* character of that phenomena. It pays to be reminded that it's always available to us to be *both* constructive and non-reductionist. After all, explanatory circles needn't be vicious, so long as they're sufficiently wide.

The disposition to swing to either extreme is, at least in part, the result of a series of exaggerated dialectical responses. Metanormative theory began (or if not that, it came to be widely recognized as a *bona fide* philosophical subdiscipline) with G. E. Moore's opposition to the (inchoate) normative reductionisms of the nineteenth century. The trouble is that the language with which Moore states his objection oversteps the mark, producing a theory that's more obscure than it needs to be. Meanwhile, the constructive theorist, hoping to escape that obscurity, overcompensates in the other direction—she's too emphatic in stating the shortcomings of non-reductionist austerity, and thereby gives the impression of having embraced reductionism. This radicalizing dialectical exchange has endured for quite some time. It has on occasion prevented contemporary metanormative realists and constructivists from appreciating the extent to which the metanormative conversation has advanced in sophistication, and grown more intramural. It may be that they've

already struck upon the correct mixture of immanent and transcendent elements, but being disposed to hear echoes of Moore in contemporary realism, and echoes of Moore's naturalist interlocutors in constructivism, they're prevented from noticing that this is the case.

We can redraw the map of the contemporary metanormative scene—in particular, of the putative conflict between T. M. Scanlon's realism and Christine Korsgaard's constructivism—in view of the fact that metanormative theory can capture both the immanence and transcendence of normative reasons. I suspect that both constructivism and realism are only a few steps away from establishing a unity of their own, a kind of constructivist-realist hybrid, which marries the immanence of the former to the transcendence of the latter. Korsgaard's constitutivist insight opens a path for theory to follow from agency to normativity. This is a solution to my Janus-faced problem, but only if it's softened somewhat to allow for genuinely transcendent normative authority. A softened constructivism must surrender its claim that normativity is constructed, but this is not so drastic a move as it may seem. It relinquishes remarkably little from the Korsgaardian programme, especially given the work that there remains for Kantian self-legislation to do in a hybrid conception.

In any case, as a preliminary to explaining exactly *how* a metanormative theory might meet both the immanence and transcendence constraints, I have to show such a theory to be possible. From the start, pressure from the transcendence constraint drives metanormative theory to insist on the primitiveness of the idea of *normative reason*, and this primitiveness appears, on its surface, to drain that theory of all explanatory force. I will show that this is only an appearance; we needn't indulge the temptation to follow primitiveness to austerity. In just the same way that we have recourse to a constructive normative non-reductionism, we have recourse to a constructive normative primitivism. Note that the idea of *normative reason* is particularly amenable to non-reductive explanation. As we've seen, the ideas of *normative reason* and of *agency* are meshed

together—this is a relation that we can exploit for explanatory power, even given the irreducibility of both ideas.

In the first and second chapters, I introduce Scanlon's conception of *normative reason* as an exemplar of normative primitivism, and then articulate the immanence constraint in two formulations, first as demanding that we grant the possibility of practical thought and then as demanding an explanation of that possibility. The third chapter locates the immanence constraint among a set of traditions with which it overlaps, namely, metanormative internalism and constructivism, and moral contractualism. The fourth chapter returns to Scanlon, but this time in juxtaposition with Moore. Unlike Scanlon's primitivism, Moore's is austere, leaving it vulnerable to complaints from the immanence constraint, not least complaints that dwell on the special role reserved for rational intuitionism within the austere primitivist picture, the upshot of which is that practical thought is simply posited, and, as such, inexplicable. The comparison with Moore identifies the main features of the purported slide from primitiveness to austerity, and what we might do to avoid it.

While the joint satisfaction of the immanence and transcendence constraints is the ultimate aim, this dissertation concentrates on the immanence constraint alone. The first two chapters defend that constraint, and the last two offer the beginnings of a plan to meet it. All the while, I take the transcendence constraint for granted. My argument for the immanence constraint culminates in three criteria for success in metanormative theory: a theory of practical content, a first-order theory of practical reasons, and a theory of agency. I don't develop any of these theories here, though I do commit to certain starting-points.

My allegiances on the subject of practical content are to Donald Davidson, specifically to the conception, expounded by Robert Myers, that extends Davidson's triangulation argument to the normative domain. According to the triangulation argument, thought and talk originate in discursive

exchanges between animals as they together engage with an element of their shared world. A consequence of the extension of this argument to the normative domain is that normative entities must be construed as both natural (*contra* Scanlon) and real (*contra* Korsgaard). In the course of my argument for the immanence constraint, I represent adherence to Myers' elaboration of Davidson's theory as optional. This is a with a view to leaving open a space for a modified Scanlonian conception. While, formally speaking, it likely is optional, I can't see any alternative to it, not without repudiating the immanence constraint (and this, I suspect, can only be managed by embracing an all-anaesthetizing quietism).

The second criterion, namely, the completion of a first-order theory of practical reasons, is a regulative principle in the Kantian sense—the pursuit of a formulation of the unity of all practical reasons, and of the principles that categorize them and their relations, serves as a vector for the improvement of normative thought, though the ideal can never in practice be realized. What's needed, then, isn't an instantiation of the ideal, but a programme for its pursuit. Here I favour John Rawls' method of reflective equilibrium. Incidentally, though I touch on a number of apparently epistemological concerns throughout the dissertation, concerns raised by intuitionism and reflective equilibrium among other things, I prefer to frame these things as positions on the possibilities—or impossibilities, as the case may be—of first-order theory. Thus understood, Rawls charts an avenue for normative theory-construction, and Moore denies that any such theory can be devised.

A theory of agency ought to figure at the centre of an account of normative immanence. Unhappily, I don't tackle the subject directly. Still, in one way or another, it's present all throughout the dissertation. My argument for the immanence constraint takes its start from the idea of *necessitation*, that is, the idea of *the relation that channels prescriptive force from normative entity to agent*. Necessitation is connected to an essential feature of agency, its fallibility, which is indispensable to the resolution of the apparent tension between the two faces of the idea of *normative reason*. Being

essentially fallible, agents are confirmed by normative entities that extend ahead of them, and lead them along. Thus, it inheres in agency that it never arrives at its end, its immanence being essentially transcended. Necessitating entities, answering to this feature of agency, must issue their prescriptions from a space that agency can't reach.

The theory of practical content provides the outlines of the idea of *necessitation*. The task that falls to the theory of agency is to fill those outlines in. The theory I envision enlists normative reasons as the agent-facing entities of the normative domain, the ones that direct agents in the manipulation of their material world. The agent falls between these two, the normative and the material, and operates as the channel through which the first intervenes in the second. There's no absorbing the agent into either domain, since she's transcended by the one, and the instrument for the transcendence of the other. She must have a foot in both camps, as a kind of liminal thing, operating in a world she must regard as incomplete, and, as a (half-)denizen of that world, incomplete herself.

The theory of agency has immediate consequences for moral theory, and though, at this early stage, I can only gesture at the shape a theory of agency might take, this is enough to mark certain features of the nature of our reasons to respect ourselves and others *qua* agents. The facts of necessitation and agential fallibility generate reasons that are captured rather cleanly by Scanlon's contractualist conception, according to which moral wrongness answers to what each agent could reasonably reject. And in any case, necessitation presents a strong consideration in favour of an agency-mediated moral theory, one that denies that morality arises directly from our engagement with ground-level reasons, asserting instead that it arises from our self-consciousness as necessitated animals and from our recognition that others are likewise necessitated. One substantive result of such a view is that representations of reasons are entitled to moral regard even when those representations are likely mistaken. But I leave the elaboration of that theory for another time.

8

I gesture to the potential for expansion into moral theory in part to reinforce the case for constructive theory—among other things, it shows that the *reason-agent* circle can be widened still further to encompass at least some moral ideas—but I also mean to acknowledge certain earlier ambitions of mine, the distant ancestors to this dissertation, which deal in comparatively concrete features of our experience, but which I had to set aside in order to clear the stage of some of its more esoteric debris. Naturally, the practical philosopher's enterprise (or anyway, *this* practical philosopher's enterprise) has the exclamation 'moral equality!' at its centre. Metatheoretical considerations as to, say, the possibility of constructive theorizing, arise at its periphery. But it happens sometimes that, though the practical philosopher's journey begins at that centre, she is driven by discursive pressures to the periphery, so that she must undertake a long voyage home. Ideally, upon returning she finds herself, in virtue of her extensive travels, equipped with some new means of exclaiming 'moral equality!', one that better conveys its profundity and sublimity. This dissertation traces the first several steps of my own return.

1. THE IMMANENCE CONSTRAINT, FORMALLY CONSTRUED

1.1 THE RATIONAL ANIMAL'S COMPLAINT

Scanlon opens the first chapter of *What We Owe to Each Other* as follows:

> I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favor of it. "Counts in favor how?" one might ask. "By providing a reason for it" seems to be the only answer.[1]

Any proposal to assign a given thing to the category *primitive* warrants a cautious line of approach. But the subject of Scanlon's primitivism[2] is to be reckoned a special site for difficulty: classifying the idea of *normative reason* as primitive threatens to displace or disfigure—or anyway disappoint— another idea, one adjoining the idea of *normative reason*, namely, the idea of *rational animal.*

This danger isn't specific to Scanlon's conception; the complaint originating in the idea of *rational animal* is directed against reasons-primitivism as such. In effect, what the complaint demands is a foothold for the animal in the idea of *normative reason*. She must be provided with a knot on the surface of the idea, some marker with which to orient herself, something to give substance to the possibility of her exhibiting her rationality in action. To deny her that is to deny her the means to manifest her nature. And this, in turn, is to endanger the intelligibility of the idea of *normative reason* itself—normative reasons are essentially addressed to rational animals *qua* agents, so that, in the absence of the latter, there can't be any of the former.

This suggests a success condition for conceptions of *normative reason*: the excellent conception accommodates the rational animal. This is the immanence constraint. The task of meeting it applies pressure to the idea of *normative reason*, pressure in the direction of texture, of distinguishing marks. It's only by answering to this pressure that metanormative theory can preserve the idea of *normative*

---

[1] Page 17. The term 'reason' used here is intended in its "standard normative sense" (*ibid*, page 19).
[2] Scanlon prefers the name 'Reasons Fundamentalism' (see the first lecture of his *Being Realistic about Reasons*).

reason. But it's difficult to see, at least initially, how a reasons-primitivism could accomplish this, since it must begin from a primitive, which would seem to be (or at least its name suggests) something stark, even featureless.

The rational animal's complaint can be elucidated this way: if all that we can say of a given thing is that it's primitive, and little else besides—that is, if in our asking after that thing, our spade, as the slogan goes, is just about immediately turned—then there's no philosophy for that thing.[3] Its primitiveness puts it beyond thinking, and there's nothing quite so stultifying for the rational animal as that. The radically austere primitive offers nothing for her rationality to grip. There's no prospect of her reconciling her nature with it (at any rate, not insofar as she's rational). It might be said that she must adjust to the world as she finds it. But this isn't how she finds it. Normative reasons figure at the centres of our lives as thinkers and doers. They must be richly detailed and many-faceted things if they're to animate those lives in the ways that they in fact do. They can't be simple bedrock, blank and unspeaking, refusing all analysis. That just isn't how they arise for us.

The rational animal has a legitimate complaint. She's entitled to a conception of *normative reason* that's hospitable to her nature. And she's entitled, in her philosophical moments, to cast around until she finds one. Moreover, she's entitled to the efforts of metanormative theorists; they must endeavour to enable her to devise a hospitable conception of *normative reason*. But this is just what suggests the immanence constraint. It also suggests this facet of the immanence constraint: conceptions of *normative reason* must be non-austere. If nothing else, metanormative theory must disavow austerity, and take steps to avoid it.

---

[3] Korsgaard characterizes her 'normative question'—"[w]e are asking what *justifies* the claims that morality makes on us" (*The Sources of Normativity*, pages 9-10)—as "a call for philosophy, the examination of life" (*ibid.*, page 9). I mean to channel this same call in my articulation of the immanence constraint.

1.2 THE THIN FORMULATION OF THE IMMANENCE CONSTRAINT

I say all this in order to erect a hurdle for reasons-primitivism to clear. Certain reasons-primitivisms can clear it. Scanlon's is among them; given a sympathetic reading, his conception can be shown to be non-austere, and so, to this extent at least, benign with respect to the plight of the rational animal. All the same, we must work to find our way to this sympathetic reading—I trace a path to one in chapter 4. Other normative primitivisms aren't so capable. Moore's goodness-primitivism is the most consequential instance of metanormative austerity. He says "[i]f I am asked "What is good?" my answer is that good is good, and that is the end of the matter. Or if I am asked "how is good to be defined?" my answer is that it cannot be defined, and that is all I have to say about it."[4] Going forward, I'll use Moore as my representative for the kind of metanormative austerity that triggers the rational animal's complaint.

Reasons-primitivism isn't the enemy, not ultimately. In fact, reasons-primitivism is indispensable to the rational animal. This is because it contributes to meeting the sister to the immanence constraint—the transcendence constraint—which requires that it be possible to assess the whole of agential nature given normative entities that lie beyond that nature. Relatedly, the rational animal is essentially oriented to aspects of the normative domain about which she can go wrong. Primitivism contributes to both the preservation of this last clause, the possibility of her going wrong, and the satisfaction of the transcendence constraint by establishing that normativity is irreducible to what we in fact do. Still, we need some assurance that, in meeting the transcendence constraint, we don't estrange the rational animal. The immanence constraint filters out the normative primitivisms that open too wide a gap between reasons and animal.

At minimum, the constraint prohibits our developing an austere primitivism. This and the next chapter together establish a formulation of the immanence constraint that expresses this

---

[4] *Principia Ethica*, §6.

prohibition. But to begin, I distinguish between two formulations, one that's thin, and another that's thick; the first is defended in this chapter, and the second is defended in the next. The thin formulation inheres, I argue, in the idea of *normative reason* itself, and is therefore unassailable. The trouble is that it's too formal to challenge austerity directly. To accomplish that, it relies on its thicker instance. As I see it, the truth of the first formulation suggests the truth of the second. Or anyway, once we recognize the extent of our commitment to the thin formulation, the refusal to follow it to the thick formulation should strike us as desperate, a move only to be countenanced as a last resort. Though the argument for the thick formulation isn't indisputable, it suffices to establish the immanence constraint as something that exerts real—that is, substantive, concrete—force on the direction of normative theorizing.

The argument for the thin formulation proceeds in this way: I begin, in subsection 1.2.1, with a preliminary defence of the constraint as applied to conceptions of *normative reason*. In the two subsections following that, I settle a few issues pertaining to the scope of the constraint. Then, throughout section 1.3, I defend the main premise of the argument, namely, that the idea of *practical thought* is an essential constituent of the idea of *normative reason*. My reasoning takes the form of a *reductio*—conceptions that omit the idea of *practical thought* lack the resources to explain how practical reasons relate to agents, and so, can't provide a comprehensive account of the idea of *normative reason*. Among other things, normative theory must explain why deontic entities, like reasons, can't issue their prescriptions to arational entities; only agents can receive normative direction. It's the possibility of our engaging with some consideration in thought that makes us uniquely suitable for assessments of rationality and irrationality, and this is what establishes us as a locus, the only locus, for the application of deontic force. It's what supplies the contexts in which normative reasons arise. And if these contexts answer to our rational nature, normative reasons must do so as well.

13

Of course, given so formal a basis, the thin formulation can't give full expression to the rational animal's complaint. For that matter, it can't do much to capture the immanence that is transmitted to normative reasons across their essential relations to rational animals. This is what recommends the introduction of its thicker instance. It's this second formulation that's tasked with expressing the rational animal's demand for a foothold in the idea of *normative reason*, and, related to this, for a metanormative project that sets itself to, among other things, accommodating her nature. But the materials with which to vindicate the thick formulation don't come into view until we map certain of the essential connections between *reason* and *animal*, connections involving other ideas, like *practical thought*, and *rationality*. So, we make our start here, in the architecture of the idea of *normative reason*, where thought can be shown to perform a central role.

1.2.1 The immanence constraint and the idea of *normative reason*

I begin with a relatively narrow formulation of the immanence constraint: *conceptions of* normative reason *must, if they are to succeed in capturing the idea of* normative reason*, establish the possibility of practical thought about normative reasons.* This formulation (along with each of the other formulations I present throughout this dissertation) interprets the immanence constraint as a success condition—any successful conception of *normative reason* necessarily meets it. It doesn't by itself guarantee a successful conception, since there are other such conditions, like the transcendence constraint, and these must be jointly met in order to deliver success.[5] In any case, I call the immanence constraint a 'constraint' because of the limit it places on normative theorizing. It does this by calling attention to a feature of the normative domain, namely, its commitment to the possibility of practical thought. (I

---

[5] Each conception is a conception of something—a conception of some α, like *normative reason*, or *inference*—and each must meet adequacy conditions that jointly establish the intelligibility of its being about its something, its α, whatever it may be. As it happens, this first formulation of the immanence constraint amounts to an adequacy condition (trivially, all adequacy conditions are also success conditions). The thick formulation represents the immanence constraint as something more substantive, a success condition that isn't also an adequacy condition.

assume that all theorizing aspires to success, and that, in the case of normative theory, success is partly a matter of fidelity to the features of the normative domain.)

I'll have quite a lot to say about the concept of *practical thought* in what follows. For the moment, though, I'll just say that practical thought involves forming normative judgments and being motivated by those judgments in action, such that what binds judgment and action is a rational connection, rather than, say, a mechanical connection. I'll also say that there are (at least) two senses of 'rational' in play here, the first of which is classificatory, and contrasts with 'arational', and the second is evaluative, and contrasts with 'irrational'. When all goes right, practical thought issues in action that's 'rational' in the evaluative sense. But whether things go right or not, practical thought and the actions it determines are 'rational' in the classificatory sense. Classificatory rationality attaches to behaviour that's locatable within a discursive space that can sustain the first-person point of view (I recognize that this formulation is unclear, but it's the best I can manage at this stage). That space is the stage upon which evaluative rationality is attempted, and where both success and failure—that is, irrationality—can be registered. The sense of 'rational' in the term 'rational animal' is classificatory, though some degree of evaluative rationality is necessary if a creature is to be classified as rational, since no discursive space can be found for perfect irrationality. This last point is well expressed by Davidson: "[g]lobal confusion, like universal mistake, is unthinkable, not because imagination boggles, but because too much confusion leaves nothing to be confused about and massive error erodes the background of true belief against which alone failure can be construed."[6]

The chief virtue of this first formulation of the immanence constraint is that it can be straightforwardly justified from the fact that the ideas of *normative reason* and *rational animal* are

---

[6] "Mental Events", page 221. Also, "[i]rrationality is not mere lack of reason but a disease or perturbation of reason" ("Rational Animals", page 99).

intertwined. This is clear to see in Scanlon's schema for reason-relations, as presented in *Being Realistic about Reasons*. He says there that normative reasons, themselves only "ordinary facts", acquire their status as normative from their position in distinctively "normative relations".[7] The fundamental reason-relation, which obtains for each normative reason, can be schematized as "a four-place relation, R(p, x, c, a), holding between a fact p, an agent x, a set of conditions c, and an action or attitude a. This is the relation that holds just in case p is a reason for a person x in situation c to do or hold a."[8]

The details of his conception needn't detain us here. We need only register that the intelligibility of the relation is parasitic on the intelligibility of each of the ideas that is assigned a place, namely, the ideas of *fact*, of *circumstances*, of *action*, and of *agency*. Moreover, each idea must be intelligible in the performance of its role.[9] So, instances of the idea of *agency*—that is, agents—must be capable of receiving facts *qua* normative reasons as calling upon them to act in certain ways in certain circumstances, at least in the sense that it's rational for them to act from judgments as to what those reasons are. But this capacity is just the capacity for practical thought about reasons. Thus, the immanence constraint arises as a matter of course: the intelligibility of the central component of Scanlon's conception of *normative reason*, the reason-relation, derives, in part, from the possibility of practical thought.

---

[7] *Being Realistic about Reasons.*, page 30.

[8] *Ibid.*, page 31. Scanlon's schema encompasses both reasons for action and reasons for attitudes, some of which are cognitive, like beliefs. In this sense, it straddles the divide between practical and speculative reason. My focus is practical reason, so I'll omit the provision for reasons for attitudes from here on in (to simplify things, I assume that all reasons for conative attitudes can be read off of corresponding reasons for actions).

[9] As far as the ideas of *action* and *agency* is concerned, this sentence is a pleonasm, since there's no distance between the criterion that these ideas be shown to be intelligible *simpliciter* and the criterion that they be shown to be intelligible in the performance of their roles in the relation. This is because it belongs to the essence of rational agency and action that they figure in reason-relations like Scanlon's.

  We've seen that the relation of conceptual dependence that arises between the ideas of *normative reason* and *rational animal* is symmetric, so that the two bestow intelligibility on one another. The immanence constraint derives from this dependence as it runs in one direction only, from the former idea to the latter. But, in spelling the constraint out, it's inevitable that we'll stumble over that same relation running the other way—the idea of *normative reason* must admit rational animals to its relation, but then, the idea of *rational animal* is likewise compelled to find its home there.

This line of justification secures the immanence constraint as it bears on conceptions of *normative reason* that are like Scanlon's in that they i) are relational, specifically such that they ii) incorporate agency as a *relatum*, and still more specifically such that they iii) represent the agential orientation to reasons, the orientation in virtue of which agency is admitted to the reason-relation, as the exclusive province of a special variety of thinking, what I've been calling 'practical thought'. Of course, this last condition makes the application of the immanence constraint to conceptions of *normative reason* trivial. The question is why we ought to accept it.

The considerations that recommend i) and ii) as properties that characterize normative reasons are compelling, indeed irresistible, on independent grounds. We might go so far as to say that metanormative theorizing about the idea of *normative reason* must grant its relationality, and specifically, its agent-relationality,[10] if it's to be recognizable as such. At any rate, many of the extant conceptions of *normative reason* regard the claim that normative reasons are marked by i) and ii) as something like a platitude.[11] After all, normative reasons are deontic, and though some axiatic entities[12]—for instance, cardinal values—are monadic properties, all deontic entities—duties, prohibitions, oughts, etc.—are, at least, dyadic. Hence, i). And it belongs to the nature of deontic entities that they prescribe behaviour to agents, so that, necessarily, one place in each deontic relation is reserved for agency. Hence, ii).[13]

---

[10] This term, 'agent-relational', is an invention of mine. It's distinct from the more familiar 'agent-relative'. My term expresses the property of being essentially constituted by relations involving agency, while 'agent-relative' expresses the property of being essentially related to some particular agent (see Thomas Nagel's *The View from Nowhere*, page 153). Some normative reasons are agent-relative, and others are wholly agent-neutral, but all are agent-relational.

[11] The early conceptions of *normative reason*, developed by Nagel and Joseph Raz, meet i) and ii) quite tidily. Both include two-place reason-relations: Nagel's incorporates outcomes of actions, described as "events", and agents, described as "persons" (*The Possibility of Altruism*, page 47); Raz' is nearer to Scanlon's in that it incorporates facts and agents, again described as "persons" (*Practical Reason and Norms*, page 19).

Some conceptions of *normative reason* satisfy ii) obliquely, say, by assimilating agency into an action *relatum*. E.g., Terrence Cuneo's reason-relation doesn't specify a place for agency as such, but it does include a "Response R of S" *relatum*, in which S is an agent (*The Normative Web*, page 65).

[12] I use the term 'axiatic' to refer to the domain of value, and to the language that describes that domain. I take 'axiological' to refer specifically to the study of axiatic nature, just as 'deontological' refers to the study of deontic nature.

[13] In making these claims, I rely on Immanuel Kant's treatment of the nature of deontic entities. He says that "[a]ll imperatives are expressed by an *ought* and indicate by this the relation of an objective law of reason to a will that by its

I don't anticipate resistance to i) and ii), so I won't attempt to shore them up. I take for granted that to deny either is to depart altogether from the idea of *normative reason* and take up some other notion in its stead. In any case, iii) is somewhat more ambitious than the other two, and so somewhat more vulnerable. I devote the next section (1.3) to its defence. The gist of the argument is as follows: ii) must be elaborated somehow, specifically so as to identify the principle that admits agency to the reason-relation. By foregrounding agents' capacities to represent their reasons in practical thought, iii) offers one such elaboration: reasons are addressed to agents because agents can, in thinking about their reasons, exhibit their rationality in being motivated to conform to them. There are alternatives principles, but they aren't, I think, persuasive. I take it, then, that each conception of *normative reason* must incorporate i), ii) and iii), and that the narrow formulation stands: the immanence constraint conditions *all* conceptions of *normative reason*.

1.2.2 The immanence constraint and the normative domain

But before presenting the details of that argument I'll say a bit about what bearing it has on the rest of the normative domain. Clearly, it can be redeployed *mutatis mutandis* for certain other deontic entities, like duties and oughts. We've just seen that each deontic entity is agent-relational. What remains to be shown is that these other entities are intelligible only if agents can take them up as objects for practical thought. But this is easily done, since duty-relations, ought-relations and reason-relations each draft agency into the same role, involving the same receptivity to normative authority.

---

subjective constitution is not necessarily determined by it" (*Groundwork*, page 413). This establishes at least i) and ii), (granted, only for imperatives, but the lesson can be carried over to normative reasons). As for iii), Kant's ambition in the *Groundwork* is, in effect, to demonstrate that iii) is a feature of duty-relations, and to thereby establish the possibility of our exhibiting a good will by acting from practical thought about the moral law. I think he succeeds in this, and my argument follows his in several places.

It's not quite so clear, though, that this same transposition can be made for the remaining deontic entities. Permissions, privilege-rights and power-rights relate to agency by granting it the latitude to determine itself, rather than by specifying the course it's to take. To be granted latitude is not—anyway, not obviously—to engage one's capacity for practical thought. But the deontic ideas are themselves so thoroughly intermeshed that what carries for any one, carries for the rest. After all, a permission to φ is just the negation of a duty to not-φ. If the one notion answers to the immanence constraint, the other does as well. I won't attempt to map the whole of the deontic subdomain to show that this same concatenation runs through each of its provinces. It strikes me, though, that it's reasonable to suppose that it does. This justifies the expansion of the scope of the immanence constraint to all deontic entities.

As for the other normative notions—that is, the axiatic ideas,[14] like *good* and *virtue*—few are agent-relational,[15] so the immanence constraint has little basis for direct application. But there are bridges that connect the axiatic and deontic subdomains, and the constraint can be transmitted along one such bridge. Plausibly, it's essential to axiatic entities that they entail certain deontic entities, e.g., the goodness of a potential event entails a *pro tanto* normative reason to bring that event about. If this is right—that is, if each axiatic entity has a deontic entailment—then axiatic ideas too are answerable to the immanence constraint, albeit mediately.

I believe this *is* right, though I can only gesture to the shape the argument might take, namely, that there must be some manner of transaction between the two subdomains, if just because the unity of the normative domain requires it. Moreover, the passage from a positively valenced entity to a reason to act, and from a negatively valenced entity to a reason to omit, is a *prima facie*

---

[14] For my purposes, the axiatic and deontic concepts together exhaust the normative domain.
[15] Some are relational but only incidentally related to agents, like *better*. Others, like the moral virtues, are essentially agential, but are properties, rather than relations. It takes some digging to arrive at an essentially agent-relational axiatic idea. Aristotle's conception of *happiness* qualifies, as does Rawls' idea of *primary good*. If there are varieties of friendship, justice and solidarity that are specific to society between agents, these qualify as well.

natural one. If this argument, or one relevantly like it, fails, the axiatic ideas may escape the immanence constraint. But I take it as granted that some version of an essential deontic-mediated relation between axiatic entities and agency obtains, so that the argument, in basic outline, goes through. Thus, I take the whole pantheon of normative concepts to be conditioned by the immanence constraint.

Reformulating the immanence constraint at that level of generality, we get: *if a normative conception is to succeed in capturing its corresponding idea, it must establish the possibility of practical thought about the entities that instantiate it.* Call this the thin-wide formulation, in contrast to the thin-narrow formulation of the preceding subsection. (I offer this formulation with a view to defending a move I make in chapter 4, namely, my drawing on the immanence constraint to attack Moore's conception of *goodness*. This attack would misfire if the immanence constraint can't be given a sufficiently wide formulation. That said, in the context of this dissertation, it isn't strictly speaking necessary that this defence of the wide formulation succeeds. Chapter 4 isn't solely, or even primarily, designed to defeat Moore's conception (it's no longer in currency, so its (renewed) defeat would be of little consequence).)

1.2.3 Agent-independent normative entities

A last consideration pertaining to scope: some claim that there are axiatic notions that could find application even in a world forever devoid of agents. Moore, following Henry Sidgwick, devises an isolability test for the axiatic subdomain, one that limits intrinsic goodness to "what things are such that, if they existed *by* themselves, in absolute isolation, we should yet judge their existence to be good".[16] Moore's test is famously difficult to run—it seems to require that its objects be conceivable

---

[16] *Principia Ethica*, §112.

in isolation, but nothing can accomplish that since nothing is conceivable without a context. Suppose we want to assess a bucolic landscape for intrinsic value. What would it be for the landscape to exist by itself? It must, at least, have some space to occupy.[17] But here's a variant that's, at least, suggestive: each intrinsic value could obtain even if it were the case that no agency ever did. I think this sort of counterfactual captures Moore's insight about the apparent ontological autonomy of intrinsic value. But if he's right that intrinsic values have this property, this would show them to be wholly agent-independent, and so, it would seem, not even indirectly agent-relational.

The suggestion isn't that agent-independence applies across the whole set of intrinsic values. Moore acknowledges the possibility of intrinsically valuable agency; his own relatively expansive consequentialism assigns intrinsic value to society, aesthetic contemplation, and true belief, each of which is sensibly classified as a variety of agency. Obviously, intrinsically valuable agency can't be agent-independent. But even if the idea is only that *certain* intrinsic values are agent-independent, these values would have to be rather odd. As Scanlon explains, "normative claims would not have the significance that we normally attribute to them if there were no rational agents… the existence of such agents is a presupposition of the practical domain".[18] This is because normative entities depend on agency for their practicality. Without practicality, claims applying deontic concepts would be nonsensical, and claims applying axiatic concepts would have no deontic counterparts. The resulting purely axiatic discourse would be severely diminished, able only to evaluate—to say of things that they're 'good' or 'bad', 'better' or 'worse'—and even then, without any of the prescriptive or justificatory force ordinarily reserved for these terms.[19] If we're to be at all faithful to the richness

---

[17] Moore likes to lean on his idea of *organic unity* (*ibid*., §§18-22; this is roughly the notion of an axiatic gestalt, of a value's being more than the sum of its parts) when working through his test, but it's hard to see how it's any help here.
[18] *Being Realistic about Reasons*, page 22.
[19] I elaborate on the role of practicality in the normative domain in subsection 1.3.3.

of normative discourse as we know it, we must secure some degree of agent-dependence throughout the normative domain.

Still, this is compatible with a segment's being agent-independent. Anyway, if we're to prevent the immanence constraint from trespassing on normativity's special jurisdiction—its authority to prescribe to agency without its being at all answerable to that agency—we shouldn't foreclose the possibility that some normative entities are agent-independent. In making allowances for this possibility, we ensure that the immanence constraint doesn't compromise normative transcendence.

It's a relatively simple thing to open a space for agent-independence, even given immanence: interpret the constraint's phrase 'possibility of practical thought' such that it's satisfied by the possibility of practical thinkers, that is, of agents who might engage in practical thought. That is, in the counterfactual world in which a value obtains but no agents do, the constraint should be read so as to require that another counterfactual be true, namely, that if agents were introduced to this world, that value would be practically thinkable by those agents. Understood this way, the immanence constraint can tolerate normative entities that are only hypothetically practical. The sense of possibility in play is conceptual—even if the causal nexus were such as to make agency substantively impossible, so long as the prospect of agency remains conceivable, the immanence constraint raises no objection. I believe this delivers everything that the advocate for agent-independent value could want.

Note that the constraint is already subjunctivized once, in that it requires that normative entities be think*able*, but not necessarily thought. This accommodates normative transcendence by establishing the possibility of our misrepresenting those entities, or failing to represent them at all.[20]

---

[20] Transcendence requires that we capture, to the extent that we're able, this Kantian observation from pages 407-408 of the *Groundwork*:

Subjunctivizing a second time, so that normative entities need only be conceivably practical, and not necessarily actually practical, accommodates transcendence still further. But this second step toward normative transcendence doesn't weaken the constraint. Immanence only tethers normativity to practical thinkability as such, and not to what we in fact do, or even whether we in fact are.

1.3 DEFENCE OF THE THIN FORMULATION

I said that iii) is a property characterizing normative reasons such that the agential orientation to those reasons is expressed by a special variety of thinking. The variety I mean is practical thought, which is exhibited, insofar as we are rational, in our representing a normative entity in such a way as to be (defeasibly) motivated to conform to it (or rather, to conform to it if it's deontic, or to conform to its deontic counterpart if it's axiatic). In this section, I justify iii) by showing that it's the *only* workable elaboration of ii). We've seen that any adequate normative conception grants ii), and is thereby saddled with the task of identifying the principle according to which it applies, i.e., the one that explains ii) as a non-arbitrary feature that characterizes, in its different guises, the whole of the normative domain. If iii) is the sole basis for this principle, then we must accept it, on pain of leaving ii) unelucidated.

I confess from the outset that my argument for iii) is circular. It depends on intuitions about the circumstances that must obtain if an agent is to be admitted to the reason-relation, intuitions that

---

nothing can prevent us against falling away completely from our ideas of duty and can preserve in our soul a well-grounded respect for its law other than the clear conviction that, even if there never had been actions arising from such pure sources, what is at issue here is not whether this or that happened; that, instead, reason by itself and independently of all appearances commands what ought to happen; that, accordingly, actions of which the world has perhaps so far given no example, and whose very practicability might be very much doubted by one who bases everything on experience, are still inflexibly commanded by reason; and that, for example, pure sincerity in friendship can be no less required of everyone even if up to now there may never have been a sincere friend, because this duty—as duty in general—lies, prior to all experience, in the idea of a reason determining the will by means of a priori grounds.

Later (in subsection 2.2.3 and section 4.4), we'll see that the theory of practical content can't allow for moral failure as total as this, but this doesn't prevent us from preserving the spirit of the Kantian image of normative transcendence, whereby normative entities prescribe without their being sensitive to what we've done.

are themselves only articulable via appeal to the idea of *practical thought*. That is, the argument takes

for granted that the ideas of *normative reason* and *practical thought* are intertwined. This isn't as

embarrassing as it may at first appear since, at this depth, the most that can be accomplished is to

map the constituents of the idea of *normative reason*. And, perhaps, in drawing up a tidy map, we

vindicate the intuitions we use to draft it. Still, my argument occupies a position that's, *prima facie*,

rather awkward, in that no one who refuses its intuitions could be moved by it, and anyone who

accepts its intuitions has no need for it.

Even this awkwardness is, I think, only superficial. As it happens, I can't envision a

philosopher who would deny iii)—not without resorting to caricature.[21] Maybe Derek Parfit so

abhors (what he takes to be) the psychologizing of reasons that he'd refuse to countenance any

essential connection to thought. I say this in light of his astonishing interpretation of Bernard

William's desire-based conception of *normative reason*, according to which that conception isn't, in

fact, a conception of *normative reason* at all, but is rather a conception of some nearby homonym.[22]

Such a remarkable claim needs special explanation, and this can perhaps be given by attributing to

Parfit the view that the reasons that figure in thought are altogether distinct from the ones that

command normative authority. But as far as I can tell his objection only extends to the proposal

that normativity is essentially connected to what we *in fact* think. As we've just seen, the relevant

idea of *practical thought* deals in possibility, and not actuality. And in any case, I suspect that even

Parfit would allow that, though normative force is in no way answerable to agency, normative

---

[21] But consider the discussion of judgment externalism in section 3.2.

[22] Parfit says about Williams' conception, as articulated in papers like "Internal and External Reasons", that: "[w]hen Williams makes claims about reasons for acting, he may seem to be using the phrase 'a reason' in the indefinable normative sense that we can also express with the phrase 'counts in favour'… This interpretation is, I believe mistaken. Williams did not understand this concept of a reason" (*On What Matters*, vol. II, pages 433-434). Also: "[t]hough Williams and I used the same normative words, we used them in different senses. We were not really, as we assumed, disagreeing" (*ibid.*, page 448). I say more about Parfit's conception of *rationality* in subsection 1.3.2, and about Williams' reasons-internalism in section 3.1.

reasons for action are essentially addressed to agents in such a way as to make the performance of those actions rational, and this, in effect, acknowledges iii).

So, for lack of a genuine interlocutor, it isn't clear what the criterion for a successful defence of iii) might be. I offer the argument all the same because, as I see it, what makes the argument worthwhile isn't its capacity to persuade, but its operating as a kind of expository device, an exercise with which to illuminate the extent to which the idea of *rational animal* figures among the underlying machinery of the idea of *normative reason.* It reveals the formal bases of the immanence of that idea.

1.3.1 The extensional reading of the reason-relation

For the most part, the argument proceeds by refuting the following proposal: Scanlon's conception states that "*p* is a reason for a person *x* in situation *c* to do or hold *a*". Why not read this schema in purely extensional terms? Read that way, it needn't presuppose the possibility of some *x*'s *representing p* as her reason to *a.* In fact, it needn't presuppose any capacity for practical thought at all, so that there's nothing there to justify the immanence constraint, even on its narrowest formulation. It's clear enough that this extensional reading isn't what Scanlon intends,[23] but that by itself doesn't settle the issue, since, for all we've seen, he might've done otherwise.

On my view, the best response is as follows: the extensional reading revises the idea of *normative reason* to such an extent as to eliminate it. After all, normative reasons can be taken up as objects for deliberation, they can be cited as *explanantia* for our actions, and as a basis for appraising behaviour.[24] Each of these—guidance, explanation and evaluation—presupposes that reasons

---

[23] The evidence is easy to find. Among other things, Scanlon says "it is the point of view of internal reasoning that is primary in an investigation of reasons and normativity" (*Being Realistic about Reasons*, page 14), that is, the point of view of "reasoning about what reasons one has oneself" (*ibid.*, page 13; the idea of *internal reasoning* is Gilbert Harman's). (Of course, the idea of *internal reasoning* might itself be given an extensional reading, but this would be a profoundly distortional reading.)

[24] I elide the distinctions between explanatory, operative and normative reasons. This isn't misleading since the normative variety is primary; the others are derived from it.

operate on agents by being represented by them. Eject this presupposition, and we're liable to confuse a scratch in the throat for a reason to cough.

To be sure, the issue is complex. After all, at times, a scratch in the throat *is* a reason to cough. And on the other side, it can happen that, upon receiving a doctor's instruction to cough, it's fear of the diagnosis that causes the cough, rather than the agent's representing and responding to the normative force of her reason to cooperate with her doctor.[25] In the first case, we have agency where we'd expect mere impulse, and in the second, we have impulse in place of agency. But these eccentric cases don't make trouble for the boundary between agency and impulse. If anything, they underscore it by tracking our expectations as to when practical thought is and isn't present.

Of course, the extensionalist line might be expanded to assimilate practical thought as well. This would trigger a new complaint about having lost *that* notion, but then, this complaint would have to be framed in terms of some other idea, which might itself be assimilated, and so the dialectic would wind along. It's useless to pursue that conversation very far. In the end, we must acknowledge that, while the idea of *practical thought* can be reframed in normative and agential terms, it can't be reframed in simpler terms than those. Anyone who undertakes to reduce the whole gamut—rationality, agency, normativity—has simply overlooked a region of the world. At a certain stage, the only available response is to say of that region "look, it's here", and leave it at that.[26]

So, I set aside the global extensionalist line, and focus on the local one. For the most part, reductive proposals aspire to leave discourse about their corresponding ideas wholly unaltered. They are defeated in this ambition if they entail any revision of that discourse. Evidently, the extensionalist line entails far-reaching revision, so to this extent, it fails. But my objection goes further than that: it holds of the extensional reading that it smuggles in eliminativism with respect to

---

[25] This is Donald Davidson's 'deviant causal chain' problem; see "Freedom to Act", page 79.
[26] That is, the only response is to espouse non-reductionism about *practical thought*. But see section 2.1.

the concepts that constitute normative talk.[27] The revision of the idea of *normative reason* under consideration is so extensive as to effectively jettison that idea from all action-explanatory discourse. And given the centrality of that idea in action-explanation as it stands, this would result in a radical transformation in the discourse; it'd have to swap out its entire set of *explanantia* and *explananda*, leaving us with a wholly different sort of talk.

What makes the extensional reading so distortional is its neglect for the fact that the idea of *conformity to reason* is distinct from that of *action from reason*,[28] and that the sense of 'reason' in play is to be understood as giving priority to this latter idea. We have other terms—they're axiatic rather than deontic—with which to account for the former idea. The relevant contrast between the *conformity to* and *action from* relations can be specified via a further idea, that of *rationality*. This idea may be absent in cases of mere conformity, but must be present in cases of action from a reason. As we'll see in the following three subsections, the clue to finding the idea of *practical thought* within the idea of *normative reason* lies here, in the fact that rationality is essential to action from a reason.

1.3.2 Two conceptions of *rationality*

Consider two rival conceptions of *rationality*. On the first conception, which Scanlon favours, an agent's thinking and doing is rational only if it aligns (non-accidentally) with her judgments as to what she ought to think and do.[29] The second is simpler; it holds that an agent's thinking and doing is rational only if it aligns with what she ought to think and do. This is an extensional conception of

---

[27] Peter Railton suggests that the fact that a reductionist project is revisionist doesn't by itself defeat that project, not if the revision is "tolerable", i.e., if it doesn't eliminate its concept ("Naturalism and Prescriptivity", pages 159-161). He may well be right about that, but the reductionism considered here is, I think, eliminativist, and so, problematic even by Railton's lights.

[28] That this distinction is real is, I think, plain. It's akin to the reality of the distinction between Kant's two standpoints (*Groundwork* 451-453), as well as the one between those cases in which Anscombean 'why?' questions are apt and those in which they're not (see *Intention*, §§5-6).

[29] "Irrationality in the clearest sense occurs when a person's attitudes fail to conform to his or her own judgments" (*What We Owe to Each Other*, page 25).

rationality—the relevant sense of 'aligns with' is satisfied by mere conformity, so that an agent can meet this condition simply by doing as normative entities prescribe, no matter how, or even whether, she depicts those entities. I'll call the first the 'mediated conception' since it involves a mental intermediary, like a judgment. The extensional conception involves no such intermediary, and so I'll call it the 'direct conception'.

Here are two preliminary observations: both conceptions deal in evaluative rationality rather than classificatory rationality, and both offer necessary conditions rather than sufficient conditions. That last point deserves emphasis; John Broome seems to attribute a biconditional conception to Scanlon, according to which he identifies both necessary and sufficient conditions for rationality.[30] I believe Scanlon's conception is a good deal less ambitious than that, in that it offers material for a single conditional only, one establishing that, if someone is evaluatively rational, then she thinks and acts as she judges that she ought to think and act. Moreover, as I understand it, his conception isn't designed to generate a comprehensive inventory of each of the principles of rationality. It remains a possibility, for all he's said, that, e.g., it's irrational (under conditions conducive to adequate attention and memory) to simultaneously accept contradictory claims no matter the content of one's judgment, i.e., even if one judges that one ought to accept the contradiction.[31] This same possibility is available to adherents of the direct conception.

Because our business is necessary, and not also sufficient, conditions, it's possible to endorse both conceptions, that is, to hold that it's necessary for rationality that our thoughts and actions *both* proceed from our judgments and align with our reasons. Still, I depict the two conceptions as rivals.

---

[30] "Does Rationality Consist in Responding Correctly to Reasons?", page 362.
[31] Scanlon might also accept Davidson's principle of continence: "perform the action judged best on the basis of all available relevant reasons" ("How Is Weakness of the Will Possible?", page 41). With respect to action, this principle is quite a lot more demanding than the mediated conception, since it confines rationality to the performance of what we judge *best* rather than simply what we judge in some way worth doing. But, while the mediated conception is likely too weak by itself, the principle of continence might be too strong. Much depends on how 'judged best' is interpreted. I won't attempt to address these issues here. Our immediate concern is whether rationality involves mental mediation, and for that purpose, the mediated conception will serve.

That's for two reasons: first, I assume that the advocate of the direct conception maintains that the idea of *rationality* can be captured entirely in extensional terms, and is drawn to the direct conception for that reason. The mediated conception is intensional in the sense that the relevant mental intermediaries represent their normative objects in particular ways, such that swapping out co-referential terms may affect the truth-conditions of the claims that express those intermediaries. The devotee of extensional rationality rejects the mediated conception on that ground. Second, my guess is that Scanlon is himself opposed to the direct conception, and that this is at least in part because it produces awkward results (as we'll soon see), which he sought to avoid by opting for the mediated conception instead. To bring this out, we can add to each conception a clause denying that the other is a necessary condition for rationality.

In practice, the two conceptions diverge in this way: on the mediated conception, if I judge that the cat's being on the mat is decisive reason to praise her, then my failing to praise her is irrational. This is so even though, *ex hypothesi*, the cat's being on the mat is decisive reason to leave her undisturbed.[32] On the direct conception, praise would be irrational since the reasons that bear on my conduct prohibit it. Meanwhile, my opposition to my own judgment may be rational if it produces the mandated result. This remains the case even if I must resort to sabotaging myself in order to ensure that my judgment doesn't determine my behaviour.[33] Perhaps more remarkably, on the direct conception I would qualify as rational (at least with respect to this issue) even if I don't

---

[32] By 'decisive reason' I mean the one that determines what I ought to do in a given circumstance. In the case I have in mind, I likely have *some* reason to praise the cat, but it's defeated by this other reason, the one deriving from the value of leaving her undisturbed. All reasons are at least *pro tanto*, that is, "compelling reasons unless outweighed by other, better reasons, but they can be outweighed without losing their force or status as reasons" (*What We Owe to Each Other*, page 50). So, my reason to praise the cat retains its normative force even though I ought not to act from it. Its force can be given this counterfactual formulation: it would have been decisive if it hadn't been defeated by the other reason, the one prohibiting my disturbing the cat.

[33] I have in mind Nagel's account of the bizarre consequences of certain desire-based theories of *practical reason*. According to these theories, if it happens that I have a desire to realize some future state of affairs that I know I will later have a desire to prevent, "I may have reason to do what I know I will later have reason to try to undo, and will therefore have to be especially careful to lay traps and insurmountable obstacles in the way of my future self" (*The Possibility of Altruism*, page 40).

disturb the cat only because I'm unaware of her being on the mat, because I'm sleeping, say, or across town.

I take this example to show that the direct conception goes wrong in at least three places—it's in one respect too demanding, and in two other respects too liberal: it's too demanding because it classifies putatively rational behaviour as irrational, and it's too liberal because, for one, it classifies putatively irrational behaviour as rational, and, even more troublingly, it classifies arational phenomena as rational. The first two objections deal in evaluative rationality, and the last deals in classificatory rationality. I consider the first objection next, before turning to the third objection in subsection 1.3.3. I leave the middle objection until section 1.4.

Broome articulates the first objection in this way: "[s]uppose your reasons require you to *F*, but you are ignorant of those reasons. Suppose you are not at fault in being ignorant; you have no evidence of the reasons. If you do not *F*, you might nevertheless be rational."[34] Recognizing this possibility, Scanlon confines his charges of irrationality to cases involving contradictions in one's own reasoning, as manifested in, say, one's judgment that one ought to φ on the one hand, and one's opting not to φ on the other.[35] It's this appeal to reasoning, as manifested in normative judgment, that makes his conception mediated, rather than direct.

Of course, normative judgments are by their nature trained on the normative domain. Thus, Scanlon's conception of *rationality* is like the direct conception in that it's, at least ultimately, about normative entities like reasons. The difference is that it takes his conception two steps to reach the

---

[34] "Does Rationality Consist in Responding Correctly to Reasons?", page 352.
[35] *What We Owe to Each Other*, page 25. That there's a contradiction here, between a cognitive attitude (a judgment) and a conative attitude (a motivational state), is a matter of controversy. Scanlon seeks to secure the relevant sense of 'contradiction' by characterizing (perhaps not all, but at least many) conative attitudes as "judgment-sensitive attitudes", that is, "attitudes that an ideally rational person would come to have whenever that person judged there to be sufficient reasons for them and that would, in an ideally rational person, "extinguish" when that person judged them not to be supported by reasons of the appropriate kind" (*ibid.*, page 20). The device of 'an ideally rational person' serves to model the fact that cognitive-conative 'contradiction' (or incoherence, tension, mismatch, whatever it is) amounts, at least in most cases, to irrationality.

normative domain, as opposed to just one. The introduction of the mental element doesn't do much to alter the normative orientation of rationality. All the same, it's a significant change. Broome says, I think correctly, that it's "a fundamental feature of rationality… that your rationality depends on the properties of your mind."[36] Because judgments are rendered, so to speak, within the mind, mediated conceptions are well-positioned to capture this feature, while direct conceptions must neglect it.

In the course of his articulation of his conception of *rationality*, Scanlon identifies Parfit as his foil.[37] It's worth mentioning that Parfit's treatment of rationality in *Reasons and Persons* isn't tailored to defend the extensional reading of the reason-relation. For that matter, Scanlon's treatment in *What We Owe to Each Other* isn't tailored to repudiate it. Rather, both are, in their different ways, deployed with a view to defeating Williams' desire-based theory of *normative reason*.[38] Moreover, Parfit's view is a good deal more nuanced than the direct conception I've drafted. All the same, his animating concern, together with Scanlon's response, is relevant here: Parfit argues that the mediated conception is itself too liberal because it raises no complaint against behaviour that's clearly absurd, like failing to assign normative significance to pain experienced on future Tuesdays. That is, on his view, it's conceivable that one can exhibit indifference to the quality of one's experiences on future Tuesdays without acting contrary to any of one's judgments, or for that matter, to any aspect of one's thinking.[39] Since this seems to show the mediated conception to be compatible with obvious irrationality,[40] it warrants, he argues, a shift toward the direct conception,

---

[36] "Does Rationality Consist in Responding Correctly to Reasons?", page 352. Scanlon endorses this principle at page 14 of *Being Realistic about Reasons*.

[37] *What We Owe to Each Other*, page 26, citing a passage from Parfit's *Reasons and Persons* at page 119.

[38] A central feature of Scanlon's objection to Parfit's view is that it gives ammunition to Williams' argument that external reasons claims can only be "bluff" (see *What We Owe to Each Other*, pages 27-28, and Williams' "Internal and External Reasons", page 111).

[39] *Reasons and Persons*, pages 123-124.

[40] If, as I've maintained, the mediated conception makes no claim to comprehensiveness, Parfit's objection requires that we assert that there's no capturing the irrationality of future Tuesday indifference except via recourse to the direct conception. I'm not certain how this last move could be sustained, especially given the Korsgaardian option of finding it

or anyway, toward a conception according to which charges of irrationality can emanate from certain normative entities—like the prudential reason to avoid the onset of future pain, no matter when it occurs—without mental mediation.[41]

Scanlon holds to the mediated conception, even in the face of Parfit's future Tuesday indifference example.[42] A part of his rationale is that, ordinarily, our evaluations of agents' responsiveness to reasons is two-tiered: we approve of agents that are properly sensitive to the reasons that bear on their conduct, but, failing that, we may yet find some virtue in agents who are sensitive to their judgments, even their mistaken judgments, about their reasons. We reserve the term 'rational' for this second variety of agential success.[43] If the error in future Tuesday indifference doesn't arise at this lower, more formal tier, we can find it in the higher, more substantive tier. But if we do, the error is properly characterized as a case of normative misjudgment, rather than irrationality.

As I've said, fidelity to one's own judgments isn't by itself sufficient for rationality. Plausibly, one must also be disposed to honour certain of the logical entailments of one's judgments, at least in circumstances in which those entailments are painted in bright colours.[44] But, crucially,

---

in the nature of agency that rationality requires that we preserve our practical integrity over time (*Self-Constitution*, pages 185-188). Still, it can be granted that there's at least a surface plausibility to Parfit's strategy.

[41] Parfit's conception in *Reasons and Persons* assesses desires, and only desires, for irrationality. As such, it's narrower in scope than the direct conception. He introduces a new conception, one with wider scope, in the first volume of *On What Matters*, which is something like a hybrid of the mediated and direct views—partly mediated because it holds that "[w]hile reasons are given by facts, what we can rationally want or do depends on our beliefs" (page 111), but partly direct because it holds that some beliefs may be deemed irrational on wholly reasons-given grounds (*ibid.*, page 121).

[42] *What We Owe to Each Other*, pages 29-30. Parfit's example is infamously artificial—it's difficult to envision what a person who's indifferent to future Tuesdays could be like, let alone to evaluate that person for rationality. Scanlon raises concerns along these lines at *ibid.*, page 29. Sharon Street wrestles with the example throughout pages 281-292 of "In Defence of Future Tuesday Indifference", and concludes, *contra* Parfit, that either there's no irrationality in it (page 288), or there is, but it's of a more familiar kind (page 291), rather like Nagel's self-sabotage cases. Either way, the mediated conception encounters no special difficulty.

[43] It's because of this two-tiered structure that Scanlon separates "the form our thinking must take if we are to avoid the charge of irrationality" from "what we have most reason to do" (*What We Owe to Each Other*, page 30).

[44] I interpret Scanlon's treatment of rational dispositions (*ibid.*, pages 23-24) as licensing this extension to the mediated conception. The 'painted in bright colours' clause is designed to forestall the proposal of another too-demanding conception of *rationality*, according to which one is irrational if one, e.g., fails to recognize the validity of a logical proof, even a very complicated one. Perhaps some logical errors are irrational, but many aren't. Scanlon's example involving

whatever other rational requirements we might uncover belong to this lower tier. Moreover, I suspect each exhibits the feature Broome identifies as belonging to the idea of *rationality*, namely, dependence on the properties of one's mind. As we've seen, if I judge that the cat's being on the mat is decisive reason to praise her, I must do so or else warrant a charge of irrationality. Likewise, if I judge that I ought to praise the cat if *p*, and also judge that *p* (and a set of *ceteris paribus* clauses obtain, e.g., I'm not distracted, the two judgments are made so close in time as to eliminate the possibility of my having forgotten one of them, etc.), then again, I must respect *modus ponens* and praise the cat or else warrant another charge of irrationality. Both, and I would guess all, rational requirements can be fitted under the umbrella of an expanded mediated conception. But if in fact I have decisive reason to leave the cat alone, and I neither judge this to be the case, nor can I find my way to it via some nearby entailment, then I have only misjudged, and am not therefore irrational. There are no rational requirements that call for the direct conception.

1.3.3 Rationality and intension

I believe the two-tiered model for agency appraisal is so compelling, just in terms of its capturing the idea of *rationality*, that it can by itself defeat the direct conception and confirm the mediated conception. I doubt, though, that I can count on everyone to agree. The trouble is that the model is vulnerable to the charge that it begs an important question, in that it draws on (what I take to be) the *prima facie* plausibility of the claim that it's irrational *not* to praise the cat when I judge I ought to, even if in fact the relevant reasons prescribe otherwise. It may be argued on behalf of the direct conception that this plausibility is only apparent (some might deny even that). The model still warrants our attention, if just because, if it's right, it has a lot to tell us about what it is to be a

---

Fermat's Last Theorem reinforces this point (*ibid.*, page 26). See also Philip Pettit and Michael Smith's "External Reasons", pages 153-154.

rational animal. But if my aim is to defeat the direct conception—and in this section, it is—I need a different line of objection.

I said several paragraphs ago that the direct conception goes wrong in another place—it classifies arational phenomena as rational. Since its criterion is simple conformity to the relevant normative entities, and since this *conformity to* relation is wholly extensional, the direct conception is compelled to call me rational even if I realize the outcome prescribed by those entities entirely by accident. Try a new example: suppose that I ought to remove the cat from the mat, and that, by falling down the stairs, I bring this about (let's say the fall produces a frightening noise, and this causes her to find cover under the couch). It's enough, as far as the direct conception is concerned, to have the requisite entities on both sides—on the one side, an agential doing, namely, a falling down the stairs, and on the other, an entity prescribing that same doing, albeit described in terms of causing the cat to depart from the mat. There's nothing in the direct conception to block its approval, nothing that attends to whether or not my agency is causal *in the right way*.[45] This is a familiar pitfall for extensional accounts: they can collect all the right players together, but they can't make them adhere to the rules.

Thus, though the direct conception guarantees that each instance of rationality is accompanied by an instance of agency, it doesn't have any means of excluding cases in which the two are only arbitrarily related. Though I remain an agent mid-fall, my agency isn't manifested in the fall. My causal contribution is, we might say, wholly physical, and not agential; it adds up to movement, but not action. In fact, that I'm an agent is altogether incidental here. For the purposes of frightening the cat, I might just as well have been a stone.

---

[45] This is another example of Davidson's deviant causal chain problem. This phrase, 'in the right way', is taken from his "Freedom to Act" at page 79.

Or suppose that what prompts the cat to relocate is a ray of sunlight that's shifted to the opposite corner of the room over the course of the afternoon. Sensibly, the direct conception confines attributions of rationality and irrationality to agents, but it's only due to the, as it were, *ex machina* intervention of this provision that the direct conception doesn't call the ray of sunlight rational. And what's most damning of all: it can't account for this provision. It has no resources with which to do so. Its connection to agency is entirely *ad hoc*.

This line of objection shows the direct conception to misunderstand classificatory rationality. And a misunderstanding of this kind is, I think, fatal. It's also one from which there's no easy escape. Recall that, on the direct conception, if an agent is rational, her thinking and doing aligns with what she ought to think and do. I've interpreted the term 'doing' such that it has no essential attitudinal component; it's in this sense that my inadvertently falling down the stairs is a doing. Of course, we might instead have reserved the term 'doing' for actions. Reinterpreting the direct conception in this way offers some protection against the charge that it must call the arational rational, since it's in the nature of actions that they fall on the rational, rather than the arational, side of the classificatory divide. But this protection, such as it is, comes at the cost of renouncing the ambition that I earlier assigned to the direct conception, which is to articulate the idea of *rationality* in purely extensional terms.

And anyway, the protection it provides doesn't amount to much. Let's say that I throw myself down the stairs in order to give expression to some frustration or other. That my fall is deliberate doesn't diminish its efficacy in causing the prescribed result—the cat still departs from the mat. So, according to the direct conception, my action is to that extent rational. There are two possibilities: if I judge that I ought rather to vent my frustration in some less ridiculous way, I'm *akratic* in throwing myself down the stairs. In that case, the direct conception goes wrong in

characterizing irrational behaviour as rational.[46] But more importantly, if my throwing myself down the stairs is by my lights altogether justified, then it may well be rational for me to do it. But then the complaint is that the direct conception finds my rationality in the wrong place—it finds it in my causing the prescribed outcome to obtain, and not in the intention underlying my action. Again, we've assembled the right players: a rational action on one side, and an entity prescribing that that action's outcome occur on the other. But the two aren't related in the right way.

I mention this last point to complete my indictment of the direct conception, but also to convey this lesson: we can make the doing *relatum* as narrow as we please—limit it to actions, hem it in with counterfactual clauses, and on and on. As long as the relation is extensional, the threat of errant cases remains. There's no dispensing with the phrase 'in the right way'.

Return to the mediated conception. Rays of sunlight can't issue judgments, so they provide nothing for the mediated conception to assess. And if, as seems likely, the capacity for judgment suffices for agency, then to this extent the mediated conception vindicates the provision confining attributions of rationality and irrationality to agents. But it does better than that, since it doesn't attend to the capacity for judgment just as such—after all, I retain that capacity as I fall down the stairs. What matters is the *exercise* of the capacity, what I in fact judge. In the first example I sketched, the one in which my fall is inadvertent, I made no judgment that pertains to my fall, so that, again, there's nothing for the mediated conception to assess. This shows that, by opting for that conception, we arrive at a principled basis for excluding the cases we want to exclude. But even this doesn't get to the crux of the issue. What if there's an amendment available to the direct

---

[46] In response to this worry, the direct conception could be supplemented with a second principle, one classifying all *akratic* behaviour as irrational. There's nothing embarrassing in depending on an anti-*akrasia* principle; anyway, the mediated conception needs one too. I wonder, though, how the direct conception and an anti-*akrasia* principle could be represented as belonging to a single uniform conception of *rationality*. The relationship between the irrationality of *akrasia* and the rationality of acting from one's normative judgments is quite tight (though perhaps not so quite as tight as it initially appears, since not all, and in fact remarkably few, failures to act from judgments are *akratic*). But there's no obvious relationship at all between the irrationality of *akrasia* and the direct conception.

conception, one that specifies exercises of agential capacity in extensional terms, and then limits assessments of rationality and irrationality to the cases thus specified?[47] We've just seen that, impressive though it otherwise would be, this wouldn't help. Simply registering that an agential capacity has been exercised doesn't deliver classificatory rationality. If my fall is deliberate, in order to determine whether it's rational or irrational—and crucially, to see what, if anything, *makes* it rational or irrational—we need to step inside my agential capacity, and observe it as it's working. That is, we need to occupy my perspective.

The chief merit of the mediated conception is its intensionality. This is what enables it to sustain *action from* relations, as opposed to mere *conformity to* relations. Judgments are attitudes, so they exhibit attitudinal opacity. Let's say the cat is the deepest thinker now living— if I praise the cat, then in so doing, I praise the deepest thinker now living. The same is true of my pleasing her, or of my performing the actions that conduce to her comfort. These things—praising, pleasing, comforting—aren't opaque. But in virtue of the opacity of judgment, my judgment that the cat is on the mat doesn't entail to the judgment that the deepest thinker now living is on the mat, not unless I know that the two are identical. To say the same thing, though *ex hypothesi* the sentence 'the cat is on the mat' is truth-functionally equivalent to the sentence 'the deepest thinker now living is on the mat', my judging that the first is true doesn't entail my judging that the second is true. My ignorance as to who the deepest thinker is prevents my judgment from spilling over from the one sentence to the other.

Because judgment is so discriminating, it provides us with a means of privileging the descriptions under which objects appear to the judger. By attending to objects specifically as described, and not as objects *simpliciter*, we depict those objects as they arise for the agent, from

---

[47] An amendment of this kind must sap the resulting conception of all explanatory force. This is because the specification of agency in extensional terms can't be sustained in any explanatory discourse (see Davidson's "Mental Events"). But we can bracket this consideration for the sake of argument.

within her perspective. Thus, my judgment that it would be good to please the cat grounds my *acting from* that goodness as I place her on the mat.[48] That action, once completed, also *conforms to* the reason prescribing that I comfort the deepest thinker now living. But I don't act from *that* reason, since I don't have it in view. In just this same way, Kant distinguishes between, on the one hand, a shopkeeper's undertaking to treat his customers fairly only with a view to advancing his business, and, on the other, his doing so out of respect for their humanity. He only conforms to the consideration that prescribes respect. The description under which the content of his judgment prompts his action is entirely prudential, and so, lacking in moral worth.[49] This same description-sensitive phenomenon arises in an example of Anscombe's: though the questions "why are you sawing a plank?" and "why are you making so much noise?" ask after the same action, there's nothing remarkable in my answering "I'm building a house" to the first and "I didn't know I was" to the second.[50] My action is recognizable to me under one description, but not the other. Likewise, the explanatory force of my reason—the one prescribing that I saw the plank in order to build the house—motivates me under one description, and not the other. Each of these things—rationality, morality, action (and, as I've been arguing, the reason-relation)—shares this intensional dimension.

If we neglected the agent's perspective, and endeavoured to account for her thinking and doing entirely from the third-person point of view, we'd find ourselves unable to distinguish an action from, e.g., behaviour that only coextends with some possible action. For that matter, an agent can mouth the words that form 'I judge that the cat is on the mat', without in fact drawing that judgment. She thereby gives the impression of being rationally responsible for certain of its

---

[48] For simplicity's sake, assume that my judgment has an object, and is correct in representing that object as good. Of course, some judgments lack objects, and other have them but misrepresent them, but this only underscores that we must deal in how things appear to the agent, rather than how they in fact are.

[49] *Groundwork*, page 397.

[50] *Intention*, §6. Davidson calls this the "quasi-intensional character of action descriptions in rationalizations" ("Actions, Reasons, and Causes", page 5).

entailments, though in fact she bears no such responsibility. Likewise, she can mouth 'it'd be good to place the cat on the mat', and appear to have made a normative judgment, complete with practical implications for action, but, in fact, assign no practical force to it. There's no irrationality in this. It can even happen that she judges in the speculative sense that she ought to place the cat on the mat, without making this same judgment in the practical sense. This would be to misunderstand the term 'ought', but this misunderstanding isn't immediately irrational. We're all too practiced in switching to the speculative register so as to acknowledge the truth of urgent, but onerous, moral obligations, without going so far as to give them any practical life. *This* is irrational, but only due to the self-deception it requires, the willful forgetting, where convenient, of the meanings of certain normative terms. It isn't practical irrationality.[51]

Genuinely practical judgments acknowledge the motivational significance of their contents.[52] This orientation to action is what warrants the designation 'practical'. And since the thought of the comfort of the cat may animate me *qua* practical agent, while at the same time the thought of the comfort of the deepest thinker does not (again, assuming I'm ignorant of the fact that the two are identical), the opacity of judgment translates to the opacity of practicality. Once we recognize this feature of practicality, we have what we need to carve out the domain of practical rationality: the idea of *practical thought* undergirds the *because* in the action-explanatory nexus—the one formed between agent and action, as in *this action because of this agent*—and it's this *because* that opens the space within which practical attributions of rationality and irrationality are intelligible.[53]

---

[51] Or anyway, from the armchair we can imagine a case like this one, in which the irrationality lies wholly in self-deception and not in anything directly pertaining to motivation. But in the familiar human case, we self-deceive at least in part in order to conceal the extent to which we're morally *akratic*, so that the irrationality is partly practical after all.
[52] Though we may fail to assign motivational force to those contents proportionate to what we judge to be their normative force. This approximates *akrasia* as Davidson understands it (see "How Is Weakness of the Will Possible?").
[53] See Davidson's "Actions, Reasons, and Causes" at pages 8-9.

It also closes the gap between agent and action so as to shut out the possibility of our engineering any more troublesome examples in which alien intermediaries, like falls down the stairs, find their way into the nexus. I was able to formulate the mediated conception so that its *aligns with* relation, the one that obtains between a rational agent's judgments and actions, is modified with 'non-accidentally' because the opacity of practical thought excludes any chance alignment. This is what guarantees that our *relata* our bound together in the right way. (In the opposite case, in which I judge that I ought to φ, and then, unrelatedly, I'm motivated to φ, our assessment should be that the practicality of the judgment isn't manifested in the motivation. In such cases, there remains the possibility that, even if in fact I φ, I don't exhibit any rationality in the action since I don't do it *from* my judgment.)

Of course, our practical judgments may, and often do, misconstrue their objects. But, as we earlier observed, assessments of agency are two-tiered. Even when I'm substantively in error about my reasons, I may succeed by a second measure, one that judges on the basis of the relations between attitudes. In this way, it can happen that I'm entirely rational though my actions don't conform to the reasons that bear on my conduct.[54]

1.3.4 The reason-relation and *acting from*

With this second objection lodged against the direct conception of *rationality*, we can return to the extensional reading of the reason-relation. Perhaps what's been said thus far is already enough to defeat it. After all, we should expect the correct conception of *rationality* to figure in the correct

---

[54] That said, I can't be admitted to the domain of rationality if I'm profoundly confused about normative matters. In fact, no one capable of normative confusion is capable of profound normative confusion. This point extends one I made earlier to the effect that animals that are 'rational' in the classificatory sense can't be perfectly irrational, since their errors are only intelligible against a backdrop of coherence. The same applies here—a mistaken normative judgment is only intelligible as such from within a web of judgments that's largely on the right track about what the normative domain is like. It follows from this that there's an upper limit to practical confusion.

conception of *the reason-relation*, so that, if the former has intensional elements, the latter must as well. And we've just seen that the former has intensional elements. Still, we might not yet see how rationality figures in the reason-relation. With this in mind, in this subsection I reconstruct the reason-relation from the ground up. This reveals that, as before, we mustn't simply gather the appropriate *relata*—the fact, the agent, the action, the circumstances—together. They must be unified according to the correct principle. And, just like the idea of *rationality*, there's no specifying this principle without appeal to the idea of *practical thought* as exemplified in judgment.

Let's begin with this axiatic entity: the goodness of the cat's comfort. Axiatic entities can't prescribe directly; they rely on deontic counterparts for that. So, corresponding to the goodness of the cat's comfort, there is, at minimum, a deontic reflection of that goodness, namely, its translation from axiatic positive value to, e.g., a deontic ought-to-be.[55] If the cat's comfort is just now realized, its ought-to-be is more concretely represented as an ought-to-be-preserved, as given, say, by a reason to leave her undisturbed. But stipulate that the cat's comfort is at present a non-actual possibility. In this case, the deontic counterpart to its goodness is an ought-to-be-realized.

Strictly speaking, the goodness of the cat's comfort only has a deontic counterpart if there's an agent who can be motivated by it. But for any given axiatic entity, we must be able to discern what form its counterpart(s) might take were the relevant agent to exist. The device laid out in subsection 1.2.3 can be applied here as well: we can read the deontic counterpart off of an axiatic entity by considering the possibility of agency that's motivated by its value. So, if as a matter of fact, there's no agent who can make the cat comfortable, we need only envision an agent who can, and

---

[55] Probably, it also translates into a deontic ought-to-be-contemplated, that is, a consideration prescribing that we admire the cat. Contemplation involves thought, so the extensional reading can't be applied there. In order to marshal the strongest defense that can be assembled for the extensional reading, I investigate only the one counterpart, the productive ought-to-be. Since it has a causal basis, it should be the most amenable to an extensional interpretation. Still, we should note that this isn't the only form a deontic counterpart might take (far from it). Moreover, my frequent appeals to deontic counterparts shouldn't be taken to signal any sympathy for consequentialism.

then consider what deontic entity would arise for that agent. In our example, we'll suppose that I'm nearby and capable. But my presence isn't in the end necessary, nor is the presence of any other agent.

The hyphenated terms—'ought-to-be-preserved' and 'ought-to-be-realized'—give the impression that they refer to monadic properties. This is misleading, since oughts invariably figure in 'ought to φ' expressions. This attests to their being, at least, dyadic properties. The same is true of normative reasons—each is a reason to φ. This is what gives us i), the relationality of normative reasons. So, in order to elaborate the deontic counterpart to the goodness of the cat's comfort, we must specify its relational features, that is, we must catalogue its *relata*, and then supply the principle that combines them. One *relatum* is the fact of the possibility of the cat's comfort. This is the *relatum* assigned to the fact-place in Scanlon's relation. It's also the *relatum* that *is* the reason, as identified by claims like "the possibility of the cat's comfort is a reason to φ."[56] What remains (for the moment, anyway) is the *relatum* at the opposite end of the deontic relation, namely, φ. We must determine what it is.

The relationality of the counterpart amounts to more than an echo of the axiatic entity to which it corresponds. It introduces some novel *relatum*, something not to be found in that axiatic entity. It isn't the case that φ is only the realization of the cat's comfort, as in "the possibility of the cat's comfort is a reason for the realization of that comfort." If it were, all it would do is compile an entity out of the noun phrase 'the cat's comfort' and some modal operators, which is too near the

---

[56] Scanlon's reason-relation assigns a place to the reason itself—the fact-place is the reason-place (*Being Realistic about Reasons*, page 30). At times, I've neglected this aspect of his conception, and spoken of 'reasons' in reference to their relations. This is for the most part innocuous. Granted, reasons *qua* facts needn't be relational, but they inherit their normative authority from their positions in reason-relations, and it's in virtue of that authority that they're deontic. My claim that each deontic entity is relational should be interpreted as locating these entities at the sources of authority, rather than at the fact-places in counts-in-favour-of relations. So, for instance, the deontic counterpart to the goodness of the cat's comfort isn't the reason that is the possibility of that comfort, but the reason-relation in which that reason appears.

misleading image in which ought-to-be-realized is given the look of a monadic property. We must have more space between the *relata* if we're to provide the material of a deontic relation.

Why must we have this space? Deontic entities have a practicality that axiatic entities lack. This is reflected in the fact that the two varieties of normative entity figure differently in practical deliberation. In the normative domain, instrumental relations transmit value from possible, non-actual goods to the indispensable means to the realization of those goods in much the same way those relations transmit deontic force. But axiatic entities can't by themselves guide agency along these instrumental relations. That is, while there's an axiatic path running between a valuable end and its (instrumentally valuable) means, no practical deliberation can follow it, not in such a way as to channel being motivated by the value of the end into being motivated by the value of the means. If an agent is to move from end to means in practical deliberation, she must recognize a deontic entity binding the two together in the form of a hypothetical imperative.[57] Without deontic intervention, the deliberator can only be motivated by the one independently of her being motivated by the other.

Axiatic entities certainly can be objects of contemplation. But in the absence of suitably related deontic entities, axiatic deliberations aren't practical. I can encounter the mat *qua* valuable for its utility in comforting cats, but this isn't to conceive of it *qua* the thing by which to comfort *this* cat. I might connect the two values, but this would require a stroke of ingenuity, the product of an imaginative rather than a practical capacity. Likewise, familiarity with the instrumental infrastructure of the normative domain may enable me to map the value of the cat's comfort to the value of the mat as a means to that comfort, but this by itself is practically idle; if I come thereby to be motivated by the latter value, it'd be from the independent recognition of its own normativity. That is, I wouldn't be motivated by it *in virtue of* having been motivated by the cat's comfort. Practical

---

[57] See Kant's *Groundwork*, page 414.

deliberation can follow the path from a possible, non-actual good to the goodness of its

actualization, but this is only an analytic connection. It can't guide action.[58]

So, let's construe φ as a potential cause of the cat's comfort.[59] This establishes what will be

the action-place of the relation once an agent is brought into the picture. But for the time being, it's

only a cause-place. Suppose that placing the cat on the mat would have this causal power. Thus, we

arrive at a workable formulation of the counterpart: *the possibility of the cat's comfort is a reason to place her*

on the mat. This supplies the requisite deliberative heft—a novel *relatum*, distinct from the goodness

of the cat's comfort.

But this still isn't enough to give us practicality. We face the same difficulty that defeated the

direct conception of *rationality*; as things stand, any cause with the relevant power is subject to the

deontic counterpart we've constructed. For instance, the movement of a ray of sunlight could

manoeuvre the cat on to the mat. This makes it a candidate φ. But there are no oughts for rays of

sunlight. We've come into conflict with this principle: whatever occupies the φ-place of the reason-

relation must be intelligible as answering to the prescription expressed by the relation. In order to

respect this principle, we must delineate φ so as to include only agential causality. We might elect to

do as the direct conception did, namely, accommodate our intuitions via brute stipulation, as in 'let φ

be an agential cause'. This would give us ii), agent-relationality, which excludes the movements of

rays of sunlight on the grounds that rays aren't agents. But we've seen that this alone won't work. It

---

[58] I've assumed that instrumental relations aren't constitutive relations. They're causal, from which it follows that, among other things, their *relata* are ontologically distinct. This is what gives rise to the problem of securing the unity of their *relata*. In constitutive relations, one *relatum* inhabits the other, so that the two aren't ontologically distinct, and the unity of the *relata* doesn't raise any special problem. But, as Williams observes, constitutive reasoning can be practical ("Internal and External Reasons", page 104). In cases in which it is, a deontic entity must be present somewhere, since otherwise the reasoning would only be imaginative, and not practical. Still, the deontic component needn't come into view in the course of deliberation (though it likely will, as when, for instance, we must commit to one from a variety of options, this mat rather than that one, this flight of stairs, etc.). (Korsgaard characterizes Williams' constitutive reasoning as a variety of instrumental reasoning. Our disagreement is, I think, unremarkable, it being nothing more than a matter of misaligned stipulations; see note 194.)

[59] Again, the example develops a *productive* deontic counterpart to an axiatic entity. Not all deontic counterparts are productive, and not all involve causal relations.

isn't enough that a given cause is agential. It can only be admitted to a reason-relation on the condition that it, at the moment of admission, manifests its status as agential. My inadvertently falling down the stairs can be described—say, as an event triggered by something that (as it happens) is an agent—so that it appears to be answerable to prescription. But, of course, nothing inadvertent can occupy the φ-place. So, the boundaries we draw for φ must be narrower than the set of causes involving agents.

And again, there's no purely extensional way to narrow the set of causes to the ones that fit. We can stipulate a bit more, and limit ourselves to actions, but this doesn't address the problem, since some of the actions that produce the cat's comfort only do so incidentally (a sadistic cat might be comforted by my hurling myself down the stairs, for instance). If the deontic counterpart is to find application, we must, as before, step inside the agent's perspective, and find our φ there. But that's just to say that φ must be specified in intensional, rather than extensional terms. Put another way, our choice of φ must accommodate the opacity of practicality. To do that, it needs its own corresponding opacity, which it finds in the possibility of the practical thought of the agent called upon to perform it. This introduces the agent-place to the relation. The result is causality deployed with a view to doing as prescribed, that is, causality that represents its effect under the description of its being prescribed. Or more briefly, the result is action *from* the reason. And with this invocation of practical thought, we have iii).

1.3.5 The status of the argument

That completes my effort to defend iii), and with it, the thin formulation of the immanence constraint. To recapitulate, though the extensionalist can grant ii), namely, that the relation includes an agent-place as an essential constituent, she can't do any more than that. The *conformity to* relation can't furnish the materials for an account of classificatory rationality. It's for lack of those same

materials that it can't account for why reasons are agent-relational. Thus, the extensional reading must leave ii) unexplained. I said before that ii) demands explanation—if this is right, then we must abandon the extensional line, and embrace its intensional opposite, the one that iii) articulates in terms of practical thought. That iii) commands the requisite explanatory power is clear—because the relation involves practical thought, it must include the things capable of practical thought, namely, agents. It's to be expected, then, that they have a place reserved for them, and so we have our explanation for ii).

As I acknowledged at the beginning of this section, my argument relies on intuitions as to the scope of deontic entities like reasons (just as my argument against the direct conception relies on intuitions about where attributions of rationality and irrationality are appropriate). In principle, every intuition can be revised. But these are intuitions that figure at the platitudinous centre of the idea of *normative reason*, and as such, their revision is liable to eliminate that idea. Assuming that we're reluctant to embrace eliminativism, we must take care that we leave these intuitions as they are. That they are core—that is, so central as to threaten elimination if revised—is, I think, plain. At any rate, my case is helped by the fact that they're corroborated by other intuitions that also have the look of platitudes: *prima facie*, rays of sunlight and inadvertent falls down the stairs can't enter reason-relations; also *prima facie,* they can't be guided, whether in deliberation or otherwise, nor can they be justified or explained via appeal to normative reasons. These intuitions form a mutually vindicating network. And if we make adjustments in every place, we can't expect in the end to be left with something still recognizable as the idea of *normative reason*.

The principle under which these intuitions fall can't be articulated without appeal to the idea of *practical thought*. In fact, the basis for the intuition that there are no reasons addressed to rays of sunlight or to inadvertent falls is just this: reasons are only addressed to things that can think them. Given that I've helped myself all the while to this principle (for which I offer no further

justification) it's hardly surprising that I've arrive at iii), since that only states the same thing from the other side, namely, that agents enter reason-relations by way of their capacity for practical thought. As I've said this justificatory structure is circular.

Is this disappointing? Whether we think so determines to what extent we can be satisfied by constructive primitivism.[60] Such a conception can't step outside the idea of *normative reason* without betraying its primitivism. Still, it hopes to justify via systemic connection, i.e., by way of the harmonious arrangement of related ideas in a single conception. Only those who already appreciate the distinguishing features of the whole array of ideas—*normative reason*, *rational animal*, *rationality*, *action*, etc.— can feel the force of this style of justification. But this doesn't challenge its status as justification.[61]

At any rate, I take myself to have established that the reason-relation is characterized by each of i), ii) and iii). And with iii) in place, the thin formulation of the immanence constraint is secure. If the scope-extending arguments in 1.2.2 succeed as well, then the thin-wide formulation is also secure, so that we can say of any successful normative conception that it includes an account of the possibility of practical thought.

---

[60] To say of a conception that it's construct*ive* isn't to characterize it as a construct*ivism*. In the context of metanormative thought, 'constructivism' refers to the view espoused in different forms by Rawls, Korsgaard, Scanlon and Street; a conception is constructive (regardless of domain) if it seeks to explain some feature of its subject matter. As it happens, all constructivist conceptions are constructive, though not all constructive conceptions are constructivist.

[61] I say more about constructive primitivism in chapter 4. Think of it this way: even those who know where everything is can benefit from a map. We must take care that we don't neglect the value of seeing everything together in one place. And anyway, in practice our knowledge of the relationships between ideas is always only partial. Maybe this is the image I want: I know my neighbourhood well enough, and I know my friend's neighbourhood too; the two neighbourhoods are adjacent to one another, but I rely on a subway to get between them. Upon consulting a map, I'm surprised to discover that a street running through both neighbourhoods is curved (indeed, I hadn't realized that I'd assume it was straight until I was corrected).

1.4 Tidying up

Before turning to the thick formulation of the immanence constraint, I'll say a bit about how iii) and

rationality are related. The mediated conception shows rationality to be partially constitutive of the

agential orientation to reasons, which, given iii), makes it partially constitutive of the reason-relation

as well. We can see this in the fact that I can't succeed in acting from a reason without also being

(to this degree) rational in so acting. This is because the reason from which I act must also figure as

content for the judgment that decides that action. Meanwhile, the judgment that I ought to act on a

given reason rationally entails my assigning myself to the agent-place of the relevant reason-relation.

This is the same constitutive relation running the other way, from reason-relation to judgment. That

these two ideas—*practical judgment* and *reason-relation*—are intermeshed in this way is a manifestation

of the intermeshing of the ideas of *rational animal* and *normative reason*. In fact, it's here that we can

see this intermeshing most clearly.

The rational requirements aren't themselves normative.[62] They may seem at times to

command deontic force, but this is an illusion. Being partly constitutive of what it is to occupy the

agent-place in a deontic relation, they can't generate their own deontic relations without triggering a

regress. Take the requirement articulated by the mediated conception: if I judge that the cat's

comfort gives me decisive reason to place her on the mat, rationality requires that I place her on the

mat. This requirement mustn't be received as a new reason, since if it was, it would provide the

material for a new judgment that I ought to do as *it* prescribes, which would, in turn, produce some

rational pressure of its own, and on and on. Evaluative rationality is a virtue, of course, but its

---

[62] Korsgaard often uses the term 'normative' to characterize the principles of practical reason (see, e.g., *Self-Constitution*, page 59). I think she's better interpreted as representing them as proto-normative, i.e., as delivering the agent to a point from which she can act from normative entities. This is a difficult exegetical issue. I can't tackle it here, but see section 3.3.

deontic counterpart takes the shape of a reason to cultivate the disposition to conform to the rational requirements. This is as close to the requirements as a deontic entity can get.

Relatedly, there are cases in which one reason directs me to act from another reason, e.g., a case in which I ought not only to place the cat on the mat, but to do so specifically because it will make her comfortable. But wherever there's a chain of higher-order reasons prescribing action from lower-order reasons, that chain must terminate at the top in a reason that's satisfied with mere conformity. It never happens that what's prescribed is practical thought all the way down.[63] This is why I've focused all this while on the agent-place of the reason-relation, and not the action-place. Practical thought figures in the constitution of the agent-place only. The action-place doesn't require *action from* (anyway, not action from the reason in which it appears). In fact, it must, at some level, be characterized in terms of *conformity to*. This is true of all deontic entities. They necessitate the agents who are to perform the relevant actions, but not the actions themselves. It's agency that introduces the intensional element.[64]

---

[63] This is why Kant doesn't *prescribe* that we exhibit the good will. Exhibiting a good will requires *action from* the moral law, but nothing can determine us to act from that law, since that would begin a regress of *action from*. The moral law only prescribes actions as extensionally conceived, so that, from its point of view, *conformity to* is sufficient. The axiatic status of the good will is plain to see, but its normative force can only be felt if it's redescribed in terms of duty. This is reflected in the fact that the agent who exhibits a good will doesn't attend to that goodness, but has her moral sights on her duty. Were she to pursue the good will described in terms of its goodness, she could never instantiate it (she'd only be a kind of moral narcissist).

  I believe quite a lot of misunderstanding is owed to neglect of this fact. Kantian deontology begins from an axiatic entity, just like any other moral theory. The difference is that Kant's favourite virtue has this idiosyncratic feature, that we acquire it by responding to deontic, rather than axiatic, normativity. This is a departure from much moral and ethical theory; it's what prevents Kant's theory from being teleological, in the Rawlsian sense of that term (*A Theory of Justice*, pages 21-22). Still, it's not so radical a departure as many take it to be.

[64] Much of my thinking on this issue derives from an argument of Prichard's: "[t]o feel that I ought to pay my bills is to be *moved towards* paying them. But what I can be moved towards must always be an action and not an action in which I am moved in a particular way, i.e., an action from a particular motive; otherwise I should be moved towards being moved, which is impossible" ("Does Moral Philosophy Rest on a Mistake?", page 12). As is characteristic for Prichard, he overstates his point, since there's a circumscribed space within which we can "will to will". Still, he's right that the chain of willings must eventually terminate in an action that isn't willed under any particular aspect.

Earlier, in subsection 1.3.2, I listed three objections to the direct conception of *rationality*. I've already expanded on the first and last, but I haven't said much about the second. That's the one that dwells on the fact that the direct conception can approve of my frustrating my own normative judgments if they misrepresent their objects. This is, the objection suggests, embarrassing. For the most part, it is, I think, rather embarrassing, but there are cases in which it may appear to capture our intuitions better than the mediated conception. Suppose that I'm disposed to misjudge in moral contexts. Recognizing this, I elect to defer to the judgments of moral experts instead, even if their judgments diverge—really, especially if they diverge—from my own.[65] If my standing inclination to act directly from my judgments persists, I might have to design a kind of regimen of rehabituation for myself. Probably, in order for this regimen to be effective, it must involve the active frustration of my will as instantiated by the offending judgments.

Cases like this one aren't unusual. They're largely innocent, or anyway, not-irrational. But they pose no immediate danger to the mediated conception, since even as I undergo my self-frustration programme, I act from a judgment, namely, the one prescribing that I defer to experts. The mediated conception only requires that my thinking and doing aligns with my judgments. It has nothing to say about how to resolve contradictions between judgments. As rational requirements go, the mediated conception is remarkably easy to meet. If I judge that I ought to φ, but also that I ought not to φ, then, as far as the mediated conception is concerned, I'm rational no matter what I do. The same is true if I make no judgment at all.

Parfit describes a similar case: I'm cornered by a thief. He knows I've hidden something valuable nearby, and provides an incentive to reveal it by threatening to kill me if I refuse. I see that the threat is genuine, but also that it depends on my capacity to respond intelligently to the incentive. If I were to faint, say, the thief would have no way to proceed. Unable to faint at will, I

---

[65] This example comes from Myers and Claudine Verheggen's *Donald Davidson's Triangulation Argument*, at page 149.

avail myself instead of a potion that causes a temporary lapse in my capacity to respond intelligently to reasons, in this way realizing the best state of affairs, in which I retain my life and the thief gets nothing.[66] Parfit claims that "[o]n any plausible theory about rationality, it would be rational for me, in this case, to cause myself to become for a period irrational."[67] But that way of framing the case gives rise to a puzzle since it suggests that I'm, for a time, simultaneously rational and irrational.[68] It's cleaner to say that there are case in which it's rational for me to suspend my agency for a time. I'm arational (not irrational) during this time, but we can find in that arationality a record of my earlier rationality, namely, the rationality exhibited in drinking the potion. Here again, I see nothing that causes trouble for the mediated conception.

But then, there's nothing in either case that warrants the name 'self-sabotage'. Nagel distinguishes genuine self-sabotage from the ordinary policing of our own desires. The former, he says, "must not be confused with the perfectly unobjectionable and not uncommon case in which someone puts obstacles in his way knowing that he will *want* something in the future which he should not have. This may induce him to put a time lock on the liquor cabinet, for example."[69] My two examples are like the time lock in that they involve a rational response to some feature of the agent or her circumstances. Really, from the point of view of evaluative rationality, there's nothing remarkable in Parfit's case at all, odd though it is. Every night I decide on a lengthy period of arationality, namely, sleep, with a view to securing, among other things, a better rested, and to that extent, more alert, disposition. The case in which I opt to defer to moral experts is closer to self-sabotage, since it involves antagonism between aspects of the will, but the conflict is too superficial to pose a significant problem for my integrity as an agent. Or anyway, it's available to us to view my

---

[66] *Reasons and Persons*, pages 12-13. The example is derived from Thomas Schelling's *The Strategy of Conflict*.
[67] *Reasons and Persons*, pages 13.
[68] I've assumed all along, and Parfit would agree, that 'rational' and 'irrational' are contraries. Parfit draws on his distinction between formal and substantive aims to solve this puzzle (*ibid.*, page 9), but I can't see how this is any help.
[69] *The Possibility of Altruism*, page 40.

moral behaviour in that case as springing from my decision to follow the advice of others, so that my agential unity is for the most part intact. What we need is a deeper rupture in my agency, something that suggests the possibility of full-blown agential fragmentation.

In the example from subsection 1.3.2, I judge that I ought to praise the cat, though in fact I have decisive reason to leave her undisturbed. Suppose I prevent myself from praising her. If all that prevents me is the thought that I might be mistaken, there's no irrationality in it. But if what prevents me is a standing disposition to assign no motivational force to my judgments, a charge of irrationality is, I think, justified. And because my irrationality (if that's what it is) derives from a mismatch in my contentful attitudes, no extensional conception of *rationality* has the means to recognize it. That's more trouble for the direct conception.

(That said, it may be that I've misdiagnosed the problem. Michael Stocker takes much contemporary moral theory to prescribe something like a bisecting of the will, whereby the faculty of moral judgment is detached from the faculty of moral motivation. This is quite a lot like what I have in mind, but he doesn't object to it on the grounds that it's irrational. His complaint is that it estranges us from the good, resulting in a bad ethics.[70] Meanwhile, Parfit is prepared to allow that rationality *requires* this same sort of thing.[71] I investigate a related idea at length in subsection 3.4.1.)

Practical thought requires the representing of a normative entity, and this representation must in some sense insert the practical thinker into the agent-place of a deontic relation. This is, I believe, just what it is for a representation to have practical content. I won't advance any substantive view as to the conceptual requisites of practical content. My suspicion is that no normative vocabulary is necessary. I needn't think 'that the cat is on the mat is *reason* to leave her alone', or 'it's *good* to leave

---

[70] See Stocker's "The Schizophrenia of Modern Ethical Theories".
[71] *Reasons and Persons*, page 24.

the cat alone while she's on the mat'. Probably, all that I need is 'the cat is on the mat'—given a

practical valence, that thought can, by itself, place me in the agent-place of the relevant deontic

relation. But, as to what distinguishes this practical 'the cat is on the mat' from the purely

speculative 'the cat is on the mat', I won't at present venture any guess.[72]

Finally, I'll make some stipulations explicit: in speaking of thought, I mean *rational thought*,

and in speaking of agency, I mean *rational agency*. These are only abbreviations; I don't mean to

foreclose the possibilities of non-rational thought and non-rational agency.[73] That said, if there are

two species of agency, rational and non-rational, they must be radically unlike one another.[74] As

we've seen, it's only rational agents that can be admitted to the agent-place of the reason-relation.

This is because the relation depends on the possibility of agents attending to their reasons under the

aspect of their normative authority, i.e., attending to them in practical thought.[75] This involves a

kind of normative intelligence, the kind that constitutes the capacity for practical thought, and that

doesn't appear in the extensional alternative (the alternative likely involves *some* kind of intelligence,

but not a specifically normative intelligence).

What about the connection between 'rational animal' and 'agent'? I assume that all rational

animals are agents. There are non-animal agents—I'm thinking of institutions, legal persons, and

the like—but I suspect they're all, at least, composed of animals. At any rate, I use the two terms

---

[72] Though I will say that I subscribe to the Davidsonian view developed by Myers and Verheggen in *Donald Davidson's Triangulation Argument*, according to which practical thought originates in causal relations with real normative entities, like normative reasons (I return to this view in section 2.3). This makes the relation between practical thought and reasons rather tight. It also shows the concept of *normative reason* to be easier to acquire than we might otherwise have supposed. And since it establishes a genetic connection, it can explain how this concept might figure in all practical thought, even if it doesn't appear in many of our occurrent attitudes. Represented that way, the intellectual view is more palatable, even appealing.

[73] Like the kind Korsgaard describes at pages 20-23 of *Fellow Creatures*.

[74] As Korsgaard elaborates at, e.g., pages 90-93 of *Self-Constitution*.

[75] Ultimately, the distinction is between linguistic and non-linguistic agency, that is, between agency that can discern semantic aspects, and agency that can't. But given a Davidsonian outlook on these issues, these come to the same thing. See Davidson's "The Emergence of Thought". (It's worth mentioning, though, that Davidson's conclusions as to the discontinuity between rational and non-rational animals in "Rational Animals" are clearly too strong. Language opens an enormous divide between the two, but not so much that we must hold that dogs and cats can't think at all.)

interchangeably. A last point: as I said in note 8, my focus is the immanence constraint as it emanates from distinctively practical rationality. This is the shape it takes as it conditions conceptions of *normative reasons for action*. At its most general, the immanence constraint conditions conceptions of *normative reasons for attitudes* as well, but a truly comprehensive account, one that is sensitive to each of the unique considerations that arise from the different attitudes, as well as those that arise from action, is more than I can begin to envision, let alone develop. So, my phrase 'immanence constraint' refers to the constraint in its practical aspect only.

2. The immanence constraint, substantively construed

I've arranged my formulations of the immanence constraint along two dimensions, the narrow-wide dimension, and the thin-thick dimension. As stipulated, a formulation is narrow if its scope extends as far as all conceptions of *normative reason*, and is wide if it extends to the whole of the normative domain. And, as stipulated, a formulation is thin if it springs directly from the relational nature of deontic entities, and is thick if its derivation from that nature is indirect. Or, put a different way, it's thin if it's a formal feature of the idea of *deontic entity*, and thick if it's a substantive feature. The price of substance is that it requires that we venture outside the schematic relationality of deontic normativity to comparatively robust claims as to how deontic relations actually obtain.

Here's the thin-wide formulation again: *if a normative conception is to succeed in capturing its corresponding idea, it must establish the possibility of practical thought about the entities that instantiate it*. Its thin component is given by the constitutive role of practical thought in agent-relationality, which is itself an essential constituent of deontic relations. Its wide component is given by the ubiquity of these relations throughout the normative domain. The defence I assembled for the former component strikes me as rather strong. The latter component is perhaps not so secure. Given that this is the case, I reserve the right to retreat to the thin-narrow formulation, the one confined to conceptions of *normative reason*, should the need arise. In any case, both the thin-wide and thin-narrow formulations have this vice: they're thin, and as such, they have little in the way of bite, little with which to slap the wrist of the metanormative theorist should she develop a conception that neglects the plight of the rational animal.

There's an ambiguity in the phrase 'establish the possibility of practical thought about α'. In order to establish a possibility, must we explain it, or is it enough to simply assert that it's there? The best that the thin formulation can sustain is the weaker of the two interpretations. After all, it only establishes that practical thought is indispensable to deontic relations, and this doesn't ground

any obligation on theory to account for what sort of thing practical thought is, or how we engage in it (to the extent that we do). Thus, the thin formulation can't by itself demand any explanation for the possibility of practical thought. It must be satisfied with the assertion 'it's there'.

And assertion is easy. What is there in the thin formulation to prevent a philosopher who espouses a conception of some practical entity α from appealing to the intelligibility of our thinking about α, i.e., to its position in logical space, so as to deliver the possibility of such thought, as it were, for free? As far as I can tell, nothing. Of course, we're dealing in distinctively *practical* thought, so the relevant possibility must leave room for motivation. But because the immanence constraint doesn't include any substantive commitment as to how motivation issues from thought, it can raise no objection against our positing that the successful representation of a normative entity has motivational force (perhaps on the back of the observation that there's no obvious contradiction in the idea). In fact, so far we've encountered only one aspect of the possibility of practical thought that does anything to delineate the shape that the thin formulation's desideratum must take, namely that the idea of *practical thought* involves *acting from* relations as encapsulated in deontic necessitation. As a result, metanormative conceptions must show practical thought to be opaque. But this is easily done, since thought *simpliciter*, at least as traditionally understood, is by its nature opaque. We can, if we like, *conceive* of theories—very eccentric ones—that could be defeated by the immanence constraint on its thin formulation, but we're unlikely to encounter any in practice.

Philosophers at time develop conceptions of α that can't locate the idea of *thought about a* in logical space. But if they do, they've made an obvious mistake. Or anyway, because it belongs to the relevant idea of *conception* that it amounts to propositional thought, any conception of α that represents α as unthinkable is immediately self-defeating. If there are conceptions of this kind, they

face far larger problems than their having violated the immanence constraint.[76] The fact is that the thin formulation doesn't by itself exclude any of the metanormative conceptions that we might otherwise be tempted to adopt, and so it has little to contribute to the course of metanormative inquiry—or anyway, little to contribute directly.

But all this is by design. From the beginning (section 1.2), the task assigned to the thin formulation wasn't to defeat any of the conceptions in circulation in contemporary metanormative theory. Rather, it was to make inevitable—or anyway, irresistible—a further formulation, one that traces the concrete features of a successful account of the possibility of practical thought, and that thereby rules out the conceptions that fail to capture those features. Thus, it falls to the thick formulation to assign some more demanding criterion to the immanence constraint, something that *can* threaten a slap to the wrist, and in this way, articulate the rational animal's complaint.

The shift from thin to thick formulation is the shift from the weaker to the stronger interpretation of 'establish the possibility', that is, from the one that's satisfied with the recognition that *practical thought* is an intelligible idea, to the one that requires, in addition to this, an explanation of how an agent might engage in such thought, how she can find her way into the agent-place of a normative relation. Construed in this way, the thick formulation is just a demand for explanation. But we must step carefully—the rational animal's complaint *can*, I think, be understood in terms of a demand for a species of explanation, but only provided that we're mindful of the fact that certain explanations fall well clear of the mark, both in terms of fidelity to the *sui generis*-ity of the normative domain, and in terms of fidelity to the rational animal's own perspective. The unqualified use of the term 'explanation' threatens to activate scientistic prejudices according to which whatever explains is

---

[76] In a manner of speaking, the early Wittgenstein characterizes ethics as ineffable, and so, as not propositionally thinkable: "[i]t is clear that ethics cannot be put into words" (*Tractatus Logico-Philosophicus*, 6.421). But he readily accepts what follows from this claim, namely, that all ethical theory is gibberish. By renouncing the idea of *normative conception*, he renounces the project of developing any such conception, so that nothing commits him to meeting the thin formulation. (I have more to say about the early Wittgenstein in subsection 2.1.3.)

locatable in the purportedly exhaustive causal nexus that characterizes the physical domain.[77] Some

deference to the physical is warranted, but we needn't go so far as to require that all *explanantia* and

explananda be readily specifiable in physical terms. Rather, we must hear the word 'explanation' as it

figures in the thick formulation as offered according to some more flexible sense, one that doesn't

trespass on the power of normative entities to transcend agential doings, or on the agent's own

expectation that her doings will from time to time be transcended.[78]

The relevant variety of explanation won't come into view until I've established the possibility

of constructive primitivism, and with it the possibility of a metanormative conception that satisfies

both the immanence and transcendence constraints. Since I don't accomplish this until chapter 4,

we have to wait some time before we can see precisely what the thick formulation demands. For

now, the best that can be managed is to demonstrate that *some* demand for explanation can be teased

from out of the thin formulation of the immanence constraint, leaving it until later to determine

exactly what would be required to satisfy it.

The thick formulation is capable of both a negative and a positive portrayal: the negative

face prohibits austere primitivism on the grounds that, given austerity, we can only posit the

possibility of practical thought as a brute, and hence inexplicable, feature of the world; the positive

face suggests three avenues for the development of an account of the possibility of practical

thought: a theory of normative content, first-order normative theory, and a theory of agency. I

---

[77] It may also cause us to insert normative entities into domains in which they can only be explanatorily impotent. Interpreted this way, the thick formulation amounts to something like Harman's explanatory criterion (see his *The Nature of Morality*), together with its nihilistic upshot. But the thick formulation is designed to be met. Its explanatory criterion is comparatively loose, or at any rate, attuned to the peculiarities of the normative domain. (Incidentally, while I sympathize with Scanlon's response to Harman at pages 26-27 of *Being Realistic about Reasons,* I agree with Myers that even Scanlon concedes too much to the scientistic conception of *explanation*; see *Donald Davidson's Triangulation Argument*, page 174).

[78] That said, normative explanation must defer to physical explanation so as to avoid parallelism (according to which there's no prospect of contact between the two explanatory systems) and pluralism with respect to the idea of *explanation*, both highly undesirable outcomes. I interpret Davidson's anomalous monism as licensing a kind of minimal deference, whereby normative explanations are undergirded by physical explanations, but not in such a way as to delimit the one by the other.

elaborate the negative face in subsections 2.1 through 2.3, and say a bit about the positive face in subsection 2.4. These subsections map the trajectory from the thin formulation to the negative face, and then from there to the positive face. All of this is with a view to giving determinate expression to the rational animal's complaint. I don't aspire to articulate the whole of that complaint, though the immanence constraint succeeds, I think, in capturing a central portion of it.

2.1 AUSTERE PRIMITIVISM AND PRIMITIVISM *SIMPLICITER*

The negative face of the thick formulation constrains normative primitivism by requiring that it be non-austere. In order to appreciate the reach of this constraint, we need a preliminary characterization of both austerity and primitivism. So, I start there. The argument for the negative face begins in section 2.2, where, as a first step, I repeat the demand for an explanation of the possibility of practical thought, alongside the fact that, by definition, austerity can't meet it. Even when stated as plainly as that, I believe the argument has some force; my suspicion—or anyway, my hope—is that, once the explanatory poverty of austerity is brought under the spotlight of the immanence constraint, some austere primitivists will be persuaded to opt instead for a constructive alternative (provided, of course, that one is available; happily, one is, as we'll see in chapter 4). But, recognizing that there are others who aren't embarrassed by the failure to explain, in section 2.3, I defend a substantive criterion for the possibility of practical thought: the discursivity of its objects, i.e., their standing as potential subject matter for fruitful discussion between agents. Since austerity can't supply the materials for fruitfulness in discussion, austere primitivism can't deliver normative discursivity, and thus, can't deliver practical thought. This is how the thin formulation generates the thick formulation: satisfaction of the thin formulation entails satisfaction of the discursivity criterion, which, in turn, entails satisfaction of the thick formulation.

2.1.1 Innocent ontological questions and answers

The thick formulation doesn't prohibit primitivism, not just as such. Primitivism is, in a manner of speaking, the refusal to countenance the demand for a certain kind of explanation. Schematically, a conception of α is primitivist if it refuses to allow that philosophy is accountable for an answer to a 'why α?' question. Or rather, it can acknowledge that question, but only on the condition that its asker can accept a circular answer. Thus, the question that primitivism refuses to answer is better articulated as 'what non-circular explanation can be offered for α?'. It often happens that 'why α?' is asked in precisely that way, that is, with an implied non-circularity provision, or at any rate that the expectation it expresses is that there's some *explanans* for α that's wholly (though perhaps only conceptually) antecedent to α. In the context of that expectation, 'why α?' is, by the primitivist's lights, a leading question; she thinks that we ought first to ask whether there *is* any *explanans* that's antecedent to α in the relevant sense, and that the answer to *this* question is 'no'.

As with any other philosopher's ism, the term 'primitivism', though likely the best of the available options, is in certain respects regrettable. On the one hand, it elicits an image that's suggestive of the explanatory boundaries the primitivist means to draw; in calling a given thing 'primitive', we represent it as basic, as something with respect to which there's nothing prior. But on the other hand, the term is liable to mislead by implying ontological commitments that only some primitivists would accept. To say of α that it's primitive isn't necessarily to locate it among the ultimate constituents of the world, or even to grant that there's a notion of *ultimacy* that has any determinate sense. Perhaps only the posits of quantum physics aspire to that degree of ontological primacy. In any case, most extant primitivisms recognize that their primitives can, and in fact do, supervene. They may, at times, dress their primitives in language—'simple', 'indivisible', 'fundamental', 'foundational', etc.—that implies a lofty ontological ambition, but for the most part

that's unintended (or anyway, not in any unmitigated or uncircumscribed way). For instance,

Scanlon says:

> I will maintain that truths about reasons are fundamental in the sense that truths about reasons are not reducible to or identifiable with non-normative truths, such as truths about the natural world of physical objects, causes and effects, nor can they be explained in terms of notions of rationality or rational agency that are not themselves claims about reasons.[79]

Clearly, his use of the term 'fundamental' doesn't mark off anything like an ontologically privileged status for normative reasons (whatever that might mean). It only describes the irreducibility of the idea of *normative reason*. To deny that reduction of a given idea is possible is not to elevate that idea above any others, except in the sense that it blocks one path by which another idea might be elevated with respect to it. (The repudiation of scientism isn't a positive ontological doctrine.)[80]

Moreover, the primitivist's refusal to answer the 'why α?' question can be remarkably narrow, especially if she holds that circles can command explanatory power. Perhaps it only shuts down very idiosyncratic and specialized demands. After all, the primitivist about *justice*, say, can hear the question 'why *justice*?' with a certain intonation, one inviting a first-order answer, and if so, she can expound on that topic effectively without limit, it being the stuff of first-order normative theory, which is remarkably rich. Her primitivism needn't object to *that* kind of question. It's only when

---

[79] *Being Realistic about Reasons*, page 2. Scanlon presents his conception of *normative supervenience* in Lecture 2 of *ibid.*
[80] In insisting that primitivism has this potential for ontological modesty, I'm rehearsing steps taken by many metanormative realists. E.g., Nagel claims that belief in the reality of normative entities doesn't entail belief in the reality of "some mysterious further property" (*The View from Nowhere*, page 144). This is rather like the claim I just made that primitivism doesn't entail foundationalism.

  We must take care, though, that our pursuit of modesty isn't taken to such a degree that we flip around to the other side—*radical* modesty can be spooky too. In his reply to Korsgaard in *The Sources of Normativity*, Nagel says "a substantive realism need not (and in my view should not) have any metaphysical content whatever", i.e., it need not (and in his view should not) posit "moral 'entities', 'facts', or 'truths'" (page 205). This recalls Parfit's metanormative ontology, according to which "normative properties and truths have *no* ontological status. These properties and truths are not, in relevant senses, either actual or merely possible, or either real or unreal. In asking whether there are such normative truths, we need not answer ontological questions" (*On What Matters*, vol. II, 487). These claims, whatever their deflationary intentions, are more ambitious—and indeed, more counterintuitive—than realism needs to be, and though Parfit's within his rights to endorse them as components of his own conception, we mustn't saddle realism *simpliciter* with his eccentric (non-)ontology, not from the start. We ought rather to pursue a strategy of *modest* modesty, namely, maximum agnosticism with respect to ontological matters, except in those places where primitivism forces the issue (such as on the question of whether reduction is possible).

'why *justice*?' is given another intonation, one that brackets first-order theory and the whole lattice of normative and agential ideas, that she must resort to the primitivist's hallmark manoeuvre, namely, to say "backed into this corner, all that I can do is assert that the world is such that it has the idea of *justice* among its constituents". The disposition to avail herself of this manoeuvre is, we must concede, essential to her primitivism, but it's misleading to dwell on it too much, especially if in so doing we neglect the peculiarity of the context in which the disposition is activated.

Who asks 'why *justice*?' with this latter intonation? Not the deliberator, since she takes her own agency, together with its normative materials, for granted. We mustn't conflate, e.g., Korsgaard's normative question, which is sometimes channeled through a series of 'why?' questions,[81] with this ontological sense of 'why?'. Korsgaard's question is a manifestation of our reflective nature; it's essentially deliberative, that is, something asked from the agential point of view.[82] We've already seen that the agential and the normative are intertwined such that any question that arises from the one must presuppose the other as well. And, for that matter, it wouldn't occur to the Korsgaardian deliberator to ask the ontological question. Her subject is *what one ought to do*, rather than *what is*.

Nor is the ontological question asked by those in the midst of some further inquiry in which the idea of *justice* arises, like sociology. It isn't even asked by those who endeavour to develop or amend first-order normative theory. Each of these asks "why *justice*?" with the former intonation, or anyway, intonations near enough to it that they can be answered via appeals to *rightness*, *equality*, *solidarity*, *human need*, *self-respect*, the relations between each of these, and on and on, i.e., more of the stuff of first-order theory. The latter intonation belongs to a perspective that stands outside all that.

---

[81] Like at *The Sources of Normativity*, page 33.

[82] *Ibid.*, page 93. A consequence of this is that normative realists can't address Korsgaard's complaint by simply repeating their allegiances to primitivism. They must formulate a straight answer to the normative question (ideally, by drawing on the deliberative materials that figure in the entities they take to be real). I return to this point in section 3.3.

It arises from out of a kind of ontological curiosity, one that asks "why *justice*?" because it wants an explanation for why we should answer "yes" to a further question, namely, 'is there *justice*?'.[83]

The primitivist can supply a reason to answer "yes", but not one that takes the form of an explanation, since that would require recourse to an *explanans* that's somehow prior to the idea of *justice*, which is exactly what she denies. Her reason to answer "yes" to the 'is there *justice*?' question is expressed by the single-sentence assertion already mentioned: 'the world is such that it has the idea of *justice* among its constituents' (though even this is a somewhat flowery way of characterizing the primitivist's rationale, since all that it conveys is 'there is *justice*'; the primitivist, we might say, gives the simplest 'yes' that can be given). Admittedly, this is of little direct use to the inquirer, since in this context no one, apart from a martian, would ask for the primitivist's reason. We understand that the idea of *justice* has application in the world (though not so frequently as we'd like), so none of us asks "is there *justice*?" with a view to determining whether the claim 'there is *justice*' is true. If we ask, it's because we want to know whether there is any further story to be told. The primitivist answers "no".

To this extent, the ontological intonation of the 'why α?' question is innocent. Still, in asking it, we enter into a treacherous discursive space. It's a thoroughly philosophical space, in that we only undertake to explore it when in the grips of some philosophical impulse. This is no mark against it, but it does recommend our being alert to the many ways philosophy can confuse. We can be forgiven, then, if the 'why α?' question puts us on our guard. We can also be forgiven if we rely, in our efforts to navigate this treacherous space, on stilted modes of expression, like 'it's just there'. Granted, in a different context, we'd find something awkward, even a little deranged, in this sort of

---

[83] The italics are important—'is there *justice*?' isn't the same question as 'is there justice?', i.e., the question that asks whether there is any person now living who, or any institution now operating which, is just. I intend 'is there *justice*?' as shorthand for 'is the world such that something could fall within the purview of the idea of *justice*?'. We must answer "yes" to 'is there *justice*?' if we're to answer "no" to 'is there justice?', since the applicability of the idea of *justice* is a condition for the intelligibility of the second question.

reply. It'd have the look of dogma, or anyway, of unjustified confidence. Or it'd signal a refusal to engage, or a bit of dialogical bad faith, like the kind we use to conceal that we don't fully understand what we're saying. But here, in this peculiarly ontological discursive space, it's perfectly appropriate to say "it's just there". There's nothing suspicious in this line of reply.[84]

This space is home to a familiar ontological project: the drawing up of a census of the denizens of the world. The census-taker's question is 'what is there?', to which the primitivist can reply "there is (among other things) α". The primitivist must reject the ontological sense of 'why α?'—the one that requests a (non-circular) explanation of α (or rather, that asks in its roundabout way whether there is any such explanation)—on the grounds that it presupposes the existence of something that, on her view, doesn't exist. But the census-taker's question is different; there's nothing leading in it. And though the census-taker wants a reason to accept the primitivist's answer, he can't require that this reason identify some additional entity β, which is both antecedent to, and *explanans* for, α. This is because the intelligibility of his project depends on the world's being (not in practice, of course, but in principle) exhaustible, which precludes his espousing anything like an ontological infinitism, i.e., a view according to which, for every object, there's some *explanans* that's prior to it. His conception, whatever it is, must leave room for the possibility that, for any given region of the world, there's some stage at which all of the occupants are counted, and we need look no further. Perhaps the idea of *justice* has its space all to itself within the lattice of normative

---

[84] There are interpretations of the Wittgensteinian programme of philosophical therapy according to which entry to spaces like this is forbidden. The thought, I gather, goes like this: if the "aim in philosophy [is t]o shew the fly the way out of the fly-bottle" (*Philosophical Investigations*, §309), then it would seem to follow (bearing in mind that the idea is famously obscure) that it's incumbent on the philosopher to remove herself, with all speed, from the space in which 'why α?' is asked, and to exorcise whatever impulses led her to it. Primitivism is, we must admit, sympathetic to Wittgenstein's position to some degree (if just because the image of bedrock at *ibid.*, §217 may be the most evocative image of a primitive that we've yet seen), but it can't go so far as to prescribe quietism with respect to ontology, since primitivism is itself a substantive ontological position. Meanwhile, the project of exorcising our ontological impulses is likely hopeless, and anyway, it's at least a little perverse, possibly reactionary, and shows a want of philosophical courage. Caution is certainly warranted, as I've said, but there's nothing to suggest that the obstacles thrown up by the ontological 'why α?' question are more than ordinary analytic rigour can handle.

notions, so that the census-taker, upon learning that there's no one else living at that address, can simply tick his box. At any rate, this is the primitivist's position on the issue. The census-taker's project is innocent, by her lights, as long as it can register that position.

There's the appearance of an asymmetry here, that would, if it were real, pose a problem: the primitivist is able to tell a deliberative story in which she acquires her belief in α in response to some reason, e.g., she encountered α. That is, at a point in this story, she moves between two things, the first being something that arose for her as a reason to adopt her belief, and the second being her belief. But when the census-taker asks for the primitivist's reason for her belief, she replies by, in effect, restating that belief, so that all he has are two instances of 'there is α'. Without two distinct things, he has no deliberative room to move from reason to response. This gives the impression that primitivism can't be of any use to the unconverted, which, in turn, gives the impression that it's dogma after all.

But the asymmetry isn't real—the primitivist doesn't offer the same 'there is α' as both reason and response. Her position is that 'there is α' is true, and her reason is that, by her lights, there is α. This is more than just an empty disquotationalism. Consider a simple case of testimony: I ask whether the cat is on the mat, and am told that she is. I might then ask for a reason to believe it, but this would be to misunderstand what I was told. My interlocutor's answer didn't *announce* a position, but was rather *itself a reason* to adopt a position. Of course, she believes the cat is on the mat, but her aim, in giving her testimony, isn't (simply) to relay that belief. And the primitivist too can say 'it's there' as both avowal and testimony. When the census-taker first canvasses her belief, she gives it as an adherent of her position. When he requests a reason, she answers as a defender of that position. That she utters the same words in both modes is no cause for alarm. At worst, it's just another residue of the awkwardness of the discursive space in which the census-taker carries out his census. (I say 'at worst' because in most cases there's nothing remarkable in the primitivist's

repeating herself. There's a sense in which 'there is α because there is α' conveys her position exactly. A sympathetic audience can hear those words in the way that she intends. And, as always, the primitivist can step out of the ontological discursive space and provide α's coordinates at the first-order—like the primitivist about *justice*, who locates the idea of *justice* in moral theory—so that the census-taker can check for himself.)

Try a different case: suppose, upon being asked for her reason, my interlocutor replies "I can see her on the mat", and this prompts me to reason as follows: now I have a datum, a report from her sensory surfaces of a collision with something cat-shaped, cat-sized, cat-coloured, and so on. The cat's being on the mat best accounts for this datum, so I'm entitled to conclude that the cat is on the mat. Though we may, at times, work our way through something like this series of thoughts, typically we don't bother. In fact, typically it'd be highly peculiar to bother. We certainly mustn't suppose that anything as elaborate as this occurs in every case. And anyway, my fussy inference doesn't eliminate the appeal to testimony; it only shifts it back one step, since in order for the abductive reasoning to go through, I must take my interlocutor at her word that she saw what she saw. Better to receive her report, rather than the likely explanation of her report, as itself carrying the force of a reason. In just this same way, if I'm presented with the claim 'there is α', I needn't venture beyond it, to some further entity β, in order to vindicate it. Sometimes, we do draw up a chain of vindications, with α undergirded by β, and β undergirded by γ, and so on. But not every vindication takes this form. Moreover, unless we're infinitists, we have to permit our chain to reach a bottom at some stage, at a δ or whatever, beyond which point we must avail ourselves of some other kind of vindication. But if we permit as much for δ, why not for α as well?

2.1.2 Reductionism and nihilism

Still, the impression of asymmetry, among other things, has inspired suspicion. At any rate, there are figures—partisan bureaucrats, we'll say—who participate in the census-taking project and who are antagonistic, in their different ways, to primitivism. They are: the reductionist, who suspects that there's more to count than the primitivist allows, and the nihilist, who suspects that there's less. The reductionist persists in asking "why α?" with the narrow ontological intonation because she believes that, *contra* primitivism, α is such that the question in fact applies. For example, Railton argues that, for any entity ψ that exhibits the property of (non-moral) goodness, that entity exhibits an additional, putatively non-normative property, one such that, for some suitably related agent, an idealized instance of that agent would want her non-idealized instance to pursue ψ.[85] According to his reductionist programme, the truth of this counterfactual psychological claim can serve to individuate that complex relational property, so that it might operate as a reduction basis for goodness (provided that goodness can itself be reduced to goodness-for-some-agent). If he's right, then the census-taker will find that, wherever she counts an instance of goodness, it's accompanied by this other property.

The nihilist asks "why α?", as it were, ironically; she means to insinuate that, since where there's no *explanans* there can be no *explanandum*, the refusal to answer the 'why α?' question amounts to the denial that α is real. So, for instance, J. L. Mackie states that "if there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe."[86] This is a claim to which the primitivist about objective values would readily agree, but Mackie intends that it be registered as a complaint. His expectation is that, upon

---

[85] "Moral Realism", pages 173-176.

[86] *Ethics*, page 38. The error theorist about α is in one respect friendlier to primitivism about α than other nihilists, like non-cognitivists or eliminativists, since she acknowledges that α-discourse can be truth-conditional. The two diverge in that the primitivist holds that some of those sentences are true, while the error theorist denies this. What makes this denial nihilistic is the fact that it's motivated by the belief that the relevant sentences are false because they quantify over entities that don't obtain.

our clearly recognizing that primitivism about α entails the *sui generis*-ity of α, we'll come to be so disturbed by this fact as to be obliged to embrace its nihilistic opposite. That is, once we coax a 'confession' from the primitivist about α that she finds a space for α all to itself—that she depicts α as "utterly different from anything else in the universe"—we're to receive it, Mackie supposes, as something like surrender. The argument is transparently question-begging, but many find it compelling all the same.[87]

A category of entity is individuated in terms of the features it doesn't share with any other category. Thus, each is to this extent *sui generis* with respect to the rest, from which it follows that we'd lose everything—not just normative entities, but physical entities too—if we took *sui generis*-ity to be grounds for nihilism. Or rather, we would if we conceived of the world as distributed across categories. It may be that our partisan bureaucrats reject the notion that the world is marked, at the ontological level, by difference. The (non-global) nihilist, at least, can't tolerate any such difference. The reductionist can, though only to a degree, since too much difference is an obstacle for reduction.[88] The primitivist requires more difference than the reductionist can allow.

---

[87] There's room for a more charitable representation of Mackie's argument. He lists "two parts, one metaphysical, the other epistemological" (*ibid.*). According to the epistemological part, because values are *sui generis* entities, they can only be known via a *sui generis* cognitive faculty, which must, in order to represent the practicality of its objects, channel motivational energies to action (*ibid.* page 40). Since we might reasonably find the development of an account of such a faculty daunting, this argument is not so question-begging as the one I ascribe to Mackie in the text (that said, Mackie is mistaken about the nature of motivation, as Myers explains at pages 164-165 of *Donald Davidson's Triangulation Argument*).
  As for the metaphysical part, Mackie raises familiar objections from supervenience (*Ethics*, page 41), but then, these objections aren't particularly strong (after all, nearly everything supervenes). We ought to be honest here: the bulk of the metaphysical work is done by the attempted *reductio*, as I've described it. Moreover, the name Mackie gives to his argument—"the argument from queerness"—suggests that it really does draw most of its force from the fact that, if they obtained, normative entities would be queer, that is, *sui generis*, as though this by itself were sufficient to deny them.
[88] On paper, the reductionist is an altogether different creature from the nihilist. She isn't opposed to *sui generis*-ity itself. But we must take care that we're not overpolite in our dealings with the reductionist. It must be acknowledged that many of her normative instances are indistinguishable from their nihilist analogues. As we've seen (in 1.3.1), the reductionist about α aspires (or claims to aspire) to impose no revision on discourse about α. But, for the most part, the normative reductionist's proposal is so transparently revisionistic—particularly in its making no discernible effort to preserve normative prescriptivity—that it's better represented as a disguised eliminativism than as the reductionism it pretends to be.

The nihilist's partisanship, and, in many noteworthy instances, the reductionist's as well, is manifested in her espousing, among other things, an ontology that goes well beyond the (already very ambitious) commitments of the census-taking project. It enforces ontological homogeneity via a monolithic principle, one that fits everything under a single category (or more precisely, one that makes nonsense of the idea of *ontological category*). The most popular such principle is scientism: all that exists, or can exist, is discoverable by the sciences. There are other monoliths—religious ones, political ones—but none is normative. Normative primitivists are anti-monolith. They have to be, since the deontic ideas are intelligible only given the idea of *agency*, instances of which reside, at least in part, in a non-normative material domain.[89]

Understood in this light, normative primitivism is the insistence that the normative can't be assimilated into the non-normative. It asserts a kind of sovereignty, an opposition to the (would-be) ontological empires of scientism and the rest. This is where the primitivist's refusal to answer the 'why α?' question is at its most polemical. I've described the innocent refusal she gives to the innocent 'why α?' question, the refusal that flags that there's no non-circular explanation of α. But here, her refusal expresses her dissenting, from her dialectical position in the ontological space I've been mapping, against the totalizing programmes of reductionism and nihilism.

We've seen that the primitivist's innocent refusal is narrow, and is moreover the articulation of an altogether legitimate position, one that anyone might occupy in the ontological space in which 'why α?' is asked. The polemical refusal is also narrow, and also the articulation of an altogether legitimate position. The views it opposes are already highly contentious. It's not as though opposition to them is strange. And the primitivist doesn't (necessarily) characterize her

---

[89] Moreover, if a normative conception is to be at all compelling, it must recognize that there are non-normative objects to which normative properties can apply, as well as non-normative properties on which normative properties supervene. Normative primitivists are anti-monolith, but they may nonetheless be monists. The apparent tension between primitivism and monism can be negotiated by way of a device like Davidsonian anomaly.

interlocutor's positions as nonsense. Rather, she takes to be false what they take to be true—the reductionist asserts the existence of some β that explains α from outside, which the primitivist denies; the nihilist denies the existence of α, which the primitivist asserts. She isn't shifting any goalposts, or failing to engage with her interlocutors according to the rules of ordinary philosophical discourse. Her position is at least as prosaic as the reductionist's (and quite a lot less bizarre than the nihilist's).

That said, some primitivists refuse to engage *any* ontological questions. Some refuse to engage *any* questions at all, whether ontological or first-order. But it's not their primitivism, not *qua* bare ontological position, that forces their hands. If they refuse to grant the applicability of these other questions, it's due to views they espouse that go beyond primitivism. This is to say that if a primitivist refuses to account for the possibility of practical thought, it's not her primitivism that's to blame. Though, at minimum, primitivism is characterized by certain refusals, none of them touch the idea of *practical thought*. They only extend as far as a leading sense of 'why α?' and the two partisan census-takers. The immanence constraint permits *those* refusals (and the transcendence constraint demands them).

But disposed as she is to swat away the reductionist's 'why α?' and the nihilist's 'why α?', the careless primitivist might slip into a different, less discerning disposition, one that swats indiscriminately, refusing 'why α?' questions even when asked with the first-order intonation. The temptation to slip is all too familiar—beset on all sides by ontological prejudices, she can't fairly be faulted if all her attention comes to be occupied with them, even to the point that she hears every 'why α?' as painting α with a narrow sort of mystery, the putatively nihilism-confirming, spooky sort. Consequently, she loses sight of the other, wider, richer sort of mystery, the one encountered in ordinary life, which can be solved, and sometimes is solved.

In yielding to this less nuanced disposition, her primitivism is made austere: she's delivered to a point from which all that she can say is that there is something, that it's α, and that this exhausts what can be said about α. Every question that asks for more strikes her as mislaid.[90] Perhaps there is some α for which this raises no problem (though I doubt it), but in cases where α is the idea of a normative entity, this triggers the rational animal's complaint. The rational animal is owed the opportunity to exercise her rationality by inquiring fruitfully into the nature of each normative entity, a right she acquires in virtue of her being necessitated, whether directly or indirectly, by that entity. Austere normative primitivism can't respect that right, which is why the negative face of the thick formulation of the immanence constraint forbids austerity in metanormative thought.

But we need a clearer view of what the negative face forbids. Primitivism *simpliciter* is, I think, a fully determinate position, but austerity is, at least by comparison, indistinct. Concretely, it's the hardest limit that can be placed on discourse about α without eliminating α altogether.[91] But few advocates of austere primitivism would recognize their thinking in this formulation of the idea (for that matter, few would recognize that the term 'austerity' as I've been using it refers to their position). A better way to specify austere primitivism is via an apparently more evocative notion with which it's closely related, namely, the idea of *intuition*. Just like primitivism, intuition has its innocent instances. But there's one idea of *intuition* that's pernicious in that it gives the illusion of the possibility of thought, which may be exploited so as to conceal the fact that no austere conception can meet the thin formulation of the immanence constraint. Once we observe that austere primitivism is tethered to this pernicious intuitionism, we'll be able to see how the thin formulation entails a prohibition on austerity.

---

[90] Korsgaard makes a similar point about intrinsic value at page 109 of *The Sources of Normativity*.
[91] Or that's the idea anyway. It's a consequence of the argument presented in section 2.3 that the notion of this limit is confused, so that this would-be idea of *austerity* is, in fact, incoherent.

2.1.3 A genealogy of ethical austerity

But first: I just traced a genealogy of austere primitivism according to which dialectical pressures triggered a slide from primitivism *simpliciter* to austerity. I doubt this picture comes very near to the truth, or anyway, not so long as these pressures are construed as arising entirely from ontological considerations (though the picture is, I think, of use in explaining how it came to be that less radical primitivisms, like Scanlon's, acquired the appearance of austerity). We need a different genealogy, with a different array of dialectical pressures, if we're to shed any light on why philosophers like Moore embraced austerity.

In Moore's case, and in many others, the relevant pressures, though in a subsidiary sense ontological, are ethical—they deal in the idea of *intrinsic value*. The English ethical philosophy of the nineteenth century laid siege to this idea in at least two places: there was (and remains) Darwin-inspired scientism, with its proclivity for reductionism, but there were also the residues of an earlier reductionist assault, namely, Jeremy Bentham's remarkably severe hedonism, which reduces *intrinsic value* to an especially narrow conception of *pleasure*. If any genealogy for normative austerity can be developed, this hedonism is, I think, the place to start.

There's something worth applauding in Bentham's disruption of English ethical theory, if just because, like Epicurus' own vulgar hedonism, it places intrinsic value within reach of every human capable of sensation (and, for that matter, of every suitably equipped animal), and this provides the basis for an egalitarian morality.[92] After all, if value is only a vibration in a sensory apparatus, no spiritual or social boundary can be drawn around it. To this extent, Bentham displaces

---

[92] *An* egalitarian morality, but not the one we want. The Benthamite utilitarian supposes that, given axiatic vulgarity, we can derive "Bentham's dictum": "everybody to count for one, nobody for more than one" (as articulated by J. S. Mill in *Utilitarianism*, chapter 5). But what vulgarity delivers is different—it isn't each person, but each unit of pleasure (each 'hedon') that's assigned an equal weight. So utilitarianism's emancipatory potential only reaches so far. It has the virtue that, in attending exclusively to hedons, it makes the differences between the bodies that 'host' them ethically irrelevant. But this only amounts to egalitarianism because it assigns no intrinsic value to persons, so that none has more than any other just because none has any at all.

what we want to see displaced, e.g., the lingering Thomism of England's political economy. But Bentham's hedonism relinquishes too much. He conceives of pleasure as entirely sensual, exhibiting neither intentionality nor intelligence. It has two dimensions, intensity and duration, and that's it.[93] And though it's not so scientistic as the later Darwinian reductionisms, it's still quite scientistic. Most conspicuously, Bentham purports to have made value measurable via his (it must be said— highly implausible) hedonic psychology, which represents pleasure as readily quantifiable, and each instance as commensurable with each other, enabling simple aggregation (likewise for pain, and, more puzzlingly, for the two together, as though pain were a kind of reversed pleasure). This, in turn, motivates the introduction of economic analyses to the ethical domain. And then, by way of the consequentialist formula for analyzing *right* in terms of *good*, Bentham draws up his 'felicific calculus' to promise mathematical precision in moral theory.

Each of these things raises its share of problems, but for our purposes, the chief bit of trouble caused by Bentham's conception is this: it gives sense to a question that should be nonsense, namely, 'what recommends intrinsic value?'. That is, if *intrinsic value* is identified with *pleasure*, then the normative gap that separates agency and pleasure—the one we exploit in asking, e.g., "why should it matter that it would feel pleasant? how does this give me cause to pursue it?"—is driven between agency and intrinsic value. But we can't make sense of an agent who doesn't see that, at least in typical cases, intrinsic values give her reason to act. This is because the concept of *agency* requires that the representation of intrinsic value be *prima facie* motivating. As a consequence, no normative gap can be opened between agency and intrinsic value.

---

[93] Bentham also lists "[i]ts *certainty* or *uncertainty*" and "[i]ts *propinquity* or *remoteness*", later adding "[i]ts *fecundity*… [i]ts *purity*… [and] [i]ts *extent*" (*An Introduction to the Principles of Morals and Legislation*, chapter IV, sections II-IV), but these are extrinsic properties. It would seem, then, that on his conception, *pleasure* is itself a somewhat austere primitive, since all that can be determined about its instances is who feels them, at what times, and to what degrees. Even Hobbes (the other English ethical infidel) recognizes that pleasure exhibits intentionality: "[p]leasure … (or *delight*,) is the appearance, or sense of good" (*Leviathan*, chapter VI, paragraph 11).

Thus, Bentham's identity claim can't go through—the idea of *intrinsic value* can't be the idea of *pleasure*. Note that the problem doesn't derive from the fact that certain pleasures are intrinsically worthless or bad. Even if were the case that all pleasures really did qualify as intrinsically valuable, we might fail to know it, and, in light of this possibility, our asking after the goodness of pleasure would remain intelligible. It's the persistent intelligibility of this sort of question that's embarrassing for Bentham's hedonic reductionism. He must hold that no agent who understands the concepts of *intrinsic value* and *pleasure* could feel the pull of the question 'is pleasure valuable?'. But agents do feel that pull, so his hedonism is wrong.

This objection to Bentham's reductionism is a variant of Moore's open-question argument. In its original formulation, the argument states that, if conceptual axiatic reductionism were true, then we'd have to misrepresent an open question—e.g., 'is pleasure good?'—as closed.[94] As William Frankena observed, Moore's argument is question-begging in that the reductionism it attacks is precisely the position according to which the relevant question is, in fact, closed.[95] Recognizing this, Moore's descendants have sought to justify his second premise, the one according to which questions like 'is pleasure good?' are as they appear to be, namely, open. On the variant I've presented, the decisive consideration is that the agential context can't field a 'why pursue that which is good?' question without absurdity. It's in this sense that the 'what recommends intrinsic value?' question is nonsense; it's *practical* nonsense, in that it can't be put in an agent's mouth. And once we see this, the contrast with 'why pursue pleasure?' comes out (I'm assuming that, at least from time to time, the agent finds herself wondering whether pleasure isn't, in the end, worthless). This way of

---

[94] *Principia Ethica*, §13. I have no rigorous notion of what, as a semantic matter, it is for a question—a genuine question, i.e., one meeting the grammatical rules that constitute the interrogative mood—to be closed. We might say that a question is closed if it asks whether an (obviously) tautological (or contradictory) claim is true. But if we do, we'll have to allow that there are contexts in which it's sensible to ask questions that are, on this definition, closed—in a logic classroom, for instance. And if we sometimes have cause to ask closed questions, why should it worry the reductionist that she finds herself asking one?

[95] *Ethics*, pages 99-100.

formulating the argument foregrounds the agent's perspective. I take myself to be entitled to

privilege the agent's perspective in this way on the basis of the wide formulation of the immanence

constraint. It may be that this shift to the agent's perspective, and with it, to the first-order

considerations that move her, looks to be just as question-begging as Moore's own formulation. I

think it isn't, but this isn't the place to make that case.[96]

This by itself doesn't defeat the effort to reduce *intrinsic value* to *pleasure*. It defeats

conceptual, i.e., analytic, normative reductionism, but there remains the possibility of metaphysical,

i.e., synthetic normative reductionism.[97] Still, it charts a promising course. But several of the early

twentieth century defenders of the idea of *intrinsic value* weren't satisfied with this line, not by itself.

They made a further move, a desperate one: they ejected the idea of *intrinsic value* from language, first

by making the term 'good' indefinable, as in Moore,[98] and then by making it nonsense, as in the early

Wittgenstein.[99] Their worry, I think, was that language had already been annexed by science, and so,

if the idea of *intrinsic value* is to be preserved from renewed scientistic assault, it must withdraw to

some extra-linguistic space. The barbarians could have the world, they decided, except that ethics

would have a realm all to itself, forever beyond their reach.

We can see, from our privileged historical position, that they were too quick to surrender

language to science, since the project of assimilating *all* language to science—of revising each

---

[96] The manoeuvre derives from Korsgaard's anti-reductionist arguments in lecture 1 of *The Sources of Normativity*. There's a sense in which this agential variant of the open-question argument owes more to her than it does to Moore. Still, the notion of *practical nonsense* is related to Moore's point about open and closed questions, since a bit of practical nonsense can be explained in terms of the failure to recognize a kind of practical tautology, like 'goodness recommends', a failure that would be evidenced by the question 'is the claim 'goodness recommends' true?'.

[97] It's not clear which of these two varieties Bentham prefers, but conceptual reductionism is, we might say, the default, so I've assigned him to it.

[98] *Principia Ethica*, §6.

[99] "I at once see clearly, as it were in a flash of light, not only that no description that I can think of would do to describe what I mean by absolute value, but that I would reject every significant description that anybody could possibly suggest, *ab initio*, on the ground of its significance. That is to say: I see now that these nonsensical expressions were not nonsensical because I had not yet found the correct expressions, but that their nonsensicality was their very essence" ("A Lecture on Ethics", page 11). Ordinarily, the intention in characterizing ethics as nonsense is to disparage it. But the early Wittgenstein conceives of nonsense as having the potential for something like spiritual resonance, so that his claim 'ethics is nonsense' is intended as reverential.

vocabulary so as to achieve the requisite precision—stalled very quickly. Really, the early twentieth century estimates as to the potential for an all-encompassing scientific language must strike us today as exaggerated to the point of hubris. But even now, we from time to time slip into that century-old habit, the one that has us conflate the limits of science and the limits of language. And since language is our record of what's real, in renouncing value's claim to linguistic expression, we deny it any means of asserting its reality. Indeed, the austere primitivist's proposal isn't that we engineer a retreat on the part of ethics from the material domain to some occult, spiritual place (since even this might, in principle, be absorbed into the material domain via the unrelenting encroachment of the sciences). The idea is rather that ethics is to be made to inhabit no-place. To the extent that we're dazzled by this idea, we allow (uncritically, of course, since a moment's thought reveals that none of this makes any sense) that science has a monopoly on facts, so that value, if it's to escape science, must be non-factual. And, while we're at it, we might as well also allow that science has a monopoly on 'is', so that 'ought', if it's to escape science, must refuse all analysis that would have it quantifying over entities.[100]

This is somewhat closer to the correct genealogy of austere normative primitivism. I'll repeat it: there was the perception of a threat to the putative sacredness of intrinsic value, and this prompted a drastic response, the abandonment of ethical language. Science was depicted as having chased the possibility of ethical profundity from the ordinary world, so that our choice seemed to be between that profundity and ethical expression. Upon recognizing that profundity is essential to ethics (but so is expression!), the choice was made. *Intrinsic value* was classified, not just as a

---

[100] Hume's law (as presented in the last paragraph of David Hume's *A Treatise of Human Nature*, book III, part I, section I) prohibits the transition from 'is' to 'ought', that is, the attempted inference from a non-normative claim to a normative one. It doesn't establish a dichotomy between entities that fall under an 'is' category, and entities that fall under an 'ought' category. Obviously, we don't go wrong in using 'is' (in any of its senses) to characterize a normative entity, as in 'the cat's being on the mat *is* good', 'there *is* a reason to place the cat on the mat', or 'good *is* good'.

The fact-value distinction can't be derived from Hume's law. It requires something like the ludicrously scientistic notion that only that which is empirically verifiable can aspire to be a fact (or perhaps the ludicrously pompous notion that facts are too banal to capture value).

primitive, but as an austere primitive, one with presence enough to warrant awe, but not so much that science could find it.

The fit isn't exact. For one thing, the upshot of the genealogy I've sketched is a position that is, at least *prima facie*, incoherent. It has the look of being an ethics-nihilism, but it's designed to be a repudiation of nihilism (or rather, of a reductionism so revisionistic as to be effectively nihilism). Austere primitivism is my best approximation of this position. As I've said, the austere primitivist about α says three things: there is something, that something is α, and this is all that can be said about α. While these claims *are* expressed in language, the first two do the minimum required to establish α, and the last places the strictest limit that can be managed on discourse about α. Short of α-nihilism, austere primitivism is the most that can be done to prevent discourse about α from being assimilated into science.[101]

2.2 AUSTERITY AND INTUITIONISM

This returns us to the subject of intuition. If the idea of *intrinsic value* lies (to this extent) beyond language, it isn't clear how we can think it. We've seen that, due to the thin-wide formulation of the immanence constraint, the austere normative primitivist needs a way to present her primitive to agential thought. And, at least typically, she asserts that such thought is possible. The trouble is that, given austerity, there can be no mechanism that enables thought about her primitive, since none appears among the austere primitivist's three claims. It must be, then, that she supposes that

---

[101] I've framed austerity in terms of a variety of primitivism, so that austerity with respect to α is given by the view that instances of α have only one feature, namely, their being α (I expand on this idea in chapter 4). That said, there likely are austere metanormative conceptions that aren't best captured by this variety (or perhaps any other variety) of primitivism. Moore and the early Wittgenstein both favour normative austerity, but only Moore is a primitivist of the kind I have in mind. The early Wittgenstein likely isn't a normative primitivist at all. The austerity in his conception takes a different shape—I suspect it involves his (notoriously cryptic) saying/showing distinction (*Tractatus Logico-Philosophicus*, 4.121-4.1212). My focus is austerity in Moore's tradition, the kind that falls pretty cleanly under the *primitivism* heading. I set the other manifestations of austerity in metanormative philosophy aside. Still, my suspicion is that each fails by the thick formulation of the immanence constraint in largely the same way.

no mechanism is necessary, that is, that her primitive is directly available to agential thought. Call this unmediated relation of thinker to object 'intuition'—or rather, because many different things go by that name, 'radical intuition'. Radical intuition, so the story goes, can operate even in the absence of any feature in its object. It can know across vacuum, without any channel for its knowing, and without any exchange. Its object may be altogether featureless; this is no obstacle. It just knows, directly and inexplicably. A conception of *practical thought* is intuitionistic, in the relevant sense, if it represents agential access to normative entities as a matter of radical intuition. Austere primitivism must be intuitionistic if it's to satisfy the thin formulation of the immanence constraint.

Granted, no austere primitivist elaborates her position in quite this way. But then, I can't see another story that she can tell. As it happens, though many austere primitivists assign a central place to the term 'intuition', they rarely offer any detailed formulation of what they mean by it. So, for instance, Moore says that "[i]n order to express the fact that ethical propositions … are incapable of proof or disproof, I have sometimes followed Sidgwick's usage in calling them 'Intuitions'."[102] This tells us nothing positive about how these intuitions might figure in thought. After all, gibberish is also incapable of proof or disproof. What, then, distinguishes gibberish from intuition? And the comparison to Sidgwick isn't as helpful as it might appear since each uses the term 'intuition' to refer to strikingly different things. Sidgwick's 'intuition' refers to a power rather than a property of propositions.[103] Also, the outputs of that power, though they don't *require* proof, aren't, for all he's said about them, *incapable* of receiving any. This is all to say that, in order to elaborate Moore's view, I couldn't rely on his own account very far. Instead, I've had to extrapolate his view from the

---

[102] *Principia Ethica*, preface. He also says "I beg that it may be noticed that I am not an Intuitionist, in the ordinary sense of the term", but what he means is that his intuitionism extends to axiatic entities and not to deontic entities (he supposes that, in 'ordinary' cases, it's the other way around). His strategy for the determination of deontic entities is the consequentialist's formula favoured by Bentham among others, the one whereby *right* is analyzed in terms of *good*.
[103] "We have the power of seeing clearly to some extent what actions are right and reasonable in themselves … This power it is convenient to call (as it is commonly called) the faculty of Moral Intuition" (*The Methods of Ethics*, book III, chapter 1, §1).

features of his primitivism as I understand it. Thus, since his ontological view is austere, his use of the term 'intuition' must be correspondingly radical.[104]

2.2.1 Varieties of intuition

What can be said for the idea of *radical intuition*? A case might be made for the application of the idea in certain special contexts. The recognition of the truth of an analytic claim—e.g., 'a normative reason is a consideration that counts in favour of some action or attitude'—is immediate, at least provided that we have mastery of the concepts involved, as well as the opportunity to attend to them. This is because, as Kant says, the claim's predicate is "contained in [the] concept", so that the one can be read directly off the other.[105] Perhaps in recognizing an analytic truth, and specifically in recognizing its 'self-evidence', we exhibit an intuitive power, even a radical one.[106]

But the austere primitivist needs a different kind of intuition, one capable of synthetic connection. After all, if α is austere, it's unanalyzable apart from the maximally uninformative identity claim 'α is α'. Moreover, because we act from particulars, the austere normative primitivist must explain how we can think about her primitive specifically as it's instantiated in particular entities, since otherwise that thought couldn't be practical. But it's only via synthetic judgment that an entity can be subsumed under a primitive. And because Moore's primitive is non-natural, his intuitionism can't be perceptual. We might think of it as a rational synthetic intuitionism, though

---

[104] But he insists:

> [a]gain, I would wish it observed that, when I call such propositions Intuitions, I mean *merely* to assert that they are incapable of proof; I imply nothing whatever as to the manner or origin of our cognition of them. Still less do I imply (as most Intuitionists have done) that any proposition whatever is true, *because* we cognize it in a particular way or by the exercise of any particular faculty[.] (*Principia Ethica*, preface)

Doesn't this show that he explicitly disavows the radical intuitionism that I assign to him? No, it confirms that his intuitionism is radical. Unlike a perceptual or affective intuitionism, which have something to tell us about the role of the mind in intuition, there's no discerning the faculties that are in play (or even if there are any) in radical intuition. That is, there's no psychology for radical intuition. If there were, it wouldn't be radical.

[105] *Critique of Pure Reason*, page A6/B10.

[106] Even here, though, the idea doesn't apply cleanly, since any successful cognition would seem to be mediated by its antecedent acquaintance with the relevant concepts.

only if we construe the term 'rational' so that it signals an orientation—in this case, a cognitive orientation—to truth, and not as involving the operation of what we might take to be a rational power, like inference.[107]

Synthetic intuitionisms are familiar enough. In saying, e.g., "it's my intuition that middle-sized objects exist", I give expression to what might reasonably be characterized as a synthetic intuitive power. This sense of the word 'intuition' is, I imagine, a philosopher's contrivance. In less fussy contexts, we might say "it seems to me that middle-sized objects exist"; the 'it seems to me that' operator is, I think, the equivalent of the philosopher's 'it's my intuition that'. Some, like Scanlon, prefer the term 'seemings'.[108] At any rate, this sense of 'intuition' doesn't refer to radical intuition. It's far more ordinary than that. It's the sort of thing that arises in the course of a kind of exercise, one in which participants are invited to devise synthetic judgments *specifically* in the absence of any expectation that those judgments are supported by evidence, let alone proof. The function of these intuitions isn't to ground any claims, at least not immediately. After all, they're thoroughly defeasible things that don't by themselves command any justificatory power. Their role, when they have one, is to provide the beginnings of deliberation by operating as something against which to orient our thinking. In this respect, they're like what Rawls calls "provisional fixed points"—fixed in that we're to take them for granted as we set out on some stretch of exploratory thought, but provisional in that we're always free to dispense with them on subsequent expeditions.[109]

---

[107] Korsgaard explains that it was Richard Price, a contemporary of Kant's, who popularized the idea of *rational intuition*, and that he did so specifically to mark an alternative to the then-dominant empiricist conception according to which all simple ideas are cognized via sense (see her "Realism and Constructivism in Twentieth Century Moral Philosophy", page 307). Rational intuitionism cognizes simples via reason rather than sense.

For what it's worth, Kant's diagnosis would be that Price's (and Moore's) prospects are grim: "in synthetic judgments I must have in addition to the concept of the subject something else (*X*) on which the understanding depends in cognizing a predicate that does not lie in that concept as nevertheless belonging to it" (*Critique of Pure Reason.*, page A8/B11). The conjunction of austerity and imperceptibility leaves us with no recourse to this *X*, not unless Price invokes a kind of occult power (the likes of which Kant would promptly dismiss as dialectical illusion).

[108] *What We Owe to Each Other*, pages 65-66.

[109] *A Theory of Justice*, page 18.

There are different species of intuition in this deliberation-enabling region, some of which are stronger than others in that their thinkers are disposed to place more confidence in them. Seemings may be the weakest of the set. Really, they're a bit like wild guesses, except that they don't have assertoric force. Suppose I offer a wild guess as to how many cats are on the mat. If I guess right, then *I'm* right. This is the case even though I haven't *done* anything right, that is, nothing praiseworthy (after all, it was only a wild guess). But if I have an intuition in the weak 'seeming' sense about how many cats are on the mat, and my intuition is right, it doesn't follow that I'm right. Likewise, if it were the case that middle-sized objects don't exist, it doesn't follow that I'm wrong (relatedly, if I hallucinate that the cat is on the mat, when in fact she isn't, I haven't yet gone wrong, not unless I form the corresponding belief). This shows that we're not yet dealing in assertions. It may also show that we're not yet dealing in judgments (though intuitions do represent their contents as true).

Other intuitions fall under this same *provisional fixed point* category but are assertoric. They inhabit the space between mere seemings and full-blown beliefs. Some are guesses, but others are better characterized as conjectures or hunches, with corresponding increases in thinker confidence. I've from time to time used the term 'intuition' in this dissertation in order to represent *prima facie* reasons for belief, that is, representations that can't by themselves recommend belief, not without further scrutiny, but that claim to gesture in the direction of a genuine reason for belief. Even these aren't thought with such confidence as to amount to a belief, not at first—*I* believe them, but only because I take them to be confirmed by philosophical investigation.[110]

---

[110] Rawls' own provisional fixed points aren't intuitions at all. They're considered judgments, i.e., full-blown beliefs that, from the start, are thought with remarkably high confidence. Moreover, we can typically offer a rationale for our considered judgments if challenged. Or anyway, we can do more than shrug and say "this is how it seems to me".
  There are representations of Rawls' view that translate his term 'considered judgment' into 'intuition', but this only goes to show how unstable the term 'intuition' is, and how important it is to stipulate a determinate meaning for it.

Clearly, though, this genus of defeasible, deliberation-enabling synthetic intuitionism can't be what the austere normative primitivist wants. Her intuitions aren't starting points for inquiry. If they were, she'd have to allow that the apparatus of practical inquiry extends past intuition. But because her conception is austere, she can't sustain any extension in that apparatus. It has nothing else to tell the rational animal as she inquires into normativity. Rather, the entire process of knowing is to be assimilated into that first step, i.e., into the intuiting. (I suspect it follows that she must hold of her intuitions that they are capable of delivering certainty—or if not that, incontrovertibility— since she can offer no resources with which to dispute or defend a normative claim.)

2.2.2 The contrast with reflective equilibrium

Let's dwell on Rawls a bit, as a foil for radical intuitionism. He isn't an intuitionist, not even a modest one, and he routinely contrasts his view with intuitionisms of various stripes, like Ross' moral intuitionism,[111] and the epistemological intuitionisms of Moore and Prichard, among others.[112] But he would raise no objection against the relatively innocuous intuitions I described in the preceding subsection.[113] More to the point, his view is at least intuitionism-adjacent, if just because his variation on Nelson Goodman's method of reflective equilibrium is shaped in large part by aspects of Sidwick's moral epistemology.[114] We've seen that Sidgwick's term 'intuition' refers to a

---

[111] *Ibid.*, §8.

[112] *Political Liberalism*, lecture III.

[113] That said, there's little reason to believe he would place much stock on data collected on the basis of these intuitions, given that they have so little to say for themselves. For that matter, there's little reason to believe he'd applaud the use of thought experiments engineered to 'pump' intuitions, as though this were any cognitive advance. (The original position isn't a thought experiment, or anyway, not one designed to pump intuitions, and again, considered judgments aren't intuitions.) As it happens, in "Outline of a Decision Procedure for Ethics", Rawls insists that the class of considered judgments be limited to actual cases—"all judgments on hypothetical cases are excluded" (§2.5 iii). He relaxes this criterion later, though he remained, I suspect, suspicious of moral intuitions generated via that well-worn philosopher's technique: the martian example.

[114] In particular, Sidgwick's account of the conditions "the complete fulfilment of which would establish a significant proposition, apparently self-evident, in the highest degree of certainty attainable" (*The Methods of Ethics*, book III, chapter 11, §2).

power; Rawls' own approach to moral philosophy begins with the exercising of a putative power of moral cognition, the sense of justice. The findings of this power are judgments, but their role in Rawls' investigations into social justice isn't to represent moral reality directly. Rather, they provide starting materials, and are vindicated, to the extent that they are, by their position in an elaborate network of other judgments, principles, empirical knowledge, and theory-building considerations. Rawls says about Sigdwick that, "although there are elements of epistemological intuitionism in his doctrine, these are not given much weight when unsupported by systematic considerations."[115] Swap out the offending epistemological elements and you're left with something that's, in broad outline, rather like Rawls' position. (It's remarkable that Rawls and Moore diverge to the degree that they do, given how much each borrows from the same source.)

Rawlsian reflective equilibrium provides the outlines of a set of rational procedures with which to establish the truth of synthetic normative claims. As I said, it begins with the sense of justice; this power generates moral judgments, which are subjected to careful scrutiny; the surviving judgments are modelled by principles, themselves collected together into a theory, with the two, judgment and theory, revised repeatedly so as to better fit the other; finally, the preceding step is repeated for each of the "conceptions of justice known to us through the tradition of moral philosophy and any further ones that occur to us".[116] (In Rawlsian nomenclature, the process is characterized this way: beginning from the sense of justice, we move to considered judgment, then to narrow reflective equilibrium, and lastly to wide reflective equilibrium.) Each step is tailored to refine the workings of a cognitive power, itself assumed to be largely on the right track,[117] until we arrive at a set of claims that are, at least, highly credible.

---

[115] *A Theory of Justice*, note 26 at page 45.
[116] *Ibid*, §9.
[117] Rawls also assumes that the tradition of moral philosophy is largely on the right track. This is what recommends the final step of his procedure, the move to wide reflective equilibrium. Myers argues, and I agree, that we can shore up

This method is remarkably rich. It offers extensive materials for the manifestation of our rational nature in moral thought. As such, implementing Rawls' strategy would go a long way to satisfying the thick formulation of the immanence constraint. But it requires that we be given machinery with which to scrutinize our judgments, to model them in principles, and to locate them within a tradition of moral thought. Austere primitivism precludes all such machinery. This is why it's saddled with radical intuitionism. It must say that intuition conjures up, *ex nihilo*, a relation of knowledge between subject and object, entirely unaided, and entirely mysterious. This is the best that it can do.

2.2.3 Failure by the negative face of the thick formulation

Clearly, this is no use to the rational animal. As the radical intuitionist would have it, primitives needn't signal their locations to their would-be knowers. They needn't, for that matter, provide any clue. Any idleness from the thing-known can be, so the intuitionist would have us believe, compensated by the intuitive power of the knower. It is she who is to shoulder the whole of the work. Perhaps this can be done.[118] But there's nothing in radical intuitionism that can serve to accommodate our rational nature. Really, radical intuitionism *just is* the failure to meet the thick formulation. Or better, it is the abandonment of the project of meeting it.

---

these aspects of the reflective equilibrium strategy by way of a Davidsonian account of the origin of normative content. See *Donald Davidson's Triangulation Argument*, pages 174-175.

[118] It all sounds a bit magical—we might just as well grant that by concentrating very hard we could penetrate solid stone. Still, knowledge is a curious thing, and we oughtn't to rule something out just because it's bizarre.

But here's a puzzle for radical intuitionism: knowledge implies a kind of receptivity on the part of the knower, but radical intuitionism depicts the knower as too much the architect of her own knowledge. Why, then, shouldn't we construe radical intuition as the output of an imaginative rather than a cognitive power?

Perhaps the idea is that the object of intuitive knowledge isn't *entirely* inactive, since it contributes to its being known by being *there*, lying ready to be discovered. The trouble is, we don't get the idea of 'being there' for free. Ordinarily, to 'be there' is to be situated in a rich lattice of relations. But if we have a rich lattice, we don't need radical intuition; austere primitives need radical intuitionism precisely because they don't figure in any such lattice. But then what is it for them to 'be there'? It's not enough, I think, for them to inhabit logical space. They must be more substantive than that.

Given that this is the case, I imagine austere primitivists would be reluctant to endorse radical intuitionism. If they're inclined to allow that they're intuitionists at all, the variety of intuitionism they intend is likely to be somewhere between the radical and the provisional varieties mentioned in the preceding subsection. My earlier taxonomy wasn't exhaustive; there are many more middle cases from which to choose. Suppose I observe that the cat wants the mat, and upon being asked how I know this, I reply "intuition". My reply doesn't (necessarily) involve anything radical, since I might, if pressed, indicate a tilt of the whiskers or a disturbance in the tail as supplying some part of my evidence. But what, then, does my reply convey? Perhaps its role is to signal that I can't specify how it is that I know that the cat wants the mat, or anyway, that I can't do so in precise (or concise) terms. Or it may be that the context is such that 'intuition' is shorthand for a complex series of observations. It may be that I simply can't be bothered to provide some more elaborate answer.

Though any of these may be closer to the sense of 'intuition' actually favoured by austere primitivists than the radical sense I've assigned to them, each performs a relatively humdrum discursive role, nothing like what the austere primitivist needs in order to establish the possibility of contact between a thinker and an austere primitive. Granted, there remains the bizarre case in which, in saying "intuition", I mean to ground my cognition in some occult faculty, like a spiritual modality. If so, I'm confused (unless, as is likely, I'm joking, or else trying to appear mysterious, or making a point about the depth of my insight into the cat's mind). But even if we were prepared to countenance an occult faculty, it must, like any other faculty, purport to register some feature in its object. Austerity can't accommodate the supernatural any better than it can the natural.

There are many other varieties of intuition and intuitionism, too many to record here. But we don't need an exhaustive list—the pattern is clear enough: each variety of intuition is either i) radical, ii) no basis for a conception of *practical thought*, or iii) incompatible with austerity. And, to the

extent that the austere primitivist seeks to develop a conception of *practical thought*, she must, for lack of a discernible alternative, do so on the back of a variety of intuition. Thus, austere primitivism is tethered to radical intuitionism. The latter offers no explanation of the possibility of practical thought. It follows then, that the former offers no such explanation either.

The negative face of the thick formulation—the one that forbids metanormative austerity—derives from this relationship between austere primitivism and radical intuitionism. I suspect some austere primitivists could be persuaded to abandon their views by being made to acknowledge this relationship. After all, *radical intuition* is a very mysterious idea. Still, many austere primitivists would see no force in this line of objection. They may grant that the thin formulation of the immanence constraint must be respected, but deny that the negative face, and for that matter any part of the thick formulation, follows from it. We are coming very near the point at which we must allow that not all will be convinced. Still, there's a last consideration that can be exploited to show that the negative face is a consequence of the thin formulation, namely, the discursivity criterion.

2.3 PRACTICAL THOUGHT AND DISCURSIVITY

According to the discursivity criterion, if a conception is to establish the possibility of a thought, it must establish the possibility of fruitful conversation about that thought. The relationship between these two possibilities is quite tight—given the nature of thought, for any given thing, if we can think it, then we can say it, and if we can say it, we can say it to another; necessarily, there is no inarticulable cognition, and no incommunicable articulation. Conversation is fruitful if it can be informative to one of the conversing parties. This excludes transparently empty forms of exchange, like those in which parties assent to tautologies, e.g., 'α is α', or to bare testimony. Substantively, what's required is the possibility of an exchange of reasons—it mustn't be the case that conversation stops dead at 'how do you know?'. Generally speaking, the best opportunities for an exchange are

to be found at the first-order, since it's most congenial to fruitful discussion. Unusual metaphysical contexts, like the one described in subsection 2.1.1, provide fewer opportunities for fruitfulness to arise. But austerity conflicts with the fruitfulness proviso in any context. Therefore, no austere conception can meet the discursivity criterion.

This criterion is what shows the thick formulation of the immanence constraint to follow from the thin formulation.; it's the fulcrum of the argument. While ordinarily, a possibility could be established without its being explained, the possibility of thought about α is peculiar, in that it requires that α be available as subject matter for discourse. And since discourse must (if it's to be fruitful) have materials for its elaboration, it must be that the thin formulation includes this facet: it demands that metanormative conceptions supply the relevant discursive materials. These materials derive from many sources, but they must, as I will argue, include the materials that figure in explanations of the possibility of practical thought. If this is right, there's no meeting the thin formulation without meeting the discursivity criterion, and no meeting the discursivity criterion without meeting the thick formulation.

This point deserves special attention. The discursivity criterion brings out the difference that thought makes. If, for instance, we sought to establish the possibility, not of thought about α, but simply of α, we wouldn't thereby incur an obligation to explain α. That obligation—i.e., the thick formulation—derives from what's owed to the thinker, and not to reality as such. We're free to entertain the possibility of some wholly nondescript α, or even some unthinkable α.[119] Mackie might

---

[119] Or anyway, my argument is compatible with our having this freedom. I'm not certain how far this freedom goes. It may be that consciousness of the limits of thought—which is to say, consciousness of our own subjectivity—involves the acknowledgment that it's possible that some things are unthinkable. And, thinking about *the possibility of an unthinkable thing* doesn't amount to thinking about *an unthinkable thing*, so there's no obvious contradiction in it. Still, I'm not at all certain what the thought could be. Nagel addresses this issue in *The View from Nowhere* (particularly in chapter VI, where he frames it—unhelpfully, I think—in terms of realism and idealism), but despairs of dispelling the air of paradox.

object that such an α can't be located among the other entities of the universe. But this only makes it *sui generis*, which, as we've seen, provides no reason to deny its possibility.

But our subject isn't simply α. It's the idea of *normative reason* (as well as the other deontic notions, and, via deontic ubiquity, the entirety of the normative domain). And this idea has the remarkable feature of being essentially thinkable. This is what the thin formulation underscores: the special relationship between normativity and thought. And since thought comes with its own requirements, we can't entertain the possibility of a normative entity all on its own. It must be accompanied by quite a lot else.

Accounting for normative discursivity is a first step in the direction of making good on the agent-relationality of deontic entities. By recognizing the force of the discursivity criterion, we recognize that normativity can't be wholly detached from the life of the rational animal, that even though it belongs to the nature of normativity that it transcends animal life, it must nevertheless perform a role in enabling the manifestation of that life, beginning with the activity at the centre of rational life, namely, talking to one another. This is what grounds the metanormative theorist's obligation to carve out a space from which one animal can say to another "here is a reason", and be understood. Without this, her conception of *normative reason* is incomplete.

I'll proceed in this way: at minimum, satisfaction of the discursivity criterion requires that we be given some device by which to assure interlocutors (defeasibly, of course, but then we don't need a guarantee) that they're talking about the same thing. But, as I will show, this assurance can only be supplied by an account of the origin of the determinacy of our meanings, i.e., of what it is that assigns them to one thing rather than another. This same account explains the possibility of those meanings, which, in turn, explains the possibility of our entertaining them in thought. This is why meeting the discursivity criterion and meeting the thick formulation of the immanence constraint come to the same thing.

2.3.1 The aspect problem

But before I begin, I must lay my Davidsonian cards on the table. My interpretation of the discursivity criterion is pulled, almost in its entirety, from Davidson's philosophy of mind and language, particularly his triangulation argument, which establishes that, given semantic externalism, we require linguistic communication with others in order to fix aspects of our shared world, and in this way to acquire those aspects as contents for thought and language.[120] As Davidson shows, the world can't by itself furnish us with meanings. Its objects must be disambiguated first. We contribute to their disambiguation by occupying perspectives from which they arise under determinate aspects. But then, no wholly solitary person can succeed in occupying such a perspective, since she can't acquire any notion of the world as distinct from her own experiences of it. So, if she is to think and speak, she must encounter a second person, recognize her as an interlocutor, and come thereby to appreciate that her own perspective is one among many. By interacting with one another, the two can respond together to some worldly object—they can triangulate on it—and in this way specify the aspect under which it enters into their thinking and speaking.

The triangulation argument offers an explanation of the possibility of thought, and when supplemented with a normative naturalism, it offers an explanation of the possibility of practical thought as well.[121] Thus, Davidson plots a path to the satisfaction of the thick formulation of the immanence constraint. But our concern for the moment is just one part of his argument, what I'll

---

[120] The triangulation argument is developed across several papers, beginning with "The Second Person". I rely on Verheggen's reconstruction in *Donald Davidson's Triangulation Argument*, pages 16-24. By 'semantic externalism' I mean the view that our meanings are partly determined by causal transactions with our environment (Verheggen calls this 'physical or perceptual externalism' at page 4 of *ibid.*). See Davidson's "Externalisms".

[121] Normative naturalism is necessary in order to grant normative entities the power to causally determine would-be thinkers and speakers. Myers forms a normative naturalism out of Davidson's (relatively sparse) metanormative remarks. See *Donald Davidson's Triangulation Argument*, pages 172-173 *et passim*.

call, following Verheggen, 'the aspect problem'.[122] Even if we suppose that semantic externalism is

true, it can only establish a causal, i.e., an extensional, relation between the objects in our

environment and the meanings of our terms.[123] Since meanings are intensional, they require a richer

basis. So, for example, in my encounter with the cat on the mat, I am presented with materials out

of which to coin the term 'cat', but because, initially, these only impress themselves on my sensory

capacities, I don't yet have any means of privileging one aspect of the cat over the others. As we've

seen, the sentence 'the cat is on the mat' has among its (infinite) truth-functional equivalents the

sentence 'the deepest thinker now living is on the mat'. The cat *qua* cause arises, not under one

aspect, but under all of them, so that my causal transaction with her can't by itself bring it about that

I mean *cat* by 'cat' rather than *deepest thinker* by 'cat'.

A brief digression for completeness' sake: the aspect problem reveals the semantic poverty

of causal relations; this might be taken to imply that we could avoid the problem by abandoning

semantic externalism. This isn't the case—even if we opted instead for a semantic internalism, we'd

be met with the same ineliminable ambiguity. After all, such an internalism must identify mental

items that in some sense correspond to the thoughts to be explained, but that are themselves non-

contentful (for otherwise they'd require their own explanations). But since they're non-contentful,

they can't direct any thinking without their being interpreted, and there are infinitely many

interpretations that can be made to fit.[124] The trouble isn't causality, but the lack of intension; with

---

[122] In "The Emergence of Thought", Davidson represents the aspect problem as an "ambiguity concern[ing] how much of a total cause of a belief is relevant to content" (pages 129-130). As with the wider triangulation argument, I draw on Verheggen's formulation of the problem (at pages 18-19 of *Donald Davidson's Triangulation Argument*), as it's quite a lot more detailed than the Davidsonian original.

[123] Bare causal relations are extensional (see Davidson's "Mental Events", page 215). That said, the explanatory force of a causal statement, if it has any, is description-sensitive, that is, intensional. Moreover, certain causes explain via appeal to intensional entities, like attitudes. Reasons-explanations are doubly intensional in just this way (see Davidson's "Actions, Reasons, and Causes"; that their *explanantia* are intensional, and are enlisted as explanatory specifically *as* intensional, is made clear by Davidson's response to his deviant causal chain problem at the end of "Freedom to Act"). But the causality that semantic externalism brings to bear is the non-explanatory, extensional variety. It must be, since otherwise, it couldn't figure in the origin of intensionality.

[124] This is Wittgenstein's objection to semantic internalism (as articulated at §73 of *Philosophical Investigations*, and elsewhere). As Verheggen observes, Wittgenstein's objection is another version of the aspect problem.

respect to this lack, non-contentful mental entities are just like worldly entities. Thus, the aspect problem is an obstacle for the possibility of thought no matter how conceived.

We're confronted with this same problem in the normative case. The possibility of the cat's comfort gives me a reason to place her on the mat; this reason may determine me causally (in the extensional sense of 'causality'—perhaps it activates a particularly sophisticated bit of Pavlovian conditioning). But even if so, this doesn't yet equip me with a means of thinking of that reason *qua* a reason, that is, a consideration in favour of some action or attitude. The reason as cause can be redescribed in infinitely many ways, many of them non-normative—it may be the first thing an admirer of cats would observe upon entering the room—so that we have no assurance that, in responding to it, I'm responding to it as a reason rather than under this other description.

How do we single out one description among many? Not by endeavouring to narrow the causal exchange, e.g., by arbitrarily eliminating the 'deepest thinker' description. The range of interpretations in play really is infinite. We can give up talk of 'the deepest thinker', but we're left with 'the thinker one rung deeper than the second deepest', and on and on. However much we might subtract from infinity, we'll always be left with infinity. The solution isn't to avail ourselves of more and more of the world, but to appeal to the semantic activity of the thinker. She has to settle for herself that she's thinking of the cat *qua* cat rather than *qua* deepest thinker now living. As Verheggen says, "since no non-intensional magic trick will do to fix the causes, and hence the meanings, of one's thoughts and utterances, only those producing the thoughts and utterances could achieve this feat."[125]

Agency solves the aspect problem directly. Because the semantic agent is, so to speak, constituted by thought, i.e., by intensional elements, she's able to legislate for herself that she means this rather than that aspect, without thereby pretending to cobble together an intension out of

---

[125] *Donald Davidson's Triangulation Argument*, page 20.

extensional pieces. This isn't to say that she can conjure meanings *ex nihilo*; she depends on conversation with others to ensure the intelligibility of what she says, and on the world to offer up the matter of her thought—it has its array of candidate aspects, and she selects one from among them. So, the solution to the aspect problem is given by this unity of subject and world. But we're not entitled to this solution just yet, since it presumes what we sought to explain, namely, the possibility of thought. We must take a step back to determine how it is that the semantic agent acquired her self-legislating power to begin with.

But before we do, I'll say three things to situate the aspect problem within the argument connecting the thin and thick formulations of the immanence constraint. First, if we're to establish the possibility of thought, we must solve the aspect problem. This is because the idea of *thought* takes the determinacy of thought for granted, and the aspect problem challenges that determinacy. It follows from this that any conception that meets the thin formulation includes a solution to the aspect problem. Second, the aspect problem suffices to deliver the negative face of the thick formulation. Austere primitivism has no way to solve it, and since it must be solved, this means that austere primitivism is doomed. If my ambition were only to vindicate the negative face, I could stop here. But my ambition is to defend the immanence constraint construed as a demand for explanation, and with it, the positive face of the thick formulation, with its demand for theory. So, I present the aspect problem as a part of the elaboration of the discursivity criterion, and not as the basis for an independent argument against austere primitivism (though this is ready at hand, should we need it). And third, while the discursivity criterion directs us to the aspect problem by asking how two people could be assured that they're discussing the same thing, it can itself be motivated by the aspect problem, since the solution to that problem can, I think, only be found in Davidsonian triangulation, which is itself a form of fruitful discourse. My hope is that the discursivity criterion

needs no explicit defence, since it's, I think, *prima facie* compelling. But there is this Davidsonian

defence, such as it is, that can be marshalled for it.

2.3.2 The poverty of the solitaire

We want to know how the semantic agent acquires her capacity to fix the aspects of her thoughts

and meanings. This is where our social nature is brought to bear. A single subjectivity won't do,

since no single person could be conscious of herself as a subjectivity, and, as we'll see, there's no

self-legislation without self-consciousness. Note, though, that in venturing to describe the origin of

content I've entered a stretch of terrain no part of which can be mapped. Davidson explains that

> [i]n both the evolution of thought in the history of mankind, and the evolution of thought in
> an individual, there is a stage at which there is no thought followed by a subsequent stage at
> which there is thought. To describe the emergence of thought would be to describe the
> process which leads from the first to the second of these stages. What we lack is a
> satisfactory vocabulary for describing the intermediary steps… It is not that we have a clear
> idea what sort of language we could use to describe half-formed minds; there may be a very
> deep conceptual difficulty or impossibility involved. That means there is a perhaps
> insuperable problem in giving a full description of the emergence of thought.[126]

The trouble is that we expect—for we can't envision an alternative—an explanation of the

emergence of thought whereby we acquire the relevant capacities by steady expansion, one thought

after another. But each thought is constituted in part by its relations to other thoughts (as well as to

other attitudes, some of them conative). This is due to the holism of the mental; as Davidson says,

"we make sense of particular beliefs only as they cohere with other beliefs, with preferences, with

intentions, hopes, fears, expectations, and the rest… the content of a propositional attitude derives

from its place in the pattern."[127] Thus, the origin of thought can't take the form of a series of

---

[126] "The Emergence of Thought", pages 127-128.
[127] "Mental Events", page 221; also "[i]n order to believe the cat went up the oak tree I must have many true beliefs about cats and trees, this cat and this tree, the place, appearance, and habits of cats and trees, and so on" ("Rational Animals, page 98). Verheggen explains how the triangulation argument confirms the principle of the holism of the mental at page 24 of *Donald Davidson's Triangulation Argument*.

incremental advances; from the start, we must have an astonishing number of mental powers, as though all at once. There's no way to describe the creature on the threshold of language, no way to describe inchoate semantic agency. So, my intention in offering a characterization of the origin of content is that it be received as something like a metaphor, or anyway, as an image that's suggestive of how the pre-linguistic could graduate to the linguistic.

Given that the subjective/objective distinction isn't yet in view, the term 'single subjectivity' doesn't apply as cleanly as I'd like (perhaps a genuine subjectivity is always a self-conscious subjectivity). The term 'solipsist' would be more appropriate, if not for the epistemological connotations.[128] Verheggen uses the term 'solitaire'—I'll do the same. Like any animal, the solitaire must navigate the world in order to carry out a life in it. But this by itself can't compel her to conceive of the world as independent of her representations of its objects, that is, of its autonomy with respect to her representations. For that matter, it can't compel her to conceive of her representations as independent from the world. It certainly can't convey to her the possibility of her going wrong with respect to the world, since this would require that a representation be able to subsist in her, so to speak, and not in the world.

I'll elaborate on this last point: in order to see that my notion that the cat is on the mat misrepresents the location of the cat in the world (suppose she's shifted to the couch), I'd have to hold that notion up against the world and spot the mismatch. But what reality could be granted to the notion, since, given that it's mistaken, it can't find its home in the world? I must, as it were, give it lodging within myself. This is the very thing the solitaire can't do. The most that can be assigned to her is a kind of *de re* fallibility, according to which certain of her responses to worldly goings-on indicate that she's missing something salient to her interests, like when the cat doesn't recognize that

---

[128] Even so, Davidson uses it in connection to the semantic incapacity of the single subjectivity, saying "[t]he solipsist's world can be any size; which is to say, from the solipsist's point of view it has no size, it is not a world" ("The Second Person", page 119).

I have a treat in my hand. That isn't a mistake in thinking. If a glance suggests that the cat is on the mat, but further investigation reveals that the mat is empty, the best that the solitaire can manage is to suppose she's dealing with a peculiar cat, or a peculiar mat, or a peculiar set of circumstances in which cat-carrying mats are replaced with catless mats in the blink of an eye. Only the full-blown semantic agent can entertain the possibility that she's misperceived the world.

Mistakes provide the model for meanings. Both must reside within the subject rather than the world. It's this extra-worldly home that allows them to escape the aspect problem. But it follows from this that the semantic agent must operate as something like a repository for meanings, and that, in this capacity, she must preserve their integrity in just the same way the world preserves the integrity of its own entities. That is, her inner world must exhibit some stability; it must offer a solid surface into which a meaning might be etched. And she herself mustn't threaten that stability. This is why Kant's image of self-legislation applies here, in the semantic domain, much as it does in the moral domain. The shelter I give to my meanings must resist my own caprice as well as the world's caprice, which is to say that, to the extent that I'm a semantic agent, I must have a sense of what it would be to violate my own decree.

Suppose I coin the term 'cat' to refer to the cat on the mat, and then, upon being asked after the meaning of my term, I gesture to the mat, unaware that the cat has since vacated it. My meaning *cat* by 'cat' must be able to survive my mistake. It's not available to me to take the change in stride, as though the empty mat were my referent (worse, as though it were my referent all along). There is no protean meaning, nothing that refers alternately to *cat* and *mat* depending on how things stand at the moment (I might have a word for *cat or mat*, like 'cat-or-mat', but its meaning wouldn't be protean; I'd go wrong in applying it to a dog). It must be that my word 'cat' has a space to itself, within my own agency, from which it refers to *cat* whether or not the cat is on the mat, and whether or not I apply it correctly. Thus, Davidson explains that "[y]ou have the concept of a cat only if you

can make sense of the idea of *misapplying* the concept, of believing or judging that something is a cat which is not a cat."[129] Here again, the possibility of a mistake signals that we've entered the agential domain. And as a kind of semantic test, we can say that, if what we have is a meaning, then it could survive a mistake.

No one can enter the agential domain on her own. If I'm to store my meanings in my subjectivity, I must recognize that this is what it is, a subjectivity, i.e., a thing apart from the world.[130] This can only be accomplished by my being estranged from the world, which, in turn, requires the introduction of a novel element, a second subjectivity, one that is recognizable to me as such. The encounter with another *qua* subjectivity immediately opens an intersubjective space between us, one characterized by the project, however implicit and informal, of discerning the world together. Within that space, each expects the other to be struck by the world in the same way she is. This makes any disagreement jarring. In order to repair whatever breaches arise, each attributes a mistake to the other. But, in so doing, each allows that the world can be misdescribed, which is, again, to acknowledge that there are perspectives onto the world that are distinct from it in that they can record a state of affairs that doesn't in fact obtain. Then, in virtue of our identifying with one another, each comes to view herself as the bearer of a perspective, something that can go wrong.[131] In this way, each is made conscious of her subjectivity.

---

[129] "The Emergence of Thought", page 124.

[130] Again, no account of the origin of content can be formulated in precise terms. Probably, it's inevitable that each must venture near to nonsense in places. My metaphor isn't intended to suggest that there could be an otherworldly domain (there couldn't be), or that some species of dualism is true (I subscribe to Davidson's monism). Meanings, like any agential feature, have a worldly existence. Moreover, they supervene (albeit anomalously) on physical properties.
  But if we insist too quickly on these metaphysical facts, we're liable to neglect certain essential features of subjectivity. For one thing, no agent can recognize her agential features in their underlying physical properties, not without being alienated from them, i.e., stepping outside herself as a subject (a point made very clearly and forcefully in chapter 2 of Arthur Collins' *The Nature of Mental Things*). So, if we're to capture the nature of subjectivity, we must forget for the moment that it must fit in the same world as everything else.

[131] Related to this, John McDowell, in an account of Davidson's anomalous monism, stresses that in our recognizing another as a bearer of a perspective onto the world, we also recognize her as someone from whom we might learn about the world ("Functionalism and Anomalous Monism", page 337). In attributing agential fallibility to another, I enter into something like a shared cognitive project with her, within which she can correct my own perspective, which reveals my own fallibility.

2.3.3 Sociality and content

Only a thoroughly social animal could recognize an entity as a second subjectivity. It isn't enough to recognize that something isn't an object, i.e., that it isn't something to be assimilated into the mind-independent world. She must also recognize it as distinct from her, as not being a part of her own subjectivity. So, agential sociality must discriminate between agents. Moreover, if a self-conception *qua* subjectivity is to emerge from their exchange, the two must be so distinct as to make disagreement possible. And yet, too much difference would undermine the prospect of their intersubjectivity.[132] Also, the two must be so disposed to identify with one another that they couldn't be indifferent to a divergence in their perspectives. Actually, the requirement is likely stronger than that—divergence must disrupt them so deeply that they are willing to be estranged from the world in order to be reconciled with one another.

Of course, this estrangement carries with it the potential for a reconciliation with the world down the line. Cognitive success, as given by an accurate representation of the world, is one such reconciliation (and we must recognize that we go wrong in order to go right). Connected to this, our two subjects must be, as Myers puts it, systemically motivated by the aim to "get things right."[133] Or anyway, they must be sufficiently interested in determining what the world is like that they don't choose to dissolve their disagreements by fiat (by flipping a coin, say). It's because they're invested in *both* what the world is like and their relationships with others that they're disposed to triangulate. Content originates in the friction between these two commitments. The relevant social disposition, then, must drive joint inquiry into the world, and this must be an especially resilient joint inquiry, the

---

[132] As Davidson observes in the course of his explanation of language acquisition: "it is clear that the innate similarity responses of child and teacher—what they naturally group together—must be much alike... A condition for being a speaker is that there must be others enough like oneself" ("The Second Person", page 120).

[133] *Donald Davidson's Triangulation Argument*, page 161.

kind that persists even once differences emerge. Probably relatively few animals are social in precisely this way.

But even this highly unusual variety of sociality is insufficient. The disposition to register one another as discrete subjectivities only sets the scene. We don't address the aspect problem until the moment the two have conveyed their different conceptions of the world to one another. The trouble is that only language can communicate a perspective (or anyway, it's only in language that two perspectives can be seen to disagree), so that, in effect, only language can explain language. Triangulation offers no antecedent explanation of linguistic capacity; it must assume that capacity in order to explain it. The solution to the aspect problem, then, requires that each of the pieces emerge together: aspect, semantic agency, consciousness of subjectivity, and discourse. This is why Davidson's conception is non-reductionist—as he says, it "is not, in my opinion, possible" to "reduce the intensional to the extensional."[134]

It's also what suggests his image of a semantic triangle. The two animals and the relevant object of the world supply its three points; discourse between the animals forms the base relative to the apex, the object, of the triangle; the remaining two sides represent each animal's semantic orientation to the world. The term 'triangulation' might be taken to suggest that the base obtains prior to the other sides, but this isn't Davidson's intention. Each part of the triangle depends on the rest, so that none is prior to any other. The apex of the triangle must be present from the start, since otherwise, the two would have nothing over which to negotiate. And yet, as the aspect problem shows, the apex can't be determined until after it has been the subject of negotiation. Perhaps its initial representation is relatively coarse, and as conversation proceeds, it comes to be more and more refined. For instance, we might begin by referring to the goodness of the cat's being

---

[134] "Comments on Karlovy Vary Papers", page 293.

on the mat, then move to the goodness of her comfort in being on the mat, and then the beauty of her comfort, and then the instantiation of her nature as disclosed by the beauty of her comfort, etc.

Recognizing that her co-triangulator is a distinct subjectivity, each animal knows that her repository of meanings needn't match the other's. Still, each endeavours to enable the other to understand her. So, while one might use 'cat' to mean *cat*, and the other might use 'chat' to mean the same, each indicates what she means via public responses to their shared world, and in this way, lays the basis of a translation scheme. With the world as their Rosetta stone, each can adjust her vocabulary on the fly—e.g., swap out the term 'cat' for 'pisică'—and still be understood. Some degree of predictability is necessary if one is to be understood, but discourse can tolerate quite a lot of fluidity in meaning. And at any rate, the community of triangulators is a community of eccentrics, since, again, it's only in disagreement, and in consequent negotiation and reconciliation, that the world is brought into view. No stock of meanings, whether shared or otherwise, is required to make semantic agency possible, since it's communication, and not what we happen to communicate, that does the work.[135]

Still, the triangulators' disagreement is only genuine if its object is shared between them. The demands of triangulation align with the discursivity criterion's test: we want some assurance that the two animals are talking about the same thing. If one says "the cat is on the mat", and the other says "the cat isn't on the mat", this only generates the friction we need if they have the same cat in mind. We can always reconcile 'α is p' and 'α is $\sim p$' by introducing a second α, so as to recast the two claims as 'α_1 is p' and 'α_2 is $\sim p$'. Our interlocutors must be prevented from pursuing this line; they must be led to suspect that they're predicating p and $\sim p$ of the same α. This requires that they seek one another out in the context supplied by their object(s), which they do by asserting other

---

[135] That our semantic sociality is to be understood in terms of mutual comprehensibility, and not in terms of shared linguistic conventions, is demonstrated by the argument in Davidson's "A Nice Derangement of Epitaphs". See also Verheggen at page 87 of *Donald Davidson's Triangulation Argument*.

things, like 'I mean the tabby' and 'I mean the mat in the bathroom'. As I've said, the apex of the triangle needn't be maximally fine, for while I might, in a given moment, describe the cat in terms of her being on the mat, and my interlocutor might describe her, say, as the cat who wakes her every morning, this doesn't prevent our sharing a referent. Even a comparatively wide apex must exist in a context, i.e., as one node in a system of aspects, the sort of thing we might, and in fact do, talk about in myriad different ways. We can exploit this context, this store of shared meanings and thoughts, to find our way to the same place, even while disagreeing about what that place is like.

This iterated exchanged doesn't reintroduce the aspect problem. The idea isn't that the two discern the other's meaning by eliminating possibilities until only one remains. In a manner of speaking, the opposite is true: each expects uniformity in their two perspectives on the world, and this causes them to presume that their meanings are shared. In exploring their context together, they test this presumption. When they find a point of divergence—which will take quite some time, since each is reluctant to attribute a mistake to the other—they acquire semantic agency, and can from then on retain their own meanings, with their different aspects.

And, by that stage, we have our assurance of shared meaning: if triangulation takes place, it's because the triangulators have encountered one another within the context established by their inquiry into the apex of their triangle. Their presumption of shared meaning can't be right in every case, since they must disagree about some feature of that apex in order to be confronted with their own subjectivity. But the possibility of disagreement arises against a backdrop of enormous agreement, a sprawling network of intermeshed attitudes. This is another consequence of the holism of the mental; holism characterizes each subjectivity as a distinct mind, but also the intersubjectivity that connects those minds.

So, the triangulator's presumption of shared meaning must be largely right, since otherwise no triangulation could occur. And without triangulation, we couldn't engage in propositional

thought, let alone practical thought. This shows that the possibility of practical thought entails the possibility of shared meaning. It also suffices to satisfy the discursivity criterion, if just because triangulation is a paradigm of fruitful discourse. Thus, if follows that any conception that meets the thin formulation of the immanence constraint also meets that criterion.

But suppose we adopted austere primitivism about the idea of *cat*. If I say "here is the cat", and my partner disagrees, how could we proceed? Austerity prevents me from offering any follow-up. The best I can do is repeat my claim, but this wouldn't invite a reply, and after a time, it's liable to be heard as mere noise. This is how austerity makes fruitful discourse impossible. And, without the benefit of any exchange with a fellow thinker, I could never achieve the requisite self-awareness with respect to the idea of *cat* (again, assuming it's an austere primitive). What, then, could prevent its being absorbed into my subjectivity? That is, what could make it the case that my thinking amounts to thinking *about* the cat? And if we can't supply an 'about', what makes it thinking at all? Nothing; austere primitivism can't, in the end, secure the intelligibility of the idea of *thought about an austere primitive*. And without that intelligibility, it can't secure the possibility of such thought either.

Perhaps no austere primitivists would be impressed by this line of argument. But if any were impressed, they might propose, in response to it, that the origin of austere language is two-step: first we raise ourselves to the level of agency by triangulating on some relatively banal subject matter, like a physical entity, and them we exploit that agency—after all, once developed, we can deploy it as we like—to coin terms for austere primitives. I think this strategy may work to rescue certain austere primitives, but it can't help the normative primitivist. As I've just explained, due to the holism of the mental, there's no discerning a thought, even one about an ordinary physical entity, except against a backdrop formed by many other thoughts, and involving many other entities. It's always the case that a few normative entities are included in this set—the principles that unify the backdrop demand it. For, if my interlocutor is to understand me in my saying "the cat is on the mat", she

must have some sense of my ambitions—e.g., my wanting not to disturb the cat, or my wanting to praise her—which are themselves, in turn, only comprehensible in light of my representing some good that I hope to secure by them. We don't pursue thoughts in a vacuum; for the most part, we see something worthwhile in them, something that makes them salient, that recommends them as objects of thought. Triangulators must be able to find one another among those normative considerations, and if they're to do so, they must be able to follow one another in the deploying of concepts like *good* and *normative reason*.[136]

We saw in the last chapter that deontic entities include rational animals among their *relata*, and that this contributes to the legitimacy of the rational animal's insistence that her proximity to normativity be represented in metanormative theory. We can add to this that the rational animal's experience is saturated with normative concern, so that she can't hope to express that experience in language without reference, however indirect, to normative entities. This is a second point of intersection between her nature and normativity: the thing in virtue of which she qualifies as rational, namely, her thought, originates in her encounter with entities, some of which must be normative. Other entities that are peripheral to her life, e.g., quantum entities, needn't be present from the start. She can exercise her semantic agency with respect to them at a later stage, well after triangulation. But normative entities are like middle-sized objects—they're too close to home for thought about them to be postponed. As the basic stuff of her life, they supply the basic stuff of her language, and of her thought. Thus, we have a second place from which to motivate the rational animal's complaint: her rational nature derives from a story about content that includes the

---

[136] The point I'm making here is a modest one. Conative attitudes, as a class of attitude, are constituted by networks of normative beliefs among other things. It doesn't follow from this that it's essential to each conative attitude that it's undertaken under the 'guise of the good'. For all I've said, David Velleman may be right in holding that many conative attitudes have no such guise, and even that some are so perverse as to be undertaken under the guise of the bad (see his "The Guise of the Good"). These must fit into a wider conative context, one that *is* governed, in some sense, by normative entities, but this is compatible with their, as it were, running against the current of that context. At any rate, Myers shows that Davidson's holism of the mental can accommodate Velleman's errant cases (*Donald Davidson's Triangulation Argument*, pages 152-155).

possibility of contact between animals like her and normativity. Austerity undermines that story, and in this way, undermines her status as rational.

2.3.4 Some obstacles for the argument for the thick formulation

Davidson's triangulation argument is designed, in the first instance, to provide an account of the origin of content, and in this way, of thought and language. And given normative naturalism, it also accounts for the content that characterizes practical thought. This satisfies the thick formulation of the immanence constraint, but our focus in this section has been the immediate consequences of the thin formulation, and specifically, discursivity as essential to the possibility of thought. Discursivity is essential because of its contribution to the solution to the aspect problem, which would otherwise threaten the intelligibility of the idea of *thought*, and *a fortiori* of the idea of *practical thought*. Thus, we can only meet the thin formulation of the immanence constraint if we also meet the discursivity criterion. As I've defined it, austere primitivism can't meet the discursivity criterion; thus, it can't meet the thin formulation. The negative face of the thick formulation, the one that prohibits normative austerity, is thereby shown to follow from the thin formulation.

But then, discursivity only supplies the base of our triangle. We require, in addition to this, the triangle's worldly apex. This suggests a realist analogue to the discursivity criterion: metanormative conceptions must represent the world as housing the materials for practical thought, i.e., the candidate aspects to which our practical terms might refer. An astonishing consequence of this is that the possibility of practical thought demands normative externalism, which, in turn, demands normative naturalism. In one move, this wipes out wide swaths of the contemporary metanormative landscape. Given that this is the case, I expect that the appeal to Davidsons' theory of content will be met with considerable resistance.

And, for all I've said, the argument for the discursivity criterion—and with it, the argument from the thin formulation to the thick formulation of the immanence constraint—is vulnerable in several places. Most conspicuously, I haven't shown that the triangulation argument is the *sole* means of solving the aspect problem. It might be maintained that practical meanings can be fixed *ex nihilo*, or out of wholly subjective parts, without worldly intervention. Or perhaps normative ideas are innate, so that, again, no contact with the world is necessary. Even if it's granted that the possibility of practical thought relies on contributions from the world, it might be maintained that the world could perform the job entirely on its own. Maybe it can privilege certain of its aspects, so that they, in a manner of speaking, cry out to be enshrined in language.[137] Even if it's allowed that we need *both* a worldly contribution and an agential contribution, many suppose that the solitaire could fix her own meanings without drawing on her sociality or participating in an instance of linguistic communication.

Moreover, the discursivity criterion only connects the thin formulation to the thick formulation if it's essential to whatever it is that satisfies the criterion that it also explains the possibility of practical thought. As it happens, the triangulation argument enlists discourse in an essential explanatory role. But one might accept the discursivity criterion while refusing to accept the triangulation argument along with it—this would seem to open a bit of conceptual space within which to accommodate fruitful discourse without offering an explanation of the possibility of practical thought. One may even accept that austere primitivism is defeated by the discursivity criterion without supposing that the thick formulation has any hand in it, in which case, the negative face of the thick formulation would come apart from the thick formulation *simpliciter*. I suspect that

---

[137] This might take the shape of something like David Lewis' 'reference magnets'; see his "Putnam's Paradox".

discursivity can't in the end be disentangled from explanation, but I'm not prepared to defend this claim here.[138]

My ambition is only to show that the discursivity criterion is attractive, rather than to compel its acceptance. Up to this point, I've done little more than elaborate it. My hope, though, is that there's something persuasive in the picture to which it belongs, and that this translates to something persuasive in it as an element of that picture. As Rawls says, "justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view."[139] It's natural to suppose that sociality enters into the origin of content somewhere. The aspect problem and its solution in the triangulation argument reinforce this supposition. They indicate that any conception of *meaning* that fails to assign an essential role to discursivity risks offering too thin a basis for thought and language.

In any case, as I've said several times, the discursivity criterion figures in the argument for the immanence constraint as a bridge between the thin and thick formulations, along which the force of the former—which is, as I've all the while maintained, substantial—is conveyed to the latter. I will proceed under the assumption that it succeeds in this. It's true that, in order to complete the argument, I'd have to exhaust every avenue, and demonstrate that only one, the avenue I've chosen, is workable. While this can be done, it'd require a more elaborate account of Davidson's work, together with its metanormative consequences, and this would take us too far afield.[140] It's worth observing that, if, instead of rigour, the relevant measure of argumentative

---

[138] Perhaps the more pressing question is this: is the triangulation argument an indispensable feature of the argument for the immanence constraint? The argument for the immanence constraint depends on an account of the origin of content. I must select what is, by my lights, the best available account. This is the triangulation argument, so, under that guise, it's indispensable.
 The issue is complicated by the fact that the positive face of the immanence constraint demands that metanormative conceptions include a theory of practical content. I am, at this moment, developing a metanormative conception, one that, I take it, answers to that demand. The triangulation argument is indispensable, then, in this second place (if it's distinct from the first), namely, that it satisfies that feature of the positive face.

[139] *A Theory of Justice*, page 19. I'll have more to say about this form of justification in chapter 4.

[140] Anyway, the work has already been done in *Donald Davidson's Triangulation Argument*.

success is rhetorical power, the negative face is better served by the expositions of primitivism and austerity in sections 2.1 and 2.2. In the first, I argued that primitivism *simpliciter* can offer everything that might have tempted us to consider austere primitivism in the first place, and in the second, I demonstrated that austerity, being committed to highly peculiar notions like radical intuitionism, is rather unwieldy. My guess is that, if enthusiasm for austerity *can* be extinguished, it was extinguished there, in the sober presentation of what austerity really is, and of how far primitivism can get without it.

The virtue of the defence of the thick formulation offered under the aegis of the discursivity criterion isn't its rigour or its rhetorical power. Its virtue is its fecundity, its proposing a course of reasoning that can give full expression to the rational animal's complaint. As I've formulated it, this complaint is raised against metanormative theory insofar as it fails to include any reference to the rational animal's nature in its determination of normativity. The thin formulation establishes that the rational animal has a legitimate claim to a stretch of the normative domain, but this is all it does—it erects no significant obstacle for theory to clear. But if we follow the path from the possibility of practical thought to the possibility of practical discourse, we are thereby introduced to materials with which to elaborate the idea of *rational animal* so as to impose substantive constraints on the design of metanormative theory. That is, by way of the discursivity criterion, we can see that a complete picture of the normative domain assigns an essential place to the discursive animal, with all that that entails. Thus, the thin formulation is thickened, and a framework for the articulation of the rational animal's complaint falls into place.

2.4 THE POSITIVE FACE OF THE THICK FORMULATION

Our starting point in this chapter was the rational animal in her most skeletal form, an entity in some sense capable of practical thought, with no commitments beyond that; she needn't have any sort of

life, beyond her simply being there, with the requisite capacity. We've since acknowledged that we aren't after all free to stipulate the possibility of such an animal as an isolated point in logical space. Rather, this possibility comes tethered to an expansive constellation of other things. And so we begin to wrap some flesh around the bone. That is, we show the relevant idea of *rational animal*, the one for which metanormative theory is responsible, to be substantial: the capacity for practical thought necessarily belongs to a network of other capacities—thought requires communication, and this requires sociality, language, and awareness of one's self and others as subjectivities.

Thought also requires causal receptivity to the world, and more concretely, a body equipped with a sophisticated receptive apparatus, something like perception. No mere sensory mechanism will do, since this can only register worldly input as though through a veil, i.e., as the imprint of some noumenal entity. If the world is to supply matter for thought, it must supply it to an animal who can understand herself and her causes as together being in the world, which is to say that she must represent herself and her causes as public, the sort of things that are jointly locatable by triangulators within their common environment. What's needed, then, is bodies able to represent themselves and their objects within space, alongside other bodies.[141] Also, while the possibility of practical thought requires that the rational animal have a will, the possibility of discourse requires that she use it—she must have a rich supply of conations, and she must command her share of causal power, something she can deploy in pursuit of the satisfaction of those conations. Probably, managing those conations (not to mention, navigating her social dispositions) demands a system of

---

[141] At pages 129-130 of "The Emergence of Thought", Davidson' lists two ambiguities in the prospective worldly causes of our thoughts: the aspect problem, which we've covered, and the distance problem, i.e., "whether [the cause] is proximal (at the skin, say) or distal." I follow Verheggen in representing the aspect problem as sufficient to prompt the considerations that lead to the triangulation argument. Moreover, a solution to the distance problem can't by itself rescue semantic externalism. For these reasons, Davidson's inclusion of the distance problem alongside the aspect problem is somewhat misleading (anyway, many have been misled). Still, I think there's something evocative in this second problem. I gather that it's easily solved by the perceptual systems of most animals, but this just underscores what's remarkable about those perceptual systems. After all, no machine could distinguish between phenomena at its sensory surfaces and phenomena out in the world.

affective capacities as well. Among other things, she must be equipped with a reliable means of aligning her conations with her normative judgments.

More might be said. And the more that can be wrung from the possibility of discourse, the more we have to direct us in the development of metanormative theory. I think the discursivity criterion commits such theorizing to at least these three bodies of inquiry: asking after what makes practical thought possible, asking what its object is, and asking what its subject is. These correspond to i) a theory of practical content, ii) a first-order theory of practical reasons, and iii) a theory of agency, respectively. This suggests a formulation for the positive face of the thick formulation: *if a normative conception is to succeed in capturing its corresponding idea, it must advance a theory of practical content, a first-order theory of practical reasons, and a theory of agency.* Of course, success also requires that these theories be true, but the rational animal is satisfied—for now, at least—if metanormative theorists acknowledge that it's incumbent on them to include an attempt at an account of these three things in their comprehensive conceptions of *normativity.* I'll say a bit about each in what remains of this chapter.

i) As we've just seen, it's an entailment of the thin formulation of the immanence constraint that metanormative theory is responsible for a theory of normative content. This is because the possibility of practical thought is the possibility of thought with determinate practical intentionality, and the story of how thought acquires determinate practical intentionality is not so simple as to be the sort of thing we might pass over. It involves, among other things, an answer to the aspect problem, so that, in order to establish the possibility of practical thought, we must identify some means of latching thought on to particular practical entities. And the task of identifying this means falls to theory. Davidson's triangulation argument identifies one such means; I believe it identifies

the correct one, though in principle the positive face of the thick formulation can be satisfied by some other story, provided that it meets the discursivity criterion.

ii) Any reasonably comprehensive metanormative conception treats the subject of normative reasons from the second-order, but for the most part it's supposed that metanormative theory isn't answerable to normative theory, so that the second-order is autonomous with respect to the first-order, and the latter is subordinate to the former. It's also supposed that it's available to metanormative theory to deny the possibility of first-order theorizing, for instance, by denying that normative claims can be collected in such a way as to form a system, the likes of which we might undertake to map via theory. A commitment to austerity at the second-order must drive us to such a theory-pessimism, since the austere primitivist is saddled with radical intuitionism, from which it follows that she must conceive of each normative claim as wholly individual, discretely removed from each other claim. With no relations to follow from one claim to another, she can't identify any material for theory-building, and so she must give it up. In the metanormative case, this means abandoning moral and ethical theory, aesthetic theory, prudential theory, etc., and holding of the entire history of normative theorizing that it's the product of a confusion (or anyway, that it isn't what it purports to be).

But it's a consequence of the argument for the discursivity criterion that each normative claim can only arise against a vast backdrop of claims, since each is constituted by its position among countless intertwining relationships. Thus, it's in the nature of practical thought that it inhabits an environment, i.e., that it belongs to a system. And where there's a system, there's the project of mapping it. Due to the holism of the mental, the rational animal must navigate this system, at least to a degree, if she's to exercise her capacity to communicate with others. To this extent, the system underlies the possibility of practical thought, so that determinations of that possibility must extend

to the system as well. And, if in the course of her engaging with her interlocutor, the rational animal

invariably encounters reasons to esteem her fellows, then these reasons contribute to the

constitution of the possibility of practical thought, belonging as they do to the context from out of

which such thought arises.[142] And these reasons likely indicate others, for instance, a reason to

esteem herself, and more broadly, a reason to esteem persons as agents engaged in the shared

project of discerning and acting from normative reality. It's informative to collect these reasons in

one place, and to depict their relationships with one another by articulating them in the form of

general principles. But this is just what it is to compose a first-order theory.

This is all to say that an account of the possibility of practical thought must include an

account of what that thought is about. An account amounts to a theory if it exhibits systematicity,

and an account ought to exhibit systematicity if its object is marked by a system of relations. We've

already established that normative entities, when represented in thought, are marked by a system of

relations. So, we're entitled to demand of accounts of those entities that they be theories.

iii) Most important of all, if we're to address the rational animal's plight, we need a theory of agency,

one that begins from the source of her plight, the normative domain. The subject of this theory is

the animal as marked out by her place within the reason-relation. Her nature, as revealed by that

place, is essentially necessitated, i.e., essentially answerable to normative demands. She's causally

powerful, but she conceives of that power as being at the disposal of her reasons. She is, in a

manner of speaking, the vehicle of reason, its representative in the physical world.[143] But her power

---

[142] This is only an example. It may be that no consideration is made inevitable by the context in which communication takes place—indeed, this seems to be Davidson's view (as Myers indicates at page 121 of *Donald Davidson's Triangulation Argument*.). I suspect, though, that we can exploit certain essential features of discourse so as to identify a moral centre of the normative domain, one that can be captured by a contractualist moral theory. At any rate, I follow Myers in matching Davidson's metanormative conception with contractualism (see *ibid.*, pages 185-190).

[143] Crispin Wright complains in "Truth and Ethics" that the scope of normative causality is confined to one remarkably narrow corner of the universe, the part that determines animal behaviour by figuring in their thoughts. Once we recognize that this is the case, we might be tempted by an analogy with fictional entities, which are also 'causal' in only

is limited, not just physically, but agentially, since she can't ensure that its deployment is in every respect good. This is a consequence of her fallibility—she's to some extent obedient, and to some extent disobedient, to normative demands. Agential fallibility is itself an essential feature of necessitation. Or anyway, such is Kant's view:

> if the will is not *in itself* completely in conformity with reason (as is actually the case with human beings), then actions that are cognized as objectively necessary are subjectively contingent, and the determination of such a will in conformity with objective laws is *necessitation*: that is to say, the relation of objective laws to a will that is not thoroughly good is represented as the determination of the will of a rational being through grounds of reason, indeed, but grounds to which this will is not by its nature necessarily obedient.[144]

Though it's essential to the agent that she's fallible, she's nevertheless estranged from her mistakes. Some of her mistakes are owed to simple ignorance, but some are owed to something deeper, and in this second case she (to cite an appropriately perplexing Davidsonian insight) "recognizes, in [her] own intentional behaviour, something essentially surd."[145]

The rational animal doesn't belong to any biological species, not insofar as she occupies the agent-place of the reason-relation. Michael Thompson distinguishes between Aristotelian and Kantian traditions by portraying the first as captivated by the idea of *human*, and the second as captivated by the idea of *rational animal*, which might, as far as we know, be instantiated by Martians and Venusians—or dolphins and bonobos, for that matter—anything that exhibits the requisite powers.[146] The theory of agency required to satisfy the positive face of the thick formulation must

---

this way (i.e., they command no genuine causal power at all). But because the relation of normativity to the rational animal is essential to the nature of both, we oughtn't to be disturbed by the fact that the causal power of the one (and the causal receptivity of the other *qua* rational) is only manifested in this one place. In any case, the analogy with fictional entities can't be sustained, because there, the relation isn't essential—it doesn't belong to the idea of *unicorn* that it's thought, but it does belong to the idea of *normative reason* that it's thought; it doesn't belong to the idea of *rational animal* that it involves thoughts about unicorns, but it does belong to that idea that it involves thoughts about reasons.
[144] *Groundwork*, page 413. This feature of the rational animal *qua* necessitated is an important starting point for a theory of agency. It's a central feature of agency that agents go wrong.
[145] "How Is Weakness of the Will Possible?", page 42.
[146] *Life and Action*, page 7.

attend to this same Kantian subject matter. It promises to contribute something to a theory of human nature, but only because necessitation is so central an element of human life.[147]

The error of radical intuitionism is its supposing that we can achieve practical thought without any intermediary. But once we introduce an intermediary, we find the thinker in a particular place, alongside that intermediary—now she's conditioned by a worldly cause, now she's conditioned by another thinker, and the like. This tells us about her. And it follows that the agent-place of the reason-relation isn't simply a variable, and that a full account of the agent-place can't be entirely formal. It must include an investigation into what it is, substantively speaking, to enter into the agent-place, and what it is, substantively speaking, to be an agent.

I won't do more to specify the kind of theory of agency the positive face demands. I suspect the best way to do so is to provide an example, and I won't attempt that here. My ambition all this while was to extract substantive constraints on metanormative theory from the fact that normative reasons are essentially connected to rational animals via the agent-place of the reason-relation. Here we have three such substantive constraints, three areas in which philosophers must contrive to say something, since otherwise they're vulnerable to the rational animal's complaint. And, thus, the stretch of the normative domain to which the rational animal is entitled is revealed to be quite large. She isn't admitted simply to perform a formal role. Rather, she has a kind of home there.

---

[147] There's a vaguely Hegelian complaint to the effect that the idea of *rational animal* is too slim a basis for ethical theory, and that we ought instead to avail ourselves of the comparatively thick idea of *human*. This complaint had seen a resurgence in the 1980s with works like Alasdair MacIntyre's *After Virtue* and Williams' *Ethics and the Limits of Philosophy*, and has lingered with us since. I suspect that it's made to appear more persuasive than it really is because the idea of *rational animal* has been occluded by metanormative austerity. Once we renounce austerity, and attend to the positive face of the immanence constraint, we are given these three ways to thicken the idea of *rational animal* (and all without violating the transcendence constraint, as we would do if we derived an ethics from anthropology). The theory of agency does the most work—it shows that the rational agent isn't some lone self-determining subjectivity, like a Sartrean hero; the agent is that which straddles the normative and physical domains, and intervenes in the latter at the direction of the former.

3. INTERNALISM, CONSTRUCTIVISM, AND CONTRACTUALISM

I've tried to express the rational animal's complaint in terms of the immanence constraint. But it isn't clear to what extent the immanence constraint captures the whole of the complaint. Addressing this point is difficult—if we're to assess the constraint for its capacity to articulate the complaint, we'll need an independent articulation of the latter, and I confess that I can't see how to articulate the complaint without simply restating the constraint. Still, the two are, in my view, distinct. I suspect the best way forward is to appeal to a theory of agency of the kind demanded by the positive face of the thick formulation, since this frames the rational animal's plight, and, to that extent, offers some indication as to what prompts her complaint. Such a theory would provide a context against which we might ask whether the immanence constraint succeeds in channeling the complaint, and whether it exhausts the demands that the rational animal is entitled to make on metanormative thought.

But I don't have room to develop a theory of agency. Instead, I'll turn to some positions and principles in metanormative theory and moral theory that might reasonably be offered as materials for elucidating the rational animal's plight. The positions I mean are Williams' reasons-internalism, practical internalism broadly conceived, Korsgaard's constructivism, and moral and political contractualism. Apart from Korsgaard's constructivism, each position's fit with the rational animal's complaint is loose. Still, each has a lesson to offer that's at least adjacent to the complaint. The immanence constraint doesn't collect all of these lessons in one place; it only succeeds, to the extent that it does, in capturing a part of the complaint. I'm not prepared to advance any substantive conclusions about the prospects of a precise formulation of the complaint, or about the degree to which the immanence constraint must be supplemented if it's to convey the whole of the limit placed on metanormative theory by the nature of agency. This chapter is for the most part

exploratory—it's as much about situating the immanence constraint within a philosophical tradition as it is about fleshing out the content of the rational animal's complaint.

Before I begin, though, I'll say a bit about an obvious point of mismatch between the constraint and the complaint. The immanence constraint is a constraint on metanormative theory, but the rational animal's complaint isn't raised against theory, not in the first instance. Of course, it isn't raised against normativity either—after all, the normative domain must be largely as it is if there's to be a rational animal who can complain. What offends the rational animal are representations of the normative domain, the ones that neglect her position in it. There are metanormative theories among the offending representations, and the immanence constraint aligns with the rational animal's complaint in objecting to those theories. But the complaint isn't so parochial as to be directed against theorists, i.e., the self-appointed authorities who inhabit the academy, and who compose, among other things, dissertations on the subject of normativity. If it were, it'd be much less interesting than it is. And anyway, it's always available to the rational animal to ignore what academics have to say about her. In fact, for the most part, that's what she does.

It isn't available to her to ignore philosophy itself—or anyway, not provided that we distinguish between two philosophical endeavours, a wider one and a narrower one, the first an inevitable manifestation of rational life, and the second a specific instance of the first, namely, the philosophical project as carried out by professionals in the context of the university. Every rational animal is a philosopher in the wide sense, if just because it's in the nature of rationality that it encounters philosophical questions and motivates the effort to answer them.[148] She may, for all that,

---

[148] I'm not prepared to defend this claim. I take it, though, that it has strong inductive support, at least provided that our criterion for philosophizing isn't too strict. The layperson's philosophy ought to be construed on something like the model of the layperson's mathematics. Certainly, any human life involves a good deal of counting, calculation, proportioning, sequencing, and the like. Not all of it is connected to material concerns; much of it is, but much of it isn't. Lay mathematics is informed by a public mathematical culture, and that culture is, in turn, sensitive, in a way far too complex to map, to academic mathematics. Granted, there are important differences: lay philosophy isn't quite so prosaic as lay mathematics, extra-philosophical concerns exert greater control over philosophy—both in the wild and in

be estranged from, even repelled by, the narrow philosophy pursued in graduate seminars. But then, given her commitment to wide philosophy, she can't escape narrow philosophy entirely; her philosophizing is informed by a culture, and while for the most part it's the graduate seminar that answers to the caprices of the public philosophical culture, it does happen from time to time that it's the culture that answers, however obliquely, to the seminar. A. J. Ayer's book *Language, Truth, and Logic* is a good candidate for a piece of academic philosophy that spilled over into the wider philosophical world. Its publication explains, more than anything else, how the public culture landed on the notion that the philosophical landscape is organized around a kind of crossroads, with one path leading to uncritical superstition and the other leading to something like Ayer's caricature of logical positivism, with its boo-hurrah non-cognitivism.[149] It's clear, I think, that Ayer's cultural legacy has been by and large inhospitable to the idea of *rational animal*. It follows that the rational animal *qua* wide philosopher ought to attend to *some* narrow philosophy so as to guard against theorists like Ayer.

This relationship between wide and narrow philosophy is what forgives the mismatch between complaint and constraint. It also suggests a strategy. The complaint targets an aspect of the public philosophical culture, the aspect that insists on an impoverished conception of *agency*. The constraint has a narrower target, but this what enables it to include, in addition to a diagnosis of the problem, a solution to that problem. I don't think it's unreasonable to suppose that an effective

---

the academy—than extra-mathematical concerns do over mathematics, etc. All the same, the analogy serves well enough to delimit the philosophy that belongs to the rational animal taken just as such.

[149] The relationship between logical positivism and the zeitgeist of the interwar years is, of course, very complicated. There's a case to be made that, given the abject state of moral, political and theological discourse at that time, the German- (and English-) speaking world was eager to turn its attentions to the hard sciences instead. Related to this, it may be that a period of non-cognitivism was necessary to rescue normative thought from itself. If this is right, Ayer is better represented as giving a voice to an existing cultural force, rather than charting a new course for culture to take.

  I suspect there's something to this historical interpretation. Still, in selecting Ayer and his legacy as a representative of the positivistic cultural force, we came to be saddled with a bundle of philosophical dispositions that are, as we've since learned, remarkably difficult to shake. Hostility to normativity, *qua* alien to the scientific worldview, remains ubiquitous despite its having no tenable justification whatsoever. Ayer's confidence (which was, I think it's fair to say, dishonest, given the scope of his arguments) contributes to an explanation of the resilience of that hostility, even if he arrived too late to be the origin of scientism in our culture.

(though not the *most* effective) means of correcting a deficiency in the culture is academic intervention. If the public culture is dominated by views that elicit the rational animal's complaint, one way to accommodate the rational animal is to displace the academic analogues of those views in the hopes that this will translate to a shift in the wider culture. It seems to me that something like this same dynamic arises when Korsgaard identifies the 'why should I be moral?' question as a call for philosophy.[150] That call is registered at the level of the wider philosophical project, the one shared by all rational animals. Still, *The Sources of Normativity* scrutinizes the answers to that question offered from within the university, from Hobbes to Kant.[151] That's as it should be—it's incumbent on the university to address the moral poverty of the public culture, but it needn't (and often shouldn't) do so directly. Its best tool is the one it's honed over centuries, namely, academic dialectic.

Perhaps this line of response is unpersuasive. If so, I have this fallback: each of the formulations of the immanence constraint can be restated as descriptions of formal features of the normative domain, rather than as success conditions for theory. So, for instance, the thin-narrow formulation is the metatheoretical equivalent of this metaphysical claim: *the intelligibility of the idea of* normative reason *depends on the possibility of practical thought about normative reasons*; the thick-wide formulation is the equivalent of this claim: *the intelligibility of a normative idea depends on that which explains the possibility of thought about that idea.* (I'll rely on this straightforward translation scheme between metatheoretical and metaphysical claims to relate the immanence constraint to each of internalism, constructivism and contractualism, since these engage with the objects of metanormative theory, rather than with metanormative theory itself.) If we doubt that the immanence constraint can be of any use to the rational animal as a cultural counterbalance to the views that offend her, we can draw

---

[150] Again, at page 9 of *The Sources of Normativity*. See note 3.
[151] Granted, Hobbes wanted nothing to do with the university. But you know what I mean.

on this translation scheme to make the immanence constraint relevant in a different way, namely, as a metaphysical claim that the rational animal should consider in the course of her philosophizing about her own nature.

3.1 WILLIAMS' INTERNALISM

As Korsgaard explains, in the context of normative psychology, "[a]n *internalist* theory is a theory according to which the knowledge (or the truth of the acceptance) of a [normative] judgment implies the existence of a motive (not necessarily overriding) for acting on that judgment."[152] The relation of implication running from judgment to motive is weak—it sustains the *expectation* that a given agent's judgment that she ought to φ is accompanied, according to the appropriate rational relationship, by a motive to φ, but it doesn't *guarantee* that motive.[153] That said, the internalist supposes that a failure to be motivated in line with one's judgments demands an explanation of a special kind, one deriving from *akrasia* or the conviction that one's judgments are wrong, i.e., from ruptures in one's rationality. This provides the material for a different kind of guarantee: according to internalism, each normative judgment entails a disjunction—either the judger is (defeasibly) motivated or she exhibits a rupture. This is what enables Korsgaard to formulate this internalism requirement: "[p]ractical-reason claims, if they are really to present us with reasons for action, must

---

[152] "Skepticism about Practical Reason", page 315. I've modified Korsgaard's definition so as to cover normative internalism generally, rather than specifically moral internalism.

[153] Internalism only applies to normative judgments that pertain to the agency of the judger. Judgements to the effect that another person ought to φ don't ground any expectation that the judger is motivated to φ. Still, internalists may hold that vicarious judgments are capable of commanding counterfactual motivational force, so that the judgement that an agent *x* ought to φ would entail something like the judgment that I ought to φ were I to be *x* (agent-relative reasons complicate things; see note 10). This closes a gap that would otherwise open between reasons that one takes to bear on one's own conduct, and reasons that one takes to bear on others' conduct, even hypothetical or imagined others.

  Normativity prescribes to rational nature as such. Actions are prescribed under certain conditions, and to that extent we're each issued different directions, but these conditions appear *within* necessitating entities, rather than outside them—as in '*r*(if *p*, then *q*)', rather than 'if *p*, then *r*(*q*)'—so that we all encounter the same reasons. If, for instance, I recognize a reason that prescribes action from those with a special ability that I lack, rationality doesn't compel me to perform the action since the relevant conditions aren't met in my case. Still, if I'm rational, I'll feel the force of that reason's authority in the sense of being disposed to conform to it should I acquire the relevant ability.

be capable of motivating rational persons."[154] Her use of the word 'rational' conveys the disjunction—either a normative judgment is assigned the motivational force that it's due or its judger is to that extent irrational.[155]

I'll return to Korsgaard's internalism requirement in the next section. For now, my focus is the Humean branch of normative internalism, with Williams as its exemplar. The Humean branch subordinates normative reasons to the idiosyncratic features of individual agents' conative psychologies, so that a normative reason for a given agent to φ is necessarily predicated both on some antecedent conative entity, like a desire, which is lodged within that agent, and on an instrumental relation, according to which her φ-ing might fulfil, or else promote the fulfilment, of that entity. I call the branch 'Humean' because it emerges nearly fully-formed from Hume's *A Treatise of Human Nature*, book II, part III, section III.[156] This position is quite a lot more ambitious than the one Korsgaard describes. For all her requirement says, it might be that it's our capacity to judge that a reason obtains that's constrained by our receptivity to that reason as a motivating force, so that the practicality of the reason is primary, and our psychology is shaped by it.[157] But in the

---

[154] *Ibid.*, page 317.

[155] What about the example from Myers discussed in section 1.4, the one of a person who's disposed to misjudge in normative matters, and is aware of this fact? It may be that it's rational, in some sense of other, for him *not* to assign any motivational force to his normative judgments. If so, we might prefer the 'rupture' formulation of the disjunction to Korsgaard's 'rationality' formulation. My suspicion, though, is that Korsgaard would classify this person as at least somewhat irrational, given his conative disunity (see *Self-Constitution*), so that, in the end, the 'rupture' formulation is equivalent to hers.

[156] Williams' allegiance to Hume is announced at page 102 of "Internal and External Reasons". Smith's Humean theory of motivation and Street's Humean constructivism belong to this same branch. But because Hume denies that reason can be practical, he isn't himself an internalist, whether Humean or otherwise.

[157] This is the form that Nagel's internalism takes. Though he opens *The Possibility of Altruism* by saying "I conceive of ethics as a branch of psychology", what he means is that advances made at the level of ethical theory are to be registered as having consequences for psychology, and not the other way around (he goes on to say that "the view presented here is opposed not only to ethical relativism but to any demand that the claims of ethics appeal to our interests: either self-interest or the interest we may happen to take in other things and other persons").

Even Jonathan Dancy's pure cognitivism, which dispenses with desires entirely, is internalist by Korsgaard's criterion—because its belief-belief dyads are represented as themselves comprising motivation states, the view's moral psychology preserves the internalist connection between judgment (*qua* belief) and motivation (see *Moral Reasons*). This underscores the fact that normative internalism isn't the exclusive possession of the Humean metanormative tradition. (Perhaps no pair of terms invites more confusion than 'internalism' and 'externalism'. Many avowed 'externalists', like McDowell ("Might There Be External Reasons?") and Scanlon (*What We Owe to Each Other*, page 373), are to be counted internalists by Korsgaard's definition. What their use of the term 'externalism' signals is their opposition to Williams' internalism.)

Humean case, it's the capacity for motivation, depicted as in some sense psychologically fixed and prior to practical deliberation, that operates as a constraint on the possibility of a reason's obtaining. As Scanlon puts it, the idea is that "all reasons for actions have subjective conditions".[158] The immanence constraint has no affinity to this tradition. While the constraint presupposes that normative reasons are agent-relational, it doesn't envision that relation as involving anything along the lines of the Humean subordination of reason to conations (and for that matter, to passions).

This isn't to deny that Humean internalism is one way of meeting the constraint. In fact, read one way, Williams' internalist criterion—"[i]f 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"[159]—is just the thin-narrow formulation of the immanence constraint. Translated into metaphysical terms, that formulation states that each normative reason is potentially thinkable in such a way as to sustain action from it; this is equivalent to the claim that the practicality of a normative reason entails that that reason has the potential to determine an action by being 'someone's reason for acting' (i.e., her operative reason). The following passage makes it clear that the variety of explanation Williams wants must pass through an *action from* relation:

> [n]o doubt there are some cases of an agent's φ-ing because he believes that there is a reason for him to φ, while he does not have any belief about what that reason is. They would be cases of his relying on some authority whom he trusts, or, again, of his recalling that he did know of some reason for his φ-ing, but his not being able to remember what it was. In these respects, reasons for action are like reasons for belief. But, as with reasons for belief, they are evidently secondary cases. The basic case must be that in which A φ's, not because he believes only that there is some reason or other for him to φ, but because he believes of some determinate consideration that it constitutes a reason for him to φ.[160]

The extensional reading of the reason-relation described in subsection 1.3.1 might serve for secondary cases, but in the basic case, what's required is that normative explanations proceed via

---

[158] *What We Owe to Each Other*, page 363.
[159] "Internal and External Reasons", page 106.
[160] *Ibid.*, page 107.

practical thought. This is connects internalism to necessitation—the necessitated agent must be able to act from the entity that necessitates her.

These passages reflect Williams' commitment to internalism *simpliciter*. Trouble emerges once we turn to the idiosyncratic features of his version of internalism. I have two objections in mind: first, Williams tethers his internalist criterion to a remarkably narrow—and, I suspect, unworkable—conception of *action from*. This is a consequence of his adherence to a quasi-Humean conative psychology, according to which each agent has a store of conative items that are propositionally individuated prior to practical deliberation, and that function as the origins of all *action from* relations. I don't see how any successful theory of practical content can be made to cohere with this picture. The aspect problem can only be solved in practice, that is, through the use of contents in thought and speech. It follows from this that there's no determining conative content prior to practical deliberation.[161] Thus, the quasi-Humean view that practical deliberation must begin from an antecedent conative aspect is self-defeating.

The second objection is that Williams' Humeanism straightforwardly violates the transcendence constraint. If a conception is to meet that constraint, it must offer a perspective from which to assess the whole of our agential nature. Our conations belong to that nature, both as individual attitudes and, crucially, taken together as a system. Williams can only permit the assessment of a conation in light of another conation; he can't allow a judgment as to the goodness or badness of an agent's entire conative disposition.[162] Thus, he must embrace the profoundly

---

[161] Note that Williams is explicit in opting for a capacious conception of *practical deliberation*: "[t]here is an essential indeterminacy in what can be counted a rational deliberative process. Practical reasoning is a heuristic process, and an imaginative one, and there are no fixed boundaries on the continuum from rational thought to inspiration and conversion" (*ibid.*, page 110). If deliberation includes all that, and if conative content is, as Williams maintains, prior to deliberation, then we're left with very little with which to solve the conative aspect problem. (Genuinely Humean conative conceptions deny that conations are propositional (or anyway, they're not propositional in such a way as to enter into deliberation), so that no conative aspect problem arises.)

[162] Or rather, what he can't allow is that judgments framed in axiatic language have deontic counterparts that can be expressed in terms of reasons. He's ready to judge a person's conative temperament to be good or bad in various ways, but only on the condition that that judgment isn't taken to in any way necessitate that person to some course of action.

counter-intuitive conclusion that our reasons are, in a manner of speaking, held hostage by what we

happen to want. Consequently, the question "what ought I to want?" must, on his view, be

unintelligible (unless we'd accept the answer "you ought to want (some part of) what you already

want"), so that he can't address a version of Korsgaard's normative question that's reformulated for

desires, as in 'what justifies the claims that desire makes on us?'. The result is even more

embarrassing at the first-order, since, on this conception, a person who has no desire to take life-

saving medicine may have no reason to do so.[163]

This is what drove Parfit to his reading of Williams' internalism according to which Williams

simply fails to grasp the idea of *normative reason*.[164] My impression is that Williams understands the

idea perfectly well—his neglect for the transcendence constraint is better construed, I think, as a

symptom of his hostility to the suggestion that deontic entities could direct us to anything more than

the organization or specification of the projects that, in some rudimentary sense, we already have.

The source of that hostility is a suspicion of categorical normativity, especially as it figures in moral

theory.[165] Of course, I think this suspicion is mislaid. In fact, I can't fathom what could motivate it.

So, for instance, he says the following about a man who mistreats his wife: "[t]here are many things I can say about or to this man: that he is ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal, and many other disadvantageous things. I shall presumably say, whatever else I say, that it would be better if he were nicer to her" ("Internal Reasons and the Obscurity of Blame", page 39). But he denies that any of these claims entail that this man has a reason to be nicer to his wife. This invites the questions: what can Williams make of the relationship between axiatic and deontic entities, and of the unity of the normative domain? And how can he convey the practicality of axiatic entities? Even bracketing these questions, the transcendence constraint requires that assessments of our agential nature have the capacity to necessitate in precisely the way Williams denies. That is, the willingness to apply axiatic concepts isn't enough, not unless they have the appropriate deontic counterparts.

[163] A conclusion that, in a remarkable show of candour, he acknowledges (see "Internal and External Reasons", pages 105-106).

[164] See note 21.

[165] See chapter 10 of *Ethics and the Limits of Philosophy*. Oversimplifying quite a lot, we might say that Williams distinguishes the ethical from the moral by taking ethics to frame in axiatic terms what morality (attempts to) frame in deontic terms. Given his preference, in this context, for the axiatic over the deontic, he recommends that we pursue ethics, and abandon (or anyway, significantly restrain) morality. (Whether or not this is the distinction Williams has in mind, I've been using the terms 'ethics' and 'morality' in roughly this same way, that is, by drawing on the differences between the two normative vocabularies, the axiatic and the deontic. As I see it, ethics deals in a subset of axiatic entities, the ones pertaining to the agential good, while morality deals in a subset of deontic entities, the ones pertaining to the agential orientation to the good, as well as agents' orientations to one another and perhaps also to non-agential animals.)

The best I can manage is this: much moral theory in the twentieth century had had to operate within the confines set by metanormative austerity. The result was often insubstantial, sometimes to the point of vacuousness. It likely appeared, at that time, as though the problem lay with a preoccupation with categoricity, or with the privileging of the deontic over the axiatic.[166] But the problem was at the metanormative rather than normative level, and its remedy is to resist austerity by attending to the positive face of the immanence constraint.

3.2 INTERNALISM *SIMPLICITER*

There's an insight worth preserving in Williams' conception, the one expressed by his internalist criterion. As I understand it, this criterion is uncontaminated by Williams' Humeanism. In fact, I suspect it can be extracted from his conception wholly intact. I suggested that it might be read as the equivalent of the thin-narrow formulation of the immanence constraint—while this is true, it's a little misleading, since internalism is in the first instance a theory in normative psychology. As Nagel says, "[t]he names 'internalism' and externalism' have been used to designate two views of the relation between [normativity] and motivation."[167] Granted, Williams' reasons-internalism shifts the topic somewhat by focusing on normative reasons themselves, but his concern is the nature of normative reasons according to which they can be fitted into a relation with motivation.

Of course, the immanence constraint is itself animated by the question of how agents can be fitted into reason-relations, so that the reason-relation and the internalist relation deal in many of the same considerations. Still, they're distinct. The reason-relation connects genuine normativity to practical thought, which is thought that encompasses both cognition and conation. The internalist

---

[166] This line of complaint appears quite a lot earlier than *Ethics and the Limits of Philosophy*. Anscombe's "Modern Moral Philosophy" is the likely the most famous formulation. Cf. note 144.
[167] *The Possibility of Altruism*, page 7. I've modified Nagel's claim so as to cover normative internalism generally, rather than specifically ethical internalism.

relation only connects the representation of normativity (e.g., an operative reason) to conation, or more specifically, motivation.[168] These *relata* are intimately connected—for one thing, as we've seen, the representation of normativity is to be explained in terms of genuine normativity;[169] for another, the internalist relation is designed to chart an explanatory path for motivation *from* a judgment, rather than mere conformity to the content of that judgment, so that the ambition, in effect, is to unify practical cognition and conation in practical thought. But even with this overlap, the two are different enough to carve out different projects.

Still, the immanence constraint isn't indifferent to how the possibility of practical thought ties judgment to motivation. Among other things, it requires materials for *action from* relations. It's difficult to envision a means of securing those materials that doesn't include internalism. And, as it happens, none of the representative externalists, namely, J. S. Mill, Moore, Prichard and W. D. Ross,[170] can satisfy the immanence constraint. So, here again, the immanence constraint and internalism converge; they find fault (though not necessarily the same fault) with the same views.

That suffices as an account of the connection between the two. Now, we can examine the contribution that internalism might make to the project of addressing the rational animal's complaint. By specifying features of practical thought, internalism offers an avenue for the development of a theory of agency. It does this by identifying the role of motivation in the representation of normative entities. As a prospective agent encounters a normative entity, and represents it in thought, she must, absent *akrasia* or any similar impediment, be driven thereby to act as that entity prescribes. Otherwise, the practicality of the entity would be lost. The internalist

---

[168] This doesn't come through all that clearly in Nagel's formulation. He says "the presence of a motivation for acting morally is guaranteed by the truth of ethical propositions themselves" (*ibid.*). This gives the impression that truths could determine motivation without being cognized. But I suspect Nagel's term 'propositions' is intended to refer to judgments (later in that same paragraph, he says "[e]xternalism holds, on the other hand, that the necessary motivation is not supplied by ethical principles and judgments themselves"), so that his formulation aligns with Korsgaard's.
[169] That is, normative reasons are primary and operative reasons are secondary (see *What We Owe to Each Other*, page 19).
[170] Nagel mentions Mill and Moore (*ibid.*, page 8), and Korsgaard mentions Mill, Prichard and Ross ("Skepticism about Practical Reason", pages 315-316).

maintains that the agent's thinking is continuous as it moves from representation to motivation, so that practical thought involves reasoning from a conative capacity as well as a cognitive capacity. If we opted instead for an entirely cognitive conception, one depicting thought about normative entities as contemplative, i.e., without any conative upshot, we'd have to rely on an extra-rational mechanism to trigger the appropriate response. This is what the externalist does; on her view, practical thought concludes with judgment, and doesn't extend to motivation.

Practicality as a feature of reasoning is parallel to the practicality of normative entities. In both cases, what distinguishes the practical from the non-practical is an orientation to action. Aristotle provides the anatomy of the idea—practical reason operates on our conations,[171] and is telic, i.e., it "aims at an end".[172] Kant takes roughly the same view; he specifies the relation in which the will stands to its objects: it sets itself "*to make them real*".[173] To say the same thing in the contemporary vernacular, practical reasoning exhibits the world-to-mind direction of fit, so that success in practical reasoning is, at least in part, a matter of causing the world to align with one's representation, rather than bringing one's representation into alignment with the world.[174]

It does sometimes happen that we neglect the practicality of practical thought. Anscombe warns against modern philosophy's apparent preoccupation with the "*mode of contemplative knowledge*", and its tendency to occlude or assimilate non-contemplative reasoning.[175] So, for instance, a normative conception may make the mistake of modelling practical reason on a conception of

---

[171] Like our desires and choices; see *Nicomachean Ethics*, VI 1139a21-31.
[172] *Ibid.*, VI 1139a36-1139b1.
[173] *Critique of Practical Reason*, page 89.
[174] This distinction is conveyed most clearly by Anscombe in §32 of *Intention*. It's the conative dimension of practical reasoning that exhibits the world-to-mind direction of fit. The cognitive dimension exhibits the speculative direction, mind-to-world. Practical reasoning involves both, though each attends to different domains—practical cognition defers to the normative domain of the world, and practical conation intervenes in the non-normative domain of the world.
[175] *Ibid.* Korsgaard issues a similar warning, though she singles out moral realism as a locus for this preoccupation (*The Sources of Normativity*, page 44). Patricio A. Fernandez does the same in "A Realistic Practical Conclusion", and accuses Scanlon (page 118). I believe both realism generally, and Scanlon in particular, can be absolved of this charge. But this isn't the place to make that case.

speculative reason. In annexing the one to the other, the special tie to action falls from view. Normative internalism binds motivational import directly to the cognition of normative entities, so that the practicality of practical thought is preserved.

But internalism isn't in the first instance engineered to guard against that mistake. Korsgaard explains that "an internalist believes that the reasons why an action is right and the reasons why you do it are the same."[176] Externalists deny this; on their view, normative reasons are fundamentally different from agent's reasons for action, so that the one can only determine the other by way of a mediating entity—as Nagel says, "an additional psychological sanction is required to motivate our compliance."[177] This undermines the unity of practical thought as simultaneously cognitive and conative, resulting in either a bifurcated conception of *practical thought*, with two rational capacities operating independently from one another, or a contemplative conception of *practical thought*, according to which there is no practical reasoning, not in the sense of reasoning from a conative capacity. Both threaten the idea of *practical thought* by denying the possibility of action from a normative entity.

A part of what motivates externalism is the thought that cognition is unlike conation in that the first is straightforwardly propositional, and the second isn't. There's something to this thought; the difference between cognition and conation isn't simply a matter of reversed directions of fit. Really, the direction-of-fit metaphors make the two seem more alike than they in fact are. While it's true that conative states have fulfilment conditions in largely the same way cognitive states have truth conditions, conative states also involve quite a lot else beside that. And motivation isn't itself propositional (though it's necessarily connected to the realization of some end, which is capable of propositional expression). It's a different kind of thing.

---

[176] "Skepticism about Practical Reason", page 316. It goes without saying that she has in mind cases in which we succeed in representing our reasons and in acting from them.
[177] *The Possibility of Altruism*, page 7.

That said, we mustn't exaggerate the difference. Yes, conative attitudes are often only clumsily expressed by that-clauses—it's natural enough to say "I believe that the cat is on the mat", but it's relatively awkward to say "I want that the cat is on the mat". We're likely more inclined to say "I want the cat to be on the mat", a form of expression that leaves it unclear as to what the propositional content of my want might be; after all, 'to be on the mat' isn't a complete sentence. But this doesn't settle the issue. The conventions of English provide clues as to the nature of the psychological concepts, but they don't supply anything like evidence (especially in this case, since the claim 'I want that the cat is on the mat' is perfectly intelligible). And there's a strong countervailing consideration: as we've seen, the holism of the mental indicates that cognitive attitudes inhabit a lattice of inferential connections, many of which lead to and from conations. Moreover, the possibility of explanation from conative attitudes relies on this same lattice—motivation isn't itself propositional, but it's only recognizable as such against a propositionally delineated context. This rules out the possibility that cognitive and conative states are propositional in radically different ways, since they must be for the most part congruous if they're to share a system.

But we might nonetheless maintain that conative states are different enough to pose a problem for the unity of practical thought. Cognitive attitudes represent actual states of affairs, and conative attitudes represent possible non-actual states of affairs—their objects are largely the same, apart from this difference in modal operators. But if an attitude is to be practical, it must, in addition to representing its target state of affairs, in some sense connect to the process of realizing that state of affairs.[178] And this process is unlike representation, since it exhibits a kind of generality

---

[178] All conative attitudes have fulfilment conditions (as Scanlon explains, "a desire that P is fulfilled if it is the case that P" (*What We Owe to Each Other*, page 41)), but not all are practical. In order for an attitude to be practical, it must include the judgment that its fulfillment is possible via the judger's agency. There are attitude terms that mark this distinction, e.g., all choices are practical, but many wishes aren't (see Kant's *Metaphysics of Morals*, page 213).

in that it isn't committed to any maximally particular conception of how the process unfolds.[179] For

instance, if the cat isn't at this moment on the mat, and I endeavour to place her there, my conative

states includes the representation of the possible non-actual state of affairs in which she's on the

mat, but also a connection to the process of placing her there. And that process is only capable of a

loose formulation, since it doesn't fix each detail of my finding her, picking her up, conveying her to

the mat, and leaving her there. And if, once I've lifted her, she squirms in my grasp, then my

manoeuvring to carry her more comfortably will belong to the process of placing her on the mat,

even though I had no notion that she would squirm once I began.

This suggests a partly dispositional formulation of practical intentionality, like the kind

described by Smith.[180] Conative attitudes, to the extent that they're practical, include prepared states,

that is, states at the ready to respond in certain ways to certain events. These serve to map the

processes by which agents seek to realize their objects. Given this difference with cognitive

attitudes,[181] some may be inclined to doubt that a wholly unified account of practical thought is

available. Again, I take the holism of the mental to reveal the different propositional attitudes to be

so deeply intermeshed as to preclude any significant disunity. That conative states have a

---

[179] I follow Davidson in framing this difference in terms of generality (see "Actions, Reasons, and Causes", page 5). He says "[i]f I turned on the light, then I must have done it a precise moment, in a particular way—every detail is fixed. But it makes no sense to demand that my want be directed to an action performed at any one moment or done in some unique manner. Any one of an indefinitely large number of actions would satisfy the want and can be considered equally eligible as its object" (*ibid.*, page 6). This usage of the terms 'general' and 'particular' is an unhappy one, since in this context all representations are particular, even representations of generalities, like 'all cats are like mats'. But I can't see a clearer way to convey the idea. (In fact, the situation is even more complex, since the relevant process does represent *some* content, if only negatively—if I were to turn on the light with my foot, we could say that this isn't what I had in mind.)

[180] See *The Moral Problem*, pages 113-115. Myers favours Smith's conception (*Donald Davidson's Triangulation Argument*, page 140), but denies that this commits him to Smith's Humean conception of *conation*. As usual, I side with Myers.

[181] Of course, cognitive states involve dispositions too. As Scanlon explains, "a person who believes that P will tend to have feelings of conviction about P when the question arises, will normally be prepared to affirm P and to use it as a premise in further reasoning, will tend to think of P as a piece of counterevidence when claims incompatible with it are advanced, and so on" (*What We Owe to Each Other*, page 21). And as Smith explains, just as attitudes with the world-to-mind direction of fit are disposed to be extinguished once their objects obtain, attitudes with the mind-to-world direction of fit are disposed to be extinguished if their objects cease to obtain (*The Moral Problem*, page 115). But dispositions are more intimately related to the conative orientation to content than they are to the cognitive orientation to content. This claim is, I recognize, quite vague, but this isn't the place to expand on it.

dispositional element isn't cause for concern, since, again, motivation is interpreted against a

propositional backdrop (and anyway, conative states also have an immediately propositional

element, the one with the world-to mind-direction of fit).

Still, it's worth considering what shape a disunified account might take. I mentioned a purely

contemplative account before. I have in mind the externalism that Korsgaard attributes to Ross and

Prichard. She says:

> [t]hey believed that there was a distinctively moral motive, a sense of right or desire to do
> one's duty. This motive is triggered by the news that something is your duty, and only by
> that news, but it is still separate from the rational intuition that constitutes the understanding
> of your duty…. Intuitionism is a form of rationalist ethics, but intuitionists do not believe in
> practical reason, properly speaking. They believe there is a branch of theoretical reason that
> is specifically concerned with morals, by which human beings can be motivated because of a
> special psychological mechanism: a desire to do one's duty.[182]

This result is, I think, quite damning. The contemplative account must accept nihilism about

practical reason because it represents our motivating states as altogether mechanical—and hence,

unintelligent—responses to conclusions reached via the cognition of normative entities. If the

relevant mechanism is neural (as is the fashion), we can tell this story: the onset of the representation

of a normative entity supervenes on a chemical event in the brain, namely, the secretion of a

substance, and this transmits an electrical impulse to a different region of the brain, from which it

causes movements in the body. Those movements then do as the represented entity prescribes.[183]

We can call this 'action', but only in a diminished sense of the term. The initial cognitive step

proceeds at the level of rational explanation, where the usual considerations pertaining to successful

representation can be brought to bear; this far, at least, we can speak of 'reason', in its speculative

register. But the remaining steps are only explanatory at the chemical level, so that the appropriate

image is of a marionette, rather than of an agent. Thus, talk of 'reason' in its practical register is

---

[182] "Skepticism about Practical Reason", page 316.
[183] The contemplative account's picture of successful agency is distorted enough. But its picture of failed agency, of movements that *don't* do as the represented entity prescribes, must be quite a lot worse.

inappropriate. (It may happen that, as the marionette moves its body, it conceives of itself as acting from a reason, but this is a transparent case of confabulation).

The result is something like the extensionalist line considered in subsection 1.3.1. Just like that conception of the *reason-relation*, the contemplative conception of *practical thought* entails a far-reaching revisionism, so far-reaching that it smuggles in eliminativism with respect to the idea, at least as it figures in ordinary action-explanatory discourse. As before, I assume that any view with those consequences is unacceptable. The more interesting externalist account is the bifurcated conception, the one that separates normative reasoning into cognitive and conative parts, with the one determining the other via some more complicated mechanism than straight causation (whether rational or otherwise). This is Mill's strategy. Korsgaard explains:

> [Mill] firmly separates the question of the proof of the principle of utility from the question of its "sanctions." The reason why the principle of utility is true and the motive we might have for acting on it are not the same: the theoretical proof of its truth is contained in chapter IV of *Utilitarianism*, but the motives must be acquired in a utilitarian upbringing.[184]

Most bifurcated conceptions place the division between cognition and conation where Mill puts it, with rigorous proof on the one side, and practical dispositions acquired in upbringing on the other. While it's true that practical dispositions involve cognitive as well as conative elements (or better, affective elements that establish patterns of attention, and in this way, direct both cognition and conation), it's possible that their cognitive components are limited to non-normative facts, or if not that, then to normative facts unrelated to their own normative credentials. On the bifurcated conception, cognition of normative facts, and specifically of the facts that justify acquiring practical dispositions, is the domain of the opposite side of the divide, the purely cognitive side.

One rationale for bifurcation derives from the special intelligence exhibited by habits, especially habits instilled from a young age. These are capable of greater sophistication and

---

[184] *Ibid.*, page 315.

reliability than ordinary beliefs about normative truths. But because habits are so sophisticated, they can't easily be formulated so as to enter into normative dialectic. Thus, the advocate of bifurcation allocates the business of proof to less intricate capacities. But not all rationales for bifurcation rely on psychological divisions of labour. Some revolve around cases in which normative beliefs are such that they undermine our capacity to be motivated to act as those beliefs dictate. Suppose, for instance, that human nature is such that, when any one of us arrives at a moral view, he is driven to proselytize with such fervour that he trespasses on the autonomy of others. Given this feature of our nature, we'd have reason to take steps to insulate our moral dispositions (like the ones involved in our being respectful of others) against the zeal developed upon adopting moral beliefs.

I'll have more to say about this kind of bifurcation in subsection 3.4.1. For the moment, I'll only repeat my objection that it straightforwardly disrupts the possibility of *action from* relations in central cases, like the recognition of a moral proof. While the holist should grant that Mill's dispositions are indispensable to practical life, she should also maintain that the practical wisdom enshrined in those dispositions can be propositionally expressed in such a way as to figure in dialectical exchanges, albeit with these qualificaitons: to say of a virtue that it can figure in dialectic exchanges isn't to say that that virtue is codifiable. The relevant exchanges are flexible; they can convey an aspect of, or even the whole of, a virtue without claiming to have specified every detail. The sense of 'proof' intended is a loose one. As for the circumstances in which cognition frustrates conation, they're either peripheral, and to that degree not a cause for concern, or else they're ruptures in the integrity of rational agency that must be resolved.

3.3 KORSGAARD'S CONSTRUCTIVISM

Korsgaard's constructivism does more to foreground the idea of *rational animal* than any other contemporary metanormative conception. The overlap between her view and my own (really, the

debt that my view owes hers) is extensive. I don't, at this moment, feel that I'm up to the task of

cataloguing all of the ways her view articulates the rational animal's complaint. The best I can

manage is to say a bit about the possibility of a rapprochement between her view and normative

realism, and then consider a few lessons from her constructivism that might be used to guard against

objections to the immanence constraint. This should provide a glimpse, at least, of the extent to

which her constructivism mobilizes to the rational animal's cause.

It also gives me the opportunity to address a worry that lingers from subsection 2.3.4,

namely, that my argument for the thick formulation includes among its consequences the truth of

normative realism, and this, some will object, surreptitiously settles a long-standing dispute that

ought to have been tackled head-on. I should, if nothing else, expand on the inference that I take to

run from the rational animal's complaint to realism, and to indicate what it rules in and what it rules

out. Since the argument proceeds from semantic externalism with respect to practical content, it

defeats the nihilisms and reductionisms I mentioned in subsection 2.1.2, as well as a variety of non-

cognitivisms and revisionary conceptions. Perhaps more remarkably, it requires a specifically

naturalist realism, so that it defeats non-natural realisms like Scanlon's; I'll return to this point in

section 4.4. But its consequences for Korsgaard's constructivism are less clear. My best guess is

that the naturalist realism that follows from semantic externalism takes us closer to constructivism

than we might have expected, but that constructivism must nonetheless undergo substantial

modification if it's to meet the immanence constraint as I've formulated it.

I delineate normative realism as the aggregate of three claims: i) normative discourse is truth-

apt, ii) some of its claims are true, and crucially, iii) truth-values for (some) normative claims are

inquiry-independent.[185] That last claim is innocuous as it stands (or it is up to a point—I'll have

---

[185] Korsgaard distinguishes between procedural and substantive realisms (*The Sources of Normativity*, page 35); procedural realism—which is her view—accepts i) and ii), but substantive realism—which she rejects—requires the addition of the inquiry-independence proviso, iii. That proviso requires some unsnarling, as we'll soon see (cf. Street's 'A Darwinian

more to say about inquiry-independence in a moment), but it's sometimes formulated in terms of full-blown agent-independence, the result of which is a realism that fails to meet the immanence constraint. And, as it happens, there are realisms that espouse the variety of normative austerity prohibited by the negative face of the immanence constraint. It's in this connection—the one given by the appearance of a realist proclivity for austerity—that my formulation of the immanence constraint is a descendent of the arguments with which Korsgaard assails realism in the opening lecture of *The Sources of Normativity*.[186] But the question is whether every realist conception is vulnerable to these arguments. That is, is it in the nature of realism to favour austerity, or is their occasional convergence just an accident of the history of the philosophy that developed them?

It's clear enough that Korsgaard's arguments censure certain varieties of realism (and, we might add, certain varieties of primitivism) for failing to furnish the resources agents require to make their reasons perspicuous, whether to themselves or to others. Take, for instance, the paradigm deliberative context in which an agent is presented with a reason-claim, even a true one, and asks of that claim whether it is, in fact, true. Korsgaard asserts that "[t]he realist's answer to this question is simply 'Yes'. That is, *all* he can say is that it is *true* that this is what you ought to do."[187] Clearly, this is too impoverished a response. If all that our philosophy can manage is "yes, there is that reason" or "no, there is no such reason" (equivalently, "yes, the primitive applies" or "no, it doesn't"), discourse about reasons is inhibited to the point that we are refused the means with which to manifest our natures as rational animals. This is reflected in the discursivity criterion, but also in what's required in order to solve the aspect problem as it arises for practical thought.

Dilemma for Realist Theories of Value', pages 110-111), but the idea is intuitive enough. As Korsgaard says "[m]oral realism … is the view that propositions employing moral concepts may have truth values because moral concepts describe or refer to normative entities or facts that exist independently of those concepts themselves. We have the concepts in order to describe or refer to those facts" ("Realism and Constructivism in Twentieth Century Thought", page 302-303).

[186] At pages 33-34, 38-41, and 44.

[187] *The Sources of Normativity*, page 38.

My argument for the negative face of the immanence constraint can be viewed as a kind of recasting of this Korsgaardian complaint. While that complaint is directed against normative realism, I think it's instructive to hear it as though it were designed to correct a trend in normative primitivism, rather than realism. It's directed against realism too, but only insofar as it indulges in this same trend, the one that leads to austerity. I'm not certain how to diagnose the realist's inclination (if she has one) to adhere to that trend, though I grant that it's too easy to dismiss it as nothing more than an accident of philosophical history. It may be that there's something mildly bewitching in the thought that's characteristic of realism, and that this provides a temptation to slide into austerity. But even if so, this doesn't weaken realism, since we can resist that temptation. In any case, there's nothing—anyway, nothing in principle—that commits realism to austerity, which suggests that we ought to leave open the possibility that there are realisms that can escape Korsgaard's line of attack. Scanlon's realism is, I propose, capable of a formulation that needn't elicit any constructivist complaint (that said, Davidson's realism is more secure than Scanlon's, since it builds fruitful discourse in from the start).

But what would Korsgaard make of this reframing, according to which her ultimate target is austerity rather than realism, and constructivism may in the end be compatible with realism? Perhaps she'd find no fault in it; perhaps it isn't even a reframing. After all, she says that "considered in one way, constructivism and realism are perfectly compatible."[188] Constructivism is the view that "[p]ractical philosophy… is not a matter of finding knowledge to apply in practice. It is rather the use of reason to solve practical problems."[189] Read superficially, there's nothing here that contradicts realism's inquiry-independence claim. Korsgaard's primary target is the "model of

---

[188] "Realism and Constructivism in Twentieth Century Thought", page 325. See also *The Sources of Normativity*, page 108.
[189] *Ibid.*, page 321. Constructivism is a good deal more complex than this pithy formulation conveys. Its name can be misleading—it isn't the view that normativity is constructed by agency; really, it's better understood as the view that agency is constructed by normativity (thus, 'constitutivism' conveys the idea somewhat better).

application",[190] which is a wholly classificatory conception of *normative reasoning*, i.e., one that depicts such reasoning as though it were only a matter of organizing actions under the headings 'good' and 'bad', without any consideration for how these designations and their instances might figure in practical thought *qua* thought oriented to acting. But nothing foists that model on realism (though many realisms implicitly embrace it).[191]

All the same, it'd be disingenuous to suggest that Korsgaard would find no fault at all with a realism that disavows austerity. There are two reasons for this: first, her anti-austerity argument (if that's what it is) isn't the only argument she levels against realism, and second, the olive branch extended in the passage mentioned a moment ago—"considered in one way, constructivism and realism are perfectly compatible"—isn't one that any realist, even a non-austere realist, could accept. That passage continues in this way:

> 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.[192]

---

[190] As described at pages 315-317 of *ibid.*

[191] Korsgaard's attacks on realism sometimes linger on passages in which its proponents appear to lean on the view to legitimize a complacency or inertia that's suggestive of austerity—Samuel Clarke pronounces that his claims "are so notoriously plain and self-evident" that skepticism is simply unintelligible (*A Discourse of Natural Religion*, page 194; quoted at page 39 of *The Sources of Normativity*), and Nagel says of a claim about agent-neutral reasons that "I am somewhat handicapped by the fact that I find it self-evident" (*The View from Nowhere*, pages 159-160; quoted at page 41 of *The Sources of Normativity*). Korsgaard withdraws her complaint (or rather, allows that she might do so) when those same proponents attend to the relationship between agency and normative entities, e.g., when they offer fuller pictures of how thought about normative entities might lead to action, or might inform our nature as rational animals (for Clarke, at pages 31-32 of *ibid.*, and for Nagel, at page 41, note 68). This lends some credibility to my notion that it's austerity, with its bare, unjustified assertions—especially the ones invoking self-evidence—that disturbs her, rather than realism.
  For what it's worth, Korsgaard never, as far as I know, explicitly directs her anti-realist arguments against Scanlon. This may be evidence that she takes his realism to be less vulnerable to those arguments than the others. Granted, if this is evidence, it's quite weak, since the publication in which Scanlon announced his realism—*What We Owe to Each Other*—took place a few years after the publication of *The Sources of Normativity*, in which the bulk of Korsgaard's anti-realist arguments are presented. Thus, the timing isn't right for Scanlon to bear the full brunt of Korsgaard's critique. (If I had to guess, note 3 at page 302 of "Realism and Constructivism in Twentieth Century Thought" is written with Scanlon (and Nagel and Raz) in mind.)

[192] *Ibid.*, page 325.

The chief virtue of realism, as I see it, is its capacity to satisfy the transcendence constraint. The reality the realism finds for normative entities—a robust, inquiry-independent reality—gives them distance from our agential nature. We can exploit this distance to assess the whole of that nature. But if we say of normative claims that their status as normative only arises for beings faced with particular problems, we thereby subordinate normative assessment to the problems we happen to have. In order to respect the transcendence constraint, we must represent our normative situation such that we can ask whether we're attending to the right problems. That is, the relevant normative facts must be prior to what, as a matter of fact, is at this moment bothering us. Thus, if what leads us to realism is the prospect of meeting the transcendence constraint (and I assume that this is a large part of the appeal), we'll be disposed to refuse Korsgaard's offers to accommodate us.

Does that settle the issue? If I were prompted to take a firm line, I'd say that realism is unlikely to be reconciled with constructivism as Korsgaard conceives of it. But if I were permitted to stake a sketchier position, I'd suggest that the two are quite close to one another. The crux of the issue is the sense of objectivity in play. I've allowed, for the sake of simplicity, that realism is defined by the claim that normative entities are inquiry-independent. But this term, 'inquiry-independence' is, though evocative, a little clumsy. The realist proposes to fit normative entities into the world—this expression, 'the world', must be read in its weightiest sense, the one according to which the world is a certain way, and it continues to be that way whether we inquire into it or not. But as we saw in our investigations into the possibility of practical thought, the world presents itself under indefinitely many aspects, and it's up to us, at least in part, to specify the aspects to which our terms refer. Truth-values don't emerge until there are thoughts and claims that can bear them, and these don't emerge until aspects are fixed. And, on Davidson's triangulation picture, aspects are fixed in the course of the shared project of inquiring into the world. It follows that, for Davidson (and for me as a follower of Davidson's), normative truth-values aren't strictly speaking inquiry-

independent.[193] This isn't just a complication for Davidson's realism; it's hard to see how literal inquiry-independence could be sustained at all, not if we hope to develop a theory of normative content. The key, I think, is to find normative transcendence in the multi-faceted world, the one that houses an infinite array of candidate normative aspects even before we've fixed our meanings to any. But this solution can only be taken so far, since when we speak of normative entities, we almost never mean to refer to the normative manifold—we mean the determinate entities that are only individuated once aspects are fixed and practical thought is on the table.

At any rate, all of this shows that it's remarkably difficult—much more than one would expect given their reported antagonism—to specify the features of constructivism and realism that keep them apart. Korsgaard says that normativity arises in the context of a problem. But inquiry itself is, in a manner of speaking, a problem. And since nothing that we can think or say is literally inquiry-independent, then everything we can think or say arises in the context of a problem. The problem I mean is quite wide-reaching—it's conveyed by the question 'what's the world like?'.[194] That's, as it were, the realist's question. And, more importantly, while Korsgaard allows that, viewed from outside of this problem, expressivism is true (i.e., true normative claims only express the values we happen to endorse), there isn't any perspective external to inquiry, so that there's no room for expressivism to occupy, and we can dismiss it as confused.

---

[193] Myers and Verheggen investigate Davidson's conception of *objectivity* in "Realism Rehabilitated". What I say here is informed by that investigation, though I must admit that I find the matter quire perplexing, and I have very little confidence that I've properly understood their view, or for that matter Davidson's.

[194] The size of the problem is a virtue, since this is what secures the categoricity we need from a conception of *normativity*. Its categoricity is secured in the same way Korsgaard secures the categoricity of rational principles, namely, via that which is constitutive of thought (or more specifically, of action; see *Self-Constitution*, pages 28-34). The 'what's the world like?' question underlies triangulation, and is constitutive of thought in that way.

  Korsgaard's constructivism is Kantian, and, as such, its "problem is the one set by the fact of free agency. It is nothing less than the problem of what is to be done" (*ibid.*, page 322). I take it that if we limit our 'what's the world like?' question to the deontic region of the world, we get 'what is to be done?', so that, again, realism and constructivism converge. (If I'm to persuade Korsgaard of this, I must show that the 'what's the world like?' question isn't committed to the 'model of application'. But I think this can be done.)

This deep connection between normative reality and normative thought provides the beginnings of a reply to another of Korsgaard's anti-realist arguments, the one to the effect that realism must misunderstand practicality. Oversimplifying a bit, the argument suggests that realism must espouse the kind of externalism described in the previous section. But because the possibility of conation begins with an encounter with normative entities, and because the individuation of those same entities begins with the possibility of conation, the unity of practical reason—of normative cognition (really, of normative content) with conation—is assured. Practicality is present from the start, so that the realist is never given the opportunity to mislay it.

This response should also suffice to address the Anscombian worries with which Korsgaard sympathizes about the assimilation of practical reason to the "mode of contemplative knowledge".[195] The contemplation of normative entities must have an immediate motivational upshot, since, again, those entities are individuated in terms of their contributions to practical thought. (It's worth adding that the rationale for introducing speculative reason to normative contexts isn't simply that normative claims are more elegantly analyzed as truth-conditional than not.[196] The main consideration that recommends speculative reason in these contexts is the fact that truth, as it figures in speculative reasoning, offers a model for normative authority. This consideration is specific to normativity; it isn't a neutral piece of linguistic analysis, divorced from the peculiar features of its subject matter.)

Korsgaard also warns that realists can't formulate an account of the principles of practical rationality without presupposing them. So, for instance, realists can't model instrumental rationality in terms of agent-independent reasons to pursue means to our ends, since if they attempted this, they'd need to posit reasons to pursue further means to those means to our ends, and then still

---

[195] *Intention*, §32.
[196] Though I do find this fact persuasive; see Scanlon at page 1 *What We Owe to Each Other.*

further means to those means, and on and on *ad infinitum*. This is because action itself exhibits an instrumental structure,[197] so that our reasons can't instruct us to erect that same structure; if we're to act from our reasons, the structure must already be in place.[198] This lesson generalizes to all of the formal principles of practical thought. It also generalizes to my formulation of the problem that frames practical thought—we mustn't posit a reason to pursue the 'what's the world like?' question, since our interest in that question, or rather, the orientation to the world it invokes, is presupposed by the possibility of action from reasons. The same is true of the triangulator's aims to get normative matters right—this aim underlies prescription, and so, it can't itself be prescribed.[199] But this isn't embarrassing for realism. The principles and the contexts that constitute agency aren't themselves normative, and so they lie outside realism's jurisdiction. What the realist represents as real are the normative entities themselves, the things from which we act given the formal machinery that philosophers like Korsgaard have identified.

Still, that machinery is an essential component of normativity. It's the underlying machinery of agency, but also of the facet of the normative domain that faces agency, that's tuned to it. We might put it this way: just as there is a corner of the physical domain, the corner belonging to agents,

---

[197] I earlier characterized instrumental reasoning as necessarily causal (note 57), but the instrumental principle in play here encompasses both causal and constitutive means-end relations. Korsgaard describes this wider principle at pages 27-28 of "The Normativity of Instrumental Reason":

> the instrumental principle is nowadays widely taken to extend to ways of realizing ends that are not in the technical sense "means," for instance to what is sometimes called "constitutive" reasoning. Say that my end is outdoor exercise; here is an opportunity to go hiking, which is outdoor exercise; therefore I have reason to take this opportunity, not strictly speaking as a means to my end, but as a way of realizing it.

The claim that instrumental rationality is a formal principle of action is only true provided that we understand that principle in the way Korsgaard suggests.

[198] Korsgaard explains at page 54 of *ibid*.:

> [t]he instrumental principle cannot be an evaluative truth that we apply in practice, because it is essentially the *principle of application* itself: that is, it is the principle in accordance with which we are operating *when* we apply truths in practice… [W]e cannot give the instrumental principle a realist foundation. But if we cannot give a realist account of the instrumental principle, it seems unlikely that we will end up giving realist accounts of the other principles of practical reason.

[199] Relatedly, truth can't by itself supply a reason for belief. That is, I don't have a reason to believe a true claim simply because it's true (though I may have a reason to ensure that my beliefs are true). If I'm asked whether the cat is on the mat, the truth of 'the cat is on the mat'—or anyway, its truth-maker, namely, that the cat is on the mat—supplies me with a reason to believe that she is. But if I'm not asked about her, no reason arises. The agent is always deferential to the world, but what this deference demands of her agency depends on what she's doing.

that by its nature faces normativity, there's a corner of the normative domain, the one belonging to deontic entities, that by its nature faces the physical domain. In order to understand the normative domain, we must understand its agency-facing corner, which in turn requires that we understand what it is for deontic entities to be agency-facing, i.e., what it is for them to necessitate physical entities. The principle that enables them to do so is the very same principle that makes those physical entities agents. Or (more clumsily but perhaps more evocatively), it's the principle that collects a bit of normative stuff and a bit of physical stuff together in a relation, thereby making the one a deontic entity and the other an agent. We need an account of that principle, not just in order to make sense of the agent, but also to make sense of the entity.

This suggests something like the constructivist-realist hybrid I mentioned in the introduction. Realism attends to normativity *simpliciter*, and excels in modeling normative transcendence. But it struggles to find the resources to explain the agent-relationality of deontic normativity. Constructivism attends to agency, and details the rational principles that establish the possibility of agent-relationality. It excels in modelling normative immanence, but it lacks the resources to represent what is distinctive about normativity, namely, its authority.[200] It seems natural, then, to seek to unite the two in a single conception. Perhaps Korsgaard's own constructivism can't accept a realist partner. Even if so, the constructivism we find for our hybrid must be quite a lot like Korsgaard's, in that it must map the structure of what makes deontic agent-relationality possible. The only thing it must relax is Korsgaard's aversion to the realist conception of normative objectivity. (But again, it isn't clear to what extent Korsgaard really is averse to that conception, especially once it's granted that we must soften our claims about the inquiry-

---

[200] This will sound surprising given the attention Korsgaard gives to the issue of moral authority, understood in terms of the inescapability or overridingness of moral reasons. But I've assumed that the appropriate measure of normative authority is given by the transcendence constraint. It's possible for a view to succeed by my criterion (in that it allows for the scrutiny of our whole agential nature), but fail by hers (by representing moral reasons as being in certain circumstances overridable). Myers' view is an example (see pages 194-195 of *Donald Davidson's Triangulation Argument*).

independence of normative truth-values if we're to develop a workable theory of normative content.)

I've said that, if constructivism is to figure in a promising hybrid, it must relax certain of its more radical claims. Obviously, in order for the hybrid to have any chance of satisfying the transcendence constraint, it can't hold that normative entities are generated by our practical identities.[201] Much as with Williams, who could only permit the evaluation of desires in terms of other desires, Korsgaard can only permit the evaluation of identities in terms of other identities. The fact that there's a moral identity that's properly categorical, and has authority over all the rest, is no help, not as far as the transcendence constraint is concerned, since that identity inheres in agency, so that it must be assessable along with everything else.[202] Still, Korsgaard is right that practical rationality must include something like a principle of integrity, which enables agential unity via commitments to identities, with some identities establishing limits for others.[203] In fact, her ideas of *self-constitution*, *integrity*, and *practical identity* are each indispensable to any representation of the nature of agency.[204] Moreover, these ideas are of use in plotting first-order normative theory. But they mustn't be permitted to perform an ontological role, since that role is reserved for the candidate aspects of Davidson's practical externalism.

---

[201] On Korsgaard's view, a practical identity is "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking" (*The Sources of Normativity*, page 101). Her constructivism represents all normative entities (or anyway, obligations, but she likely takes this to be the material for the rest) as originating in practical identities.

[202] Korsgaard argues that the human identity—the one that establishes that we, as self-conscious animals, must answer to obligations deriving from other more contingent identities, and that has us value ourselves and others under that description—provides a basis for a categorical morality at pages 121-122 of *ibid.*

[203] See *ibid.*, page 102, but also her argument that the categorical imperative is a constitutive principle of action in section 4.4 of *Self-Constitution*. Read this way, the categorical imperative serves as a principle of integrity, one that unifies the many parts of an agent so that she can carry out an action.

[204] In a manner of speaking, Kantian self-legislation and Korsgaard self-constitution are just a matter of *acting from* normative entities, rather than merely *conforming to* them. But Korsgaard's idea expands on how this works—we don't typically act from normative entities taken singly; instead, we act from them by locating them in a web of many entities, a web which is mapped by practical identities. Thus, in acting from an entity, we also act from its web, and, in this way, adopt its corresponding identity. It's in this sense that, as she says, necessitation is self-constitution (*ibid.*, page 7).

Moreover, as it stands, Korsgaard's constructivism doesn't have the tools to develop a theory of practical content. She marshals a sophisticated account of what rational animals share with non-rational animals, and of the moment when rationality is achieved, from which there is a backwards saturation of rationality into each of the pre-rational animal capacities.[205] This must be part of the story of how the transition from non-rational to rational animals took place. But it doesn't provide a clue as to how content is acquired. And (somewhat surprisingly, given the earlier objection to realism), her constructivism can't meet the discursivity criterion. A shared moral identity ensures that rational animals have a good deal in common, but this can't by itself supply an assurance that any two rational animals are talking about the same thing.

In any case, there are a couple elements of Korsgaard's view that anticipate somes lines of objection (in particular, lines of objection favoured by realists, like Scanlon) to the immanence constraint. First, 'reason' as Korsgaard uses it (and as I use it) is a "normatively loaded term", that is, its correct applications are immediately normatively authoritative.[206] So, in the context I mentioned before, the one in which an agent asks whether a reason-claim is true, and where the realist—or rather, the realist that Korsgaard targets—can only answer 'yes', the issue is how to apply a term that is, according to both the agent and the realist, already normative. Korsgaard isn't to be read as asking after an account of the normativity of normative reasons—clearly, that question is settled analytically. Rather, she's asking for an account of how we can receive those reasons as necessitating forces, or to say the same thing, of how we agents can understand the normativity of our reasons. And this account should include materials with which to assure us that our normatively loaded terms really are correctly applied, materials we can deploy in discourse to demonstrate that our reasons-claims genuinely express the normative authority that they purport to express. This is

---

[205] See *Self-Constitution*, section 6.1.
[206] *The Sources of Normativity*., page 42. Of course, there are senses of 'reason' that aren't normatively loaded, like its explanatory sense. But both Korsgaard and I are in the first instance attending to the normative sense of the term.

with a view to discerning what it is to be gripped by a normative reason, that is, what takes place as we accept a reasons-claim as genuine.

I emphasize this point in part because, when Scanlon considers the argument conveyed in this same passage,[207] he responds by accepting her diagnosis:

> Korsgaard is quite correct about what a Reasons Fundamentalist, or at least *this* Reasons Fundamentalist, would say in a situation of the kind she imagines. According to a Reasons Fundamentalist, the relation that holds between an agent and a consideration X in such a situation just *is* the relation of *p*'s *being a reason for that agent to do a*. The "grip" that this has on the agent just is the relation: *being a reason for* him or her (or in the strongest cases, a conclusive reason). As Korsgaard puts it, quite correctly, a reasons Fundamentalist "insists on the irreducible character of normativity." The fundamental disagreement here concerns whether some further explanation can and should be given of why the agent in this situation *must* treat *p* as a reason.[208]

I worry that there's a miscommunication here. I suspect Korsgaard accepts this same analysis of the reason-relation. Her complaint isn't that some further explanation should be given at *that* level. What Korsgaard wants to see from the realist is a mechanism, baked directly into normative conceptions, that tethers agents to true reason-claims by appealing to a feature of agential nature. Understood that way, the immanence constraint only restates her case: metanormative theory mustn't simply show normative entities to be the sort of things that command the behaviour of agents. It must also show agents to be thus command*able*, to be the sort of things that receive commands. This requires an explanation given at the level of a theory of agency.[209]

---

[207] Again, at page 38 of *ibid.*

[208] *Being Realistic about Reasons*, page 10.

[209] My criterion is in one way stricter, and in another way looser, than Korsgaard's. It's stricter in that it demands explanation at three levels, the ones identified by the positive face of the immanence constraint, namely, content, the first order, and agency, while Korsgaard would be satisfied with only the last of these. But my criterion is also looser in that there are contexts in which an explanation at any one of these levels will serve. So, for instance, questions about the truth of a reason-claim can sometimes be settled by appeals to other reasons. This isn't necessarily a horizontal move, since some reasons are, in a matter of speaking, deeper than others (e.g., moral reasons often have priority over prudential reasons). But even in cases where the move is only horizontal, it might still address the question, if just because we can come to understand one reason better by recognizing how it relates to others. This makes it easier for Scanlon to satisfy the request, since it allows him to do what he's already inclined to do, namely, explain from the first-order, e.g., by pointing to particular reasons.

Adapted to the immanence constraint, Korsgaard's point can be expressed this way: there's no addressing the constraint by repeating that it's in the nature of reasons that they figure in relations that include agency as a *relatum*. It's analytic that the agent who is fully apprised of her reasons is in this way apprised of her place in the corresponding reason-relations. The question is how she comes to be fully apprised, how she inserts herself into the agent-place. That's where we'll have to direct our philosophical attention if we're to address the rational animal's complaint.

The second lesson is derived from that the fact that it's a mistake to fold Korsgaard's argument into the Humean metanormative tradition that, among other things, seeks to discredit realism by denying the possibility of unmediated motivation by desire-independent entities.[210] Korsgaard holds that, if an agent takes a reason-claim to be true, it follows immediately, absent any irrationality on her part, that she is motivated accordingly. This is a consequence of her internalism, which is akin to Nagel's, and is evident all throughout her work.[211] So, at least in the first instance, Korsgaard's worry doesn't dwell on how a realist's claims might activate agents' conative dispositions (though, as we've seen, she does worry that realism is committed to externalism). Her complaint runs deeper than Hume's—it isn't a question of the relationship between psychological apparatuses, but between agency and normativity.[212] The same is true of the immanence constraint. In fact, the tradition to which the immanence constraint belongs likely doesn't include Hume at all. We'd have better luck unearthing ancestors to the constraint in Kantian philosophy and Aristotelian philosophy, since these insist on the nature, and the distinctive features, of practical reason. Because Hume is, in the end, a nihilist about practical reason, the tradition he began in normative psychology

---

[210] This tradition overlaps with the Humean tradition in normative internalism, as described in section 3.1. Both have their source in the *Treatise*, book II, part III, section III. Still, they're different traditions. The one articulates the problem that the other proposes to solve.

[211] But see in particular "Kant's Analysis of Obligation".

[212] There are realists (Parfit is a prominent example) that misinterpret Korsgaard in this way, but Scanlon isn't among them. As he puts it, "[t]he problem is not a matter of motivation in this psychological sense, but rather of something more like normative authority" (*Being Realistic about Reasons*, page 10).

is entirely different. And, for that matter, the advocate for the rational animal in metanormative thought should regard desire-centric conceptions with suspicion. They're more likely to annihilate the rational animal's perspective than to express it.

Korsgaard's argument is often framed in terms of persuasion.[213] This is closer to the mark, since she contends that realism lacks the discursive tools to reveal the normative authority—or as the case may be, the absence thereof—disclosed by talk of normative reasons. But the gap she identifies can't be filled in with a rhetorical strategy. It helps to conceive of it in terms of something like Rawls' conception of *justification*: "[b]eing designed to reconcile by reason, justification proceeds from what all parties to the discussion hold in common."[214] This is the conception of *justification* that Rawls reserves for moral conceptions. It should be interpreted, I think, as justification by engaging a moral capacity, this being what's shared by the discussants. To fit the Korsgaardian case, it must be modified so to encompass justification by engaging a practical rational capacity. It still amounts to persuasion, after a fashion; anyway, as Rawls indicates, it's distinct from "mere proof".[215] But it's specifically rational persuasion. That's what it takes to address the rational animal's complaint as expressed by the discursivity criterion. It's not enough to offer her truth—we must endeavour to offer her reasons from which she herself can appreciate what makes a given reason-claim true.

3.4 CONTRACTUALISM AND THE MORAL COROLLARY TO THE IMMANENCE CONSTRAINT

This connects with a core tenet of contractualism. Social contract theory has opposed paternalism ever since John Locke excoriated Robert Filmer's patriarchal conception in the *First Treatise of*

---

[213] Scanlon complains that Korsgaard gives priority to "dialectical context[s]… in which two people are arguing about what reasons for action one of them has" (*ibid.*, page 12), rather than to contexts in which one deliberates "about what reasons one has oneself" (*ibid.*, page 13). See also Myers, at pages 159-160 of *Donald Davidson's Triangulation Argument*.
[214] *A Theory of Justice*, page 508.
[215] *Ibid.*

Government.[216] Filmer's conception is patriarchal in the sense of advocating for the rule of the father, but it's characteristic of the ruler-as-parent's relation to the subject-as-child that the former imposes a justice that the latter needn't (and typically doesn't) understand. This is what connects paternalism to patriarchy *qua* conception of the divine right of kings. The social contract theorist's insistence on the consent of the governed repudiates both Filmer's patriarchy and his paternalism.

Of course, the social contract theorist's criterion for political legitimacy isn't the *express* consent of each subject. It's enough that it would be rational for each to consent to the political institutions detailed in the contract. That said, each citizen is entitled to the opportunity to express her rationality in this way. Thus, the social contract must include something like a transparency or publicity provision, one that makes the design of institutions available to the rational scrutiny of the citizenry. Kant provides an articulation of this provision, what he calls "the *transcendental formula* of public right: 'All actions affecting the rights of other human beings are wrong if their maxim is not compatible with their being made public.'"[217] While it might not be feasible to meet this transparency provision in every political context, it ought to be respected by default. And where there are exceptions, citizens are likely still owed the opportunity to endorse the policies that determine the exception, e.g., when discretion is required as a matter of security. To this extent, political institutions that satisfy the provision manifest the freedom of the citizenry.

The social contract tradition belongs to the wider culture of individual freedom that endured through the Enlightenment, a culture informed by, among other things, the free-thinkers, who insist

---

[216] Opposition to paternalism predates Locke in the social contract tradition, but he provides a particularly emphatic case. Hobbes is, as always, an outlier, since he represents commonwealth by acquisition (that is, by conquest) and commonwealth by institution (that is, by common consent) as equally legitimate (*Leviathan*, XVII, 15).

[217] *Perpetual Peace*, appendix II, page 126 in *Political Writings*. See also Rawls, *A Theory of Justice*, page 115, though Rawls denies that publicity amounts to an adequacy condition (let alone a transcendental formula) on conceptions of *justice*. This marks an important shift (we might think of it as something like a return to Aristotle's claim at *Nicomachean Ethics*, I.3) in conceptions of *moral proof*. For Rawls, there's no moral argument that can deliver us decisively to its conclusion. The last step of the argument is always a judgment that can itself be made well or badly. (For an argument for transparency from the nature of normativity, see Korsgaard, *The Sources of Normativity*, page 17.)

that we determine moral, religious, and metaphysical matters for ourselves rather than deferring to authorities, like the church, or the (at that time scholasticism-dominated) academy. The free-thinkers' ideal is an unmediated relation between thinker and world. Necessitation involves a remarkably similar ideal, in that *action from* relations are also unmediated, being direct connections between practical thought and normative entities in the world. In this sense, the institutions championed by the Enlightenment align with the deontic-agential order of the normative domain. That isn't to say that those institutions inhere in that order—it's a good deal more complicated than that. Still, the two harmonize in a way that most other institutions don't.

In light of this harmony between the two, I propose that the rational animal is entitled to exercise her nature in practical thought in much the same way that the citizen is entitled to participate in political institutions. The relevant moral principle is ultimately the same in both cases, namely, respect for agency prescribes that each agent be furnished with the means to deploy her agential powers. Of course, we can't reasonably demand that our agency be enabled in every walk of life, but the closer we come to the central aspects of rational life, like politics, or simply action, the more forceful our moral claim.

I assume this principle is compelling as it stands. What's contentious is the use I want to find for it: I think this principle can operate as a success condition on metanormative theory, and, in this way, operate in tandem with the immanence constraint. After all, the rational animal's complaint isn't simply a matter of insisting on her place in our representations of the normative domain. It's also a matter of insisting on representations of her relation to that domain that are respectful of her dignity as a moral subject. Call this second dimension of the rational animal's complaint the moral corollary to the immanence constraint.[218] The constraint expresses the mutual

---

[218] It isn't a true corollary, since it can't be *deduced* from the immanence constraint, not immediately. The connection between the two is looser than that. Again, the ambition is to reveal a kind of symmetry between moral and metanormative elements. It isn't to derive the former from the latter (or vice versa) with Euclidean exactitude.

dependence of certain concepts—it finds fault with conceptions that fail to accommodate the idea

of *rational animal*, but only on the grounds that this failure makes them incoherent. It doesn't assess

the moral implications of this failure. This is left to the corollary.

The idea is best conveyed with examples. In the course of my formulation of the argument

for the immanence constraint, I touched on a few conceptions that, as I explained, generate

conceptual and metaphysical difficulties. But these conceptions also generate moral difficulties of

the kind flagged by the moral corollary. So, for instance, Korsgaard's realist, the one who can only

say "yes, there is that reason", goes wrong in this second way because he refuses us the means to see

that the reason-claim is true. The bare reassertion of its truth is of no use to us, not as we

endeavour to understand that truth.

That's bad enough, but things get worse once we elaborate his position. What is he to make

of our endeavours to understand the truth? Suppose we're not, at present, disposed to do as the

reason instructs. What prospect can we have, by the realist's lights, to chart a path from

disobedience to obedience? One possibility—not the only one, but it's so dangerous that it warrants

special attention—is that he must assume that either an agent is already fully acquainted with her

reasons, so that she has no need for a philosophy to bridge any gap, or she is lost with respect to

reasons, so that, in a manner of speaking, she's no agent at all, that is, she's hopeless; she can't be

saved. The moral consequences are obviously perverse. Respect for agency requires that we, at the

very least, recognize it.

The threat to the rational animal grows stronger once realism is matched with austere

primitivism and radical intuitionism. Such a view can encroach on the rational animal's domain by

laying claim to such things as goodness and the idea of *normative reason*, i.e., to the materials of

rational agency. And then, by placing those materials behind a veil that only intuition can part, it can

rob the rational animal of the capacity to express her nature. Of course, radical intuitionism is a

rational intuitionism, which is comparatively egalitarian as intuitionisms go, especially given that the austere primitivist denies that intuition is a power, something that different people might exercise in different degrees. But, given the fact of normative disagreement, together with the fact that radical intuitionism has no way to account for a mistake, the primitivist must, if she's honest, renounce this egalitarianism, and say that there are, on the one hand, the blessed few who get it right, and, on the other, the blighted many who get it wrong. That one has, and another lacks, the intuitive relation to normativity must, by her lights, be an austere primitive as well, since whatever is it that assigns animals to the two groups is just as inscrutable as everything else. This way, the lucky intuit that they are lucky, and the unlucky have no notion of any difference between them.

In fairness to metanormative austerity, it isn't clear that this is the position it must take on this issue. But then, it's in the nature of austerity that it offers no clue as to what shape its moral and political counterparts might take. As a result, it's difficult to attack the morality of austerity without grasping a little in the dark. With that disclaimer in place, I'll offer my best guess: austerity has a deep affinity with a social stratification that apportions moral agency to one caste while denying it to the rest. If we depict normative (and thus, moral) knowledge as requiring radical intuition, and if we depict normative ignorance as suggesting that the intuitive relation is defective or missing, then we can't effectively guard against the all-too-human temptation to claim the authority to subordinate others to our political will. Thus, though austerity doesn't itself prescribe it directly, we should expect the public political culture to receive austerity as licensing something like a Straussian worldview, according to which feudal hierarchy is both inevitable and appropriate, and the promise of democracy is only a ruse designed to mollify the otherwise ungovernable masses (and, of course, it will be maintained that the alleged truth of these claims must undergo obfuscation so that only the genius few can discern it).

The immanence complaint has plenty to object to in all this, but as I said, its objection is only conceptual, and not moral. Meanwhile, the corollary prescribes that each rational animal be provided, as far as is reasonable, with the materials for practical thought. Related to this, respect for the personality of rational animals entails, again as far as is reasonable, that these materials be allocated equally. In the context of metanormative theory, an equal distribution is easily achieved—the theory need only reject the arrangement just described, according to which the class of rational animals is divided into subclasses.[219]

There are a few complications here. First, for the most part, it's histrionic to accuse a theory of immorality; we ought rather to settle for calling it incorrect. And though some theories propagated in contemporary metanormative discourse do (though for the most part unintentionally) rob the rational animal of her right to express her nature in practical thought, they do so only in theory. Philosophical conceptions don't generate states of affairs themselves. In order to accomplish that, they rely on the development of institutions that are modelled on them. So, for instance, Aristotle's conception of human nature would, if it were to be enshrined in social institutions, result in the enslavement of the many by the few. Our moral objection to the conception derives from its projected (and historically recorded) institutional results. And the second complication: while, a theory of human nature, like Aristotle's or Hobbes', may recommend an unjust social hierarchy, and be to that extent morally risky, metanormative theory is comparatively undangerous. It takes somewhat more imagination to spot the moral risks in a theory of normative

---

[219] Of course, some are more successful in their practical judgments than others, and this success may reflect a disposition of character, something in the neighbourhood of practical wisdom or practical foolishness. The moral corollary doesn't object to distinctions drawn along those lines because they're distinctions drawn *within* the space of reasons to which all rational animals are admitted. It's in principle possible for each one to cultivate her capacity for practical thought, i.e., the gap between the wise and the not-yet-wise can be traversed. Most importantly, the wise and the not-yet-wise can (again, in principle) speak to one another about ethics and morality, and, in this way, find one another in their shared space of reasons. (But what about psychopathy? The idea of *psychopathy* is neurological, and not metanormative. The moral corollary can challenge (some) philosophy, but it can't challenge scientific fact.)

reasons, for instance. For this reason, the moral corollary isn't quite so urgent as, for instance, its political analogue. Still, it's worth exploring.

There's a third complication: metanormative conceptions and moral conceptions are developed at different orders, so that, at least *prima facie*, metanormative thought is exempt from moral censure. This requires a more elaborate treatment, which I'll delay until subsection 3.4.2. In the meantime, I think the best way to proceed is to develop another example. So, with this in mind, I'll devote a subsection to specifying a second conception that the moral corollary rules out.

3.4.1 Self-effacement in normative theory

In section 3.2, I described a variety of externalism that separates practical thought into cognitive and conative parts, with the two mediated (if at all) via some elaborate, non-rational mechanism. I briefly considered two such bifurcation accounts, the first of which prescribes something like a psychological division of labour, and the second prescribes that cognition and conative be separated so as to prevent the one from undermining the other. While I've registered my opposition to externalism as a general account of practical thought, and indicated how it fails by the immanence constraint, it'd be foolish to categorically prohibit bifurcation. After all, it sometimes happens that the ideal, namely, the unity of cognition and conation in practical thought, isn't available. Any moral objection, then, must dwell on aspects of our lives, like the political and, more generally, the practical, in which the strategies of bifurcation oughtn't to be implemented.

My focus in this subsection is the second of the two varieties of bifurcation just mentioned. I'll elaborate it via the device of self-effacement in normative theory, specifically as it figures in the consequentialist tradition. Because consequentialism purports to govern both political institutions and individual actions, it has sufficient reach to fall within the scope of the moral corollary. And

many consequentialisms (perhaps all that anyone would care to defend) are, at least, under pressure

to self-efface. Parfit elucidates the idea:

> Suppose that S [a theory of rationality] told everyone to cause himself to believe some other [rival] theory. S would then be *self-effacing*. If we all believed S, but could also change our beliefs, S would remove itself from the scene and would become a theory that no one believed.[220]

Simon Keller offers a clearer formulation, one that makes the connection to externalism explicit:

> An ethical theory tells a story about what things are valuable, about what makes acts right, about what justifies acts and gives reasons to act. An ethical theory is self-effacing, in the relevant sense, if the considerations that it posits in telling that story sometimes should not serve as motives for action, according to the theory itself.[221]

A self-effacing theory, then, denies any alignment (let alone identity) between certain of our reasons

for normative belief and the (apparent) reasons for action corresponding to those beliefs. It does

this because the content of the relevant beliefs in some sense antagonize the motivational bases of

the prescribed actions. The cases in which this kind of conflict arises are, I think, marginal, but

given certain theoretical presuppositions, they take on special importance. We've already touched

on hedonism, the view that intrinsic value is essentially pleasure. Once combined with the peculiar

features of human psychology, it generates a problem that Sidgwick calls 'the paradox of hedonism':

> A man who maintains throughout an epicurean mood, fixing his aim on his own pleasure, does not catch the full spirit of the chase; his eagerness never gets just the sharpness of edge which imparts to the pleasure its highest zest and flavour. Here comes into view what we might call the fundamental paradox of Hedonism, that the self-regarding impulse, if too predominant, defeats its own aim… of our active enjoyments generally, whether the activities on which they attend are classed as "bodily" or as "intellectual" (as well as of many emotional pleasures), it may certainly be said that we cannot attain them, at least in their best form, so long as we directly aim at them.[222]

---

[220] *Reasons and Persons*, page 24.

[221] "Virtue Ethics Is Self-effacing", page 221. Strictly speaking, Parfit's formulation doesn't lead to externalism, since he identifies the potential for a rupture within cognition, rather than between the cognitive and conative elements of practical thought. Keller's formulation owes more to Stocker's notion of moral schizophrenia (as developed in "The Schizophrenia of Modern Ethical Theories") than to Parfit's idea. Still, it's Parfit who coined the name 'self-effacement'.

[222] *Method of Ethics*, page 37. Railton is right in noting that "the "paradox" is pragmatic, not logical" ('Alienation, Consequentialism, and the Demands of Morality', page 141). It derives from contingent features of our psychology. If, counter to fact, we were constituted so that maximum pleasure is realized via the *de dicto* pursuit of pleasure, no paradox would arise.

The hedonist's normative belief—according to which only pleasure is intrinsically valuable—threatens to sap her of the motivational energy she'd have if she pursued the sources of pleasure (like friendship or beauty) described in their own terms (that is, as *friendship* or *beauty*) rather than described in terms of their expected hedonic values.[223] That is, a hedonist is unlikely to relish the pleasures of friendship so long as she registers her friend as a mere instrument to intrinsic value. If she is to pursue these pleasures with any enthusiasm—they are, after all, especially intense and enduring, so some fervour is in order even by the hedonist's lights—she must, *contra* her own hedonistic views, reconceive her friend as intrinsically valuable. But she can only do so if she renounces her hedonism, and subscribes to some rival conception, one that extends intrinsic value to certain non-pleasures, like friendship.

And so, this is what she is to do. To escape the paradox, hedonism self-effaces—it prescribes that it be renounced. Its adherents (if we can call them 'adherents') must be disbelievers.[224] Not all consequentialisms are self-effacing in this way, since not all encounter the psychological features that trigger the paradox. A money-maximizing consequentialism, for instance, could expect the same motivational support whether *de dicto* or *de re*, that is, whether it

---

[223] The hedonist's problem is more serious even than that, since the result is diminished pleasure as well as diminished motivation. But that element of the paradox doesn't lead to bifurcation since the experience of pleasure doesn't belong to practical thought (or anyway, not as the hedonist construes pleasure). The problem I mean to showcase comes through more clearly if the hedonism in question is impersonal (that is, if it's a utilitarianism), so that the prospective actor wouldn't necessarily stand to experience the pleasure she seeks to produce. This isolates the motivational problem from the experiential problem.

[224] Sidgwick's own response to the paradox amounts to something like a programme of deliberate forgetfulness, of ejecting the hedonistic conception from the conscious mind:

> We may therefore state as generally true, what has been called the Fundamental Paradox of Egoistic Hedonism, that in order to attain the end we must to some extent put it out of sight and not directly aim at it. But though this presents itself as a paradox, there does not seem to be any difficulty in its practical realization, when once the danger indicated is clearly seen. For it is an experience only too common among men, engaged in whatever pursuit, that they let the original end and goal of their efforts pass out of view, and come to regard the means to this end as ends in themselves: so that they at last even sacrifice the original end to the attainment of what is only secondarily and derivatively desirable. And if it be thus easy and common to forget the end in the means overmuch, there seems no reason why it should be difficult to do it to the extent that Rational Egoism prescribes. (*Method of Ethics*, page 133)

Perhaps this falls short of theory-effacement as Parfit understands the idea. After all, we may continue to believe the hedonist conception even if we don't observe that belief at all times. Still, Sidgwick's theory is straightforwardly self-effacing by Keller's criterion.

deploys the concept *money* in conation or not. For consequentialisms of this kind, parsimony recommends against self-effacement. But the more attractive consequentialisms, the kind that are actually developed and defended—e.g., a virtue- or beauty-maximizing theory, or a pain- or exploitation-minimizing theory—these must consider self-effacement. This is because of a sister to the paradox of hedonism, what we might think of as a paradox of maximization: the maximization project itself can sap our motivational energies. That is, we're likely better motivated to pursue beauty *simpliciter* than we are to pursue the maximization of beauty, and, to the extent that this is true, the maximization of beauty requires that we disentangle our consequentialism-directed actions from the belief that consequentialism is true.

Intrapersonal bifurcation strikes me as highly mysterious. It seems to involve an element that's analogous to self-deception, since, presumably, bifurcation is itself to be motivated by a normative belief, which must then be jettisoned from the conative apparatus it's to have produced, since otherwise it would contaminate that apparatus in exactly the way bifurcation is introduced to prevent. Davidson offers a treatment of a related problem that arises from the nature of irrationality, the lesson of which can be extrapolated to bifurcation. The difficulty is that the belief must retain its causal power while meeting two conditions: first, it mustn't exhibit a rational connection with its effect, since this would bridge the two in such a way as to make bifurcation impossible; but also, it mustn't depart from the domain of mental explanation, with its distinctively mental variety of causation (recall that the bifurcation model of externalism is distinct from the contemplative model, in that it doesn't posit a sub-personal, or wholly mechanical conative psychology).[225] The solution to this problem, as Davidson envisions it, is psychological

---

[225] See Davidson's "Paradoxes of Irrationality". He says there (at page 179) that:
> [i]n standard reason explanations … not only do the propositional contents of various beliefs and desires bear appropriate logical relations to one another and to the contents of the belief, attitude, or intention they help explain; the actual states of belief and desire cause the explained state or event. In the case of irrationality, the causal relation remains, while the logical relation is missing or distorted.

partitioning—entities within different compartments can be individuated as mental by exploiting the rationality they exhibit within their different locales, while at the same time entering into causal exchanges with one another. The idea is that mental causation across different compartments needn't manifest the rational connections that establish the possibility of mental individuation within compartments.[226]

But psychological partitioning generates a slew of new problems.[227] If we're to make sense of bifurcation, it's likely better to attend to its interpersonal, rather than its intrapersonal, instance. So, suppose we're utilitarians, and, confronted with the paradoxes of hedonism and maximization, we opt to self-efface. We could implement bifurcation at the social level by erecting institutions that delegate the task of belief in utilitarianism to a caste of legislators. Being fully aware of the content of the theory, the legislators can consult its pleasure-maximizing principle as they devise a scheme of incentives and disincentives for their subjects, and in this way model observance of that principle for the bulk of society. They can also, and likely must, contrive a noble lie so as to give the impression that the populace's *de dicto* use of normative concepts, as generated by the scheme of incentives and disincentives, embodies the correct moral principles. Thus, they ensure that their subjects achieve the requisite conviction in the scheme-engineered concepts, so that they're properly motivated by them. Unhappily, the legislators can't themselves participate in that scheme, as they are corrupted by their belief in the theory. They are instead to settle for a diminished conative capacity in order to secure the greater capacity of the ruled, who outnumber them to a significant degree.

An account of bifurcation can be modelled on an account of irrationality, like *akrasia* or self-deception, but only up to a point, since, *ex hypothesi*, bifurcation isn't irrational (of if it is, it isn't irrational in the same way). What the two share is the feature of involving explanations that are distinctly mental but that don't proceed along the rationality-delineated paths that provide the essence of the mental. (Davidson indicates that his account isn't confined to irrationality at *ibid.*, page 186.)

[226] *Ibid.*, pages 180-182.

[227] As Davidson well recognizes. He describes some at *ibid.*, pages 182-184.

It's sometimes suggested that utilitarians are compelled, given the mechanics of their theory, to be receptive to this social bifurcation strategy. Williams considers the apparent inclination for the one by the other by envisioning a society in which a "utilitarian élite" compels its non-utilitarian populace via "institutions of coercion and severe political restriction".[228] This is a source of embarrassment for utilitarianism, since social bifurcation is *prima facie* morally abhorrent. For one thing, it violates the transparency provision mentioned at the beginning of this section. It deludes the average citizen as to the moral principles that govern her life. Really, it robs her of the moral dimension of her life, that is, of the opportunity to represent and act from moral reasons.

Things aren't quite so dire for the legislative caste. But then, the truly enterprising utilitarian would endeavour to squeeze some extra motivational efficacy from the legislators by having only one generation adopt utilitarianism itself. That generation would then devise a special scheme of incentives for the legislative caste, so that, once done, subsequent generations can legislate to maintain the scheme without recognizing that this is what they're doing. This is something like what Plato's Socrates prescribes with his own vision of a noble lie; what's wanted is "one noble falsehood that would, in the best case, persuade even the rulers".[229] Of course, it would require an astonishingly complex piece of social engineering to pull this off, probably more than our species could sustain. Still, if this sort of thing could be wrung from the human being, the self-effacing utilitarian should demand it. In any case, this second arrangement is, I think, somewhat less repulsive than the initial arrangement, since at least in this case the entire population is on equal footing, with each being duped by ancestors who (we might suppose) are long dead. But the thought of an entirely deluded society—and of a moral theory that prescribes it—is sufficiently

---

[228] In *Utilitarianism: For and Against*, pages 138-139. Also: "[i]t is not surprising that one should be reminded of colonial administrators, running a system of indirect rule" (*ibid.*, page 138).
[229] *Republic*, book III, 414b-c.

horrifying to elicit the same moral objection. And if our reasoning leads us to a conclusion as obviously confused as this one, it must be because an early step was radically mislaid.

That said, it's far from obvious that utilitarianism is really susceptible to this line of complaint. After all, utilitarians are, at least typically, exemplary liberals, just as disposed to condemn social stratification and deception as their contractualist peers.[230] But we needn't pursue this point, since utilitarianism isn't the target. If no moral theorist would recommend the social arrangement I've just described, it's likely because each acknowledges that the rational animal is entitled to the opportunity to act from genuine moral reasons, and this prohibits, among other things, our intervening so as to deceive her as to what those reasons are. This is suggestive of a convergence of the metanormative and moral dimensions—that is, the immanence constraint and its moral corollary—of the rational animal's complaint.

The match isn't perfect. Nothing compels externalism to espouse bifurcation, let alone self-effacement. Thus, the metanormative objection applies more widely than the moral objection. And anyway, the rational animal's chief enemy in the context of metanormative thought is austerity, not externalism. I've focused thus far on the more tractable target—externalism is a comparatively determinate notion, and, though it doesn't say outright what its moral and political counterparts might be, it says enough to inform a guess. Structurally speaking, this utilitarian stratification is like the Straussian stratification considered earlier, and both are ultimately paternalisms of the kind the

---

[230] I'm not certain how utilitarians avoid social self-effacement, but they must, since otherwise they're met with this puzzle: our best evidence suggests that the institutions that are most conducive to pleasure (or anyway, to longevity, literacy, and other reliable markers of good quality of life) are quite a lot like the ones prescribed by the social contract tradition. But those are the very institutions that foreground a commitment to transparency. For that matter, they foreground a commitment to a culture that exalts Kant's famous maxim, his motto for the Enlightenment: "dare to use your own understanding!". But that culture is fundamentally opposed to social bifurcation.

This produces a dilemma: either utilitarianism self-effaces via social bifurcation or it doesn't. If it does, then it must give up the most effective means to realizing pleasure, so that it fails by its own maximization principle. But if it doesn't, we're returned to the original problem, the paradox of hedonism—our enthusiasm for liberal institutions would be weaker if we only pursued them in light of their potential for pleasure, so that, again, we give up the most effective means to realizing pleasure. Unless utilitarianism can dissolve this dilemma, it's self-defeating.

social contract theorists sought to defeat. What distinguishes them is their bases—the former is satisfied with *conformity to* normative reasons, rather than *action from* them, and the latter relies on radical intuitionism rather than a theory-driven picture of normative inquiry. Still, the moral corollary objects to each for largely the same reason.

3.4.2 Rawls on the unity of normative and metanormative thought

At this stage, we can identify these principles as candidate components of the moral corollary: *if a normative conception is to succeed in capturing its corresponding idea, it mustn't entail a moral or political conception that prescribes* i) *the division of rational animals into subclasses, with some who intuit, and others who don't,* ii) *anything like a noble lie, and* iii) *bifurcation.* Alternatively, the principle might demand of successful conceptions that they entail a prohibition on these things. Since some prescriptions to bifurcate are innocent, the principle must be formulated so as single out the morally pernicious instances, like the variety of interpersonal bifurcation outlined above, or the sort of wide-reaching intrapersonal bifurcation that might threaten to estrange us from political life, or from ethics.

I haven't done nearly enough to defend even the weaker formulations. But my ambition is only to provide a rough sketch. In that spirit, I'll enumerate three more candidate principles, iv)-vi):

iv) As Rawls explains, self-respect

> includes a person's sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out… self-respect implies a confidence in one's ability, so far as it is within one's power, to fulfill one's intentions… Without it nothing may seem worth doing, or if some things have value for us, we lack the will to strive for them.[231]

He also describes the social bases of self-respect, such as the circumstance of "finding our person and deeds appreciated and confirmed by others".[232] In the absence of its social bases,

---

[231] *A Theory of Justice*, page 386.
[232] *Ibid.*

self-respect is undermined, for "unless our endeavors are appreciated by our associates it is impossible for us to maintain the conviction that they are worthwhile".[233]

Rawls' concern is the role that self-respect plays in the exercise of distinctly moral powers, especially our ability to be motivated by our representations of the good. But we can also attend to the role of self-respect in practical thought. The immanence constraint is designed to determine the conceptual requisites for practical thought; we can conceive of the moral corollary as being designed to determine the social requisites of practical thought. The social bases of self-respect are, I think, particularly clear examples of those social requisites. But then, the corollary can't simply demand that the social bases be provided, since it's a constraint on metanormative conceptions, and these have no direct contact with the mechanisms that distribute the social bases of self-respect. The most a conception can do is inform a public culture in ways that contribute to or detract from the social bases. The principle I have in mind is to be formulated in that light—it prohibits conceptions that are liable to foster a culture that frustrates self-respect.

v) Rawls also says that free citizens "regard themselves as self-authenticating sources of valid claims. That is, they regard themselves as being entitled to make claims on their institutions so as to advance their conceptions of the good (provided these conceptions fall within the range permitted by the public conception of justice)."[234] The relevant contrast is with

---

[233] *Ibid.*, page 387. This is an empirical claim about human psychology, but one we can't reasonably deny. Unlike the immanence constraint, the moral corollary arises downstream of our conception of *rational animal*, after "practical anthropology", as Kant uses that term (*Groundwork*, page 388), is brought to bear. The corollary expresses the rational animal's complaint specifically as embodied by the human animal. (That said, the triangulation argument shows that sociality, in some form or other, is an essential feature of rationality. There is no possibility of a rational solitaire.)

[234] *Political Liberalism*, page 32. When Rawls first introduced the idea in "Kantian Constructivism in Moral Theory", he used the somewhat more sensational phrase "self-originating sources of valid claims" (at page 330). I suspect that what motivated the change is that the original 'self-originating' formulation gives the impression of the spontaneous generation of normative entities *ex nihilo*, as though by way of a radical choice. Constructivism is at times conflated with existentialism (of the Nietzschean or Sartrean variety), but Rawls indicates (at *ibid.*, page 355) that the two are distinct:

circumstances in which subjects' "claims have no weight except insofar as they can be derived from the duties and obligations owed to society, or from their ascribed roles in a social hierarchy justified by religious or aristocratic values."[235]

The idea is related to the point just made about self-respect—we must be equipped with a self-conception according to which we can be motivated by our own conceptions of the good. That an agent represents the good in a certain way must itself be given some degree of normative authority to determine her action.[236] This is a conceptual matter, derivable from the thin formulation of the immanence constraint. After all, if the rational animal isn't construed as self-authenticating, she couldn't act from her representations of normative entities, and practical thought would be impossible. This shows self-authentication to belong to the normative psychology underlying the idea of *practical thought*. (And, if the triangulation argument is right, it's not enough that the animal construe herself in this way; at least one other must do so as well.)

---

[t]he agreement of the parties in the original position is not a so-called "radical" choice: that is, a choice not based on reasons, a choice that simply fixes, by sheer fiat, as it were, the scheme of reasons that we, as citizens, are to recognize, at least until another choice is made. The notion of radical choice, commonly associated with Nietzsche and the existentialists, finds no place in justice as fairness [i.e., in Rawls' constructivist conception].

Korsgaard is also opposed to existentialist conceptions of agency; she represents the moral essence as prior to agential existence, as it were. But Street's Humean constructivism assigns an essential role to radical choice (see "Constructivism about Reasons" 237-238), so some constructivisms are, to this extent, existentialist.

  At any rate, the self-authenticated claims derive from citizen's comprehensive doctrines, which in their different ways participate in the pursuit of the good. They must be interpreted in terms of their performing that role. (Some comprehensive doctrines are existentialist, or in some other sense nihilistic about *the good*, but they're nonetheless to be counted as conceptions of *the good* since they occupy that same doctrinal position in citizen's lives.)

  The moral consequences of self-authentication, by contrast, *are* generated *ex nihilo*. For instance, the moment that someone opts, on the basis of her reflections on the good, into a Christian conception is the moment that those around her are obligated to enable her to attend church. The basis of that obligation is that she conceives of *the good* in the way she does, and not the good itself, so that it obtains whether her conception is true or not.

[235] *Political Liberalism*, page 33.

[236] The Kantian point that we act from the "*representation* of laws" (*Groundwork*, page 412), rather than conforming directly to the moral law, is also relevant here. If the moral law were such that it caused us (in the mechanical sense of 'cause') to conform to it, there would be no acting from it, and so, no good will. And if a social authority discerns the moral law, and legislates incentives that generate behaviour that conforms to it, again there would be no acting from it, and again, no good will. This is an important point of convergence between his moral conception and his representation of the Enlightenment—the good will is only possible given liberty of conscience.

But we can add that, as a moral matter, her representations are owed the social standing that Rawls finds for them, namely, the good faith exercise of an animal's self-authenticating power is entitled to moral protections, and, in certain cases, support. This is true even when those charged with protection or support believe, or even know, that the relevant claim is mistaken. While some claims are so confused that we're permitted to ignore them, or required to antagonize them, these cases are delineated by morality itself (as Rawls says, they fall outside the range permitted by the public conception of justice); the rationale for neglect or intervention is not that a pursuit would be mistaken, but that it would be immoral. Metanormative conceptions must leave room for these obligations, and in particular, for the fact that they derive from animals' conceptions, rather than from what those conceptions are about.

vi) The last principle is owed to a third Rawlsian idea, namely, the burdens of judgment. His ambition in developing this idea is to explain how reasonable people can disagree over long spans of time without dishonesty, inattention, or stupidity. His answer appeals to the many different ways that the normative domain is marked by complexity, so that we're liable to fall into error in our thinking about it, even when that thinking is otherwise impeccable (or anyway, not obviously confused).[237] But the lesson the burdens of judgment convey is broader: we must never conceive of a rational animal as normatively hopeless. This isn't to say that we must never throw up our hands, and abandon our efforts to persuades others. Certainly, some are too stubborn to convince; others aren't stubborn, but are, in a manner of speaking, locked in to a conception that we regard as (or know to be) wrong. But, as Rawls explains, "[i]t is unrealistic—or worse, it arouses mutual suspicion and hostility—to suppose

---

[237] *Political Liberalism*, pages 55-58.

that all our differences be rooted solely in ignorance and perversity, or else in the rivalries for power, status, or economic gain."[238]

Those differences must be explainable within the space of good faith rational discourse, the space where ordinary mistakes are made (as always, such mistakes are only intelligible against a backdrop of widespread rational success). They mustn't be explained by appeal to deep-seated—that is, extra-discursive— defects of character or judgment. The notion of a radical mistake, one that could show a person to be normatively hopeless, and that could be invoked in order to eject her from discourse, must presuppose something like radical intuition, i.e., an extra-discursive capacity that, it is alleged, the hopeless (though rational) person doesn't have. The radical intuitionist might, for instance, suppose that the relevant capacity is developed over a highly specific course of ethical education, and that there isn't any possibility of rehabilitation for those who underwent a different course. The result is that practical thought is confined to the privileged few, while the rest are, at best, made to answer to some extra-rational instrument instead, like the truncheon of the state.[239] We've seen that this leads to stratification, but even prior to that, it exhibits a moral failure in its depiction of rational agency. It's in the nature of practical thought that it's amenable to rational discourse, specifically to the kind of discourse to which any creature capable of speech is permitted entry. Metanormative conceptions must honour the moral egalitarianism that inheres in that discursive space.

---

[238] *Ibid.*, page 58.
[239] Read one way, this is Aristotle's conception:
> [i]f arguments were in themselves enough to make men good, they would justly, as Theognis says, have won very great rewards, and such rewards should have been provided; but as things are, while they seem to have power to encourage and stimulate the generous-minded among our youth, and to make a character which is gently born, and a true lover of what is noble, ready to be possessed of virtue, they are not able to encourage the *many* to nobility and goodness. For these do not by nature obey the sense of shame, but only fear, and do not abstain from bad acts because of their baseness but through fear of punishment… a bad man, whose desire is for pleasure, is corrected by pain like a beast of burden (*Nicomachean Ethics*, 1179b5-1180a12).

The list isn't exhaustive (for that matter, I'm not certain that each of its entries is distinct). As for the unity of the components of the moral corollary, the best I can manage at present is to gesture to contractualism itself. We've seen that, according to contractualism, the legitimacy of institutions is conditioned by the possibility of rational consent, together with the material opportunity to manifest one's rationality in consent, as dictated by the transparency provision, among other things. The contractualist's conception of *moral equality* takes its start from this equality in political authority, with each person operating as a check on institutional power. This is what rules out stratification and deception, and demands each of iv), v), and vi), which we can gloss as dealing in the means to choose, the moral status of that choice, and respect for those who disagree.

This is one way Rawls can be used to establish continuity between first- and second-order normative thought. But then, Rawls' contractualism, being confined to political justice, is too narrow to unify the corollary; Scanlon's contractualism is a better fit. The overarching principle of the corollary, then, is given by his conception of *moral wrongness*: "an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement."[240] Again, the idea isn't that the content of the corollary is proven by the argument for the immanence constraint, but rather that it's the natural elaboration (on a rather loose sense of the term 'elaboration') of that argument.

In order to complete my treatment of the moral corollary, I must address the third of the three complications mentioned earlier, the one according to which the 'meta' in 'metanormative conception' grants austerity immunity from first-order (in this case, moral) critique. My response appeals to a second way that Rawls points to continuity between first- and second-order normative thought—in a passage I quoted earlier in subsection 2.3.4, he says that "justification is a matter of

---

[240] *What We Owe to Each Other*, page 153.

the mutual support of many considerations, of everything fitting together into one coherent view."[241] A consequence of this idea is that theories are to be assessed as a whole—our assessments of metanormative conceptions should include the assessment of their moral and political consequences. Understood this way, blunders at the level of first-order morality can indict the whole picture, including its metanormative component. This is true even where our substantive moral claims are deduced from ontological and epistemological foundations. Of course, the reverse is also true, namely, that substantive moral conceptions are answerable to metanormative conceptions. The idea is that the relation is symmetrical—just as a compelling metanormative theory can determine our first-order moral theories, so too can a compelling first-order theory determine our metanormative theories.

Rawls' efforts to curb the influence of conceptual analysis in moral philosophy are relevant here. On his proposal, "[d]efinitions and analyses of meaning do not have a special place: definition is but one device used in setting up the general structure of theory. Once the whole framework is worked out, definitions have no distinct status and stand or fall with the theory itself."[242] A superficially persuasive bit of analysis might be defeated by the recognition that it leads to a pernicious moral conception. So, for instance, a moral theory prescribing the misery of the few as a means to the ecstasy of the many can't be screened against moral objection by its appearing to follow from analysis of the word 'good'. Instead, the failures of the former serve as clues to the failures of the latter. Likewise, success at the level of moral theory may elucidate some aspect of metanormative thought. This is what the moral corollary aims to do.

I'm not certain how to argue for Rawls' position on this issue. I can gesture at the poor track record of the rival view, the one according to which ethics and metaethics are discrete subject

---

[241] *A Theory of Justice.*, page 19.
[242] *Ibid.*, page 44.

matters, with the conclusions of the one insulated from the conclusions of the other. This view arguably arose alongside analytic philosophy, and was further entrenched by verificationism, and then ordinary language philosophy. These glossed metaethics as linguistic analysis of ethical language. The achievements of such analyses are, at best, suspect. While Ayer's verificationism came and went (in the academy, though not the public culture), his method of analysis dominated for several decades. It wasn't until the sixties, and especially the seventies, that its reign ended (which isn't to say it was defeated; it endures to this day). Rawls' contribution, in publishing *A Theory of Justice*, to the project of displacing linguistic analysis as the dominant method in moral philosophy is significant.[243] And, crucially, this helped open the moral philosophical terrain to a wealth of different methods, which reinvigorated the discipline.

That's the best case I can make—Ayer's (and his descendants') reign was stultifying and anaesthetizing; moral philosophy post-*A Theory of Justice* was, at least for a time, remarkably fecund. This isn't a particularly strong argument, since it isn't clear that Rawls' position contributed to the shift, and even if it were, it isn't clear how that could show his position to be true. On the other hand, I can't see what could be said for the rival view; perhaps, for a time, it seemed as though controversies are intractable so long as they remain at the first-order, and that the only avenue for anything approximating consensus is via a retreat to the second-order. My impression is that history has since shown both that the first-order controversies aren't so intractable as they initially appeared, and that consensus at the second-order isn't any easier to secure than it is elsewhere. At any rate, it's my view that placing moral theory and metanormative thought on a par results in both better moral theory and better metanormative thought.

---

[243] Anscombe's "Modern Moral Philosophy" struck the first blow in 1958, thirteen years before the publication of *A Theory of Justice*. But it took quite a lot of time for the tide to turn.

This idea can only be pursued so far. Substantive moral considerations can condition metanormative thought because the two overlap to such a degree that it's natural to package them together in a single theory. Still, they contribute to normative thought in very different ways, so the distinction between them ought to be preserved. Nagel describes Ronald Dworkin's "constant insistence that the only way to answer skepticism, relativism, and subjectivism about morality is to meet it with first-order moral arguments. He holds that the skeptical positions must themselves be understood as moral claims—that they are unintelligible as anything else."[244] Nagel retreats somewhat from this position: "I would not go so far as that, but I have been led to the view that the answer to them must come from within morality and cannot be found on the metaethical level."[245] Perhaps even this is too strong; it depends on how the distinction between morality and metaethics is cashed out. My preference is to follow Rawls by, first, representing moral theories as comprising both metaethical and substantive accounts, and, second, evaluating the whole.

Moreover, the discoveries of wholly independent disciplines must be accepted as fixed, no matter how grizzly they may be. Certain notions underlie the possibility of moral theory itself, so that, for instance, substantive moral truths can't operate as a check on the idea of *truth*. Nor can they operate as a check on the findings of the hard sciences. (I think, though, that the soft sciences have penumbrae that morality may enable us to navigate. The theory of human nature, for instance, isn't entirely autonomous from moral theory.) Finally, the objection embodied by the moral corollary needn't be decisive, or even particularly forceful. Probably, it can only defeat an especially flimsy metanormative conception. But, as always, flimsy conceptions abound, so that there remains plenty for the corollary to do.

---

[244] *The Last Word*, preface.
[245] *Ibid.* Nagel may mean that some apparently metaethical considerations are actually concealed (or contorted) first-order positions, and that they ought to be revealed as such in order to be engaged in their proper context. I'm sympathetic to that view.

4. CONSTRUCTIVE PRIMITIVISM

Three tasks remain to be completed: first, establish the possibility of a metanormative conception that can meet the immanence constraint on its thick formulation without in this way violating the transcendence constraint; second, plot the boundary between austere and constructive primitivism; and third, demonstrate that Scanlon's conception is an instance of the latter rather than the former. But these three tasks reduce to one, namely, the specification of constructive primitivism.

Constructive primitivism is simply primitivism that's constructive. That is, the relevant idea is given by its two constituent ideas, *primitivism* and *constructiveness*, together with the principle for their combination. I'll start with the idea of *primitivism*. I intend that my use of 'primitivism' align with Scanlon's use of 'primitive' at the opening of chapter 1 of *What We Owe each Other*. Thus stipulated, 'primitivism about α' is interchangeable with 'non-reductionism about α'; the former says positively what the latter says negatively. We might be inclined (or anyway, I'm inclined) to hear the word 'primitive' as suggesting more than this, a kind of ontological heft, or a special claim to objectivity. But this would rule out quasi-realist primitivism about α, i.e., the view that the truth of claims made about α have no ontological entailments. My sense is that Scanlon wouldn't want to exclude quasi-realism at this stage, since, at least ontologically speaking, his own realism is a near neighbour.[246] He would allow, I think, that primitivism about α requires that α be instantiated, but only on the condition that the relevant sense of 'instantiated' is a soft one.

As I've already indicated, irreducibility is to be understood in terms of circularity. Thus, analysis of primitives is necessarily circular; the concepts with which such analyses begin invariably reappear later on. If we discover that an idea has a gap in its circle, from which it can be given in

---

[246] *Being Realistic about Reasons*, page 52. He says at *ibid.*, page 24, that "our ontological commitments in this general sense do not represent a claim on our part about what *the* world contains, in any meaningful sense of "the world."" I registered my dissatisfaction with this kind of non-metaphysical realism in note 78.

altogether different terms, this rules it out as a candidate for primitivism. Strictly speaking, any variety of irreducibility will do, though our concern all this while has been irreducibility as a consequence of *sui generis*-ity. The connection is obvious enough—if something is unlike everything else, it follows that it can't be explained in terms of anything else, not unless its *sui generis* elements, in a manner of speaking, accompany the explanation. Normativity is characterized by several *sui generis* elements, one of which being the transcendent capacity that underlies the transcendence constraint. This is why normative conceptions must be non-reductive if they're to have any hope of meeting that constraint.

Constructiveness is a much simpler idea. A conception is constructive if it attempts to explain. But, as I warned in chapter 2, we must take care that our use of the term 'explain' doesn't activate any deep-seated prejudices. The spectre of the scientistic monolith looms over normative *sui generis*-ity, as does the philosopher's impulse to homogenize, to buy ontological unity cheaply via brute assimilation. In order to guard against these things, we must widen our conception of *explanation* so that it can recognize explanatory circles. This wider conception of *explanation* is what enables us to tie *constructiveness* and *primitivism* together in a single conception.

I've already touched on how the conception is to be widened in subsection 1.3.5: within the philosopher's toolkit, there is, in addition to reductive explanation, explanation by way of systemic connection. Any reasonably comprehensive conception develops many different pieces—e.g., views about the idea of *normative reason*, and about other ideas that arise in its neighbourhood, like *value, wrongness, moral motivation*, etc.—with each one bound together in a system. Each piece has some local set of considerations to recommend it, a distinctive set of problems that it solves. But being a component of a system, it must also cohere with the rest of the solutions offered throughout that system. And what harmony there is among the different pieces can play its own explanatory role. So, for instance, it can happen that some arguments are only decisive in virtue of their fit with

arguments given elsewhere. This shows that there's the potential for explanatory power residing in how the pieces snap into place alongside one another, as in a jigsaw.

This variety of explanation should be familiar—it's P. F. Strawson's connective model of philosophical analysis. I present that model in the next section. And then, with the idea of *constructive primitivism* elucidated, and the path to the joint satisfaction of the immanence and transcendence constraints charted, I turn to the classification of our two non-naturalists, Moore and Scanlon. I've maintained all this while that Moore is an austere primitivist; I defend this claim in section 4.2. And then, with Moore as foil, I make the case for interpreting Scanlon as a constructive primitivist in section 4.3. Finally, in section 4.4, I consider Scanlon's prospects for meeting the immanence constraint.

4.1 STRAWSON'S CONNECTIVE MODEL

Strawson's remarks on analysis establish the possibility of philosophical explanation that's non-reductive, which, in turn, establishes the possibility of constructive primitivism. But before turning to those remarks, I'll say a bit about the provenance of the relevant sense of the term 'constructive'. In the course of Verheggen's characterization of Davidson's semantic non-reductionism (which was canvassed in section 2.3), she observes that this non-reductionism "is particularly noteworthy because it is not quietist", that is, "[i]t clearly demonstrates that constructive theorizing about meaning and content remains possible even though reductive explanations of them do not."[247] We've already seen that Davidson's semantic conception is non-reductive—the triangulation argument is plainly circular, since it only establishes the possibility of language by positing linguistic communication, so that the two, the phenomenon and its condition, arise simultaneously. And yet,

---

[247] *Donald Davidson's Triangulation Argument*, page 2. Quietism is a species of austere philosophy; I say more about it in section 4.4.

his view is constructive in that it shows, among other things, that sociality is a prerequisite for thought—hardly a trivial result.[248] In this way, it carves out a middle position between the austere and reductionist extremes. That middle position is constructive primitivism.

This is the model that metanormative theory should emulate. Verheggen expands on the possibility of constructive semantic philosophy: "[i]f intensional concepts cannot be explained in a reductive way, it seems that one promising alternative explanation is in terms of the relations among them."[249] The normative primitivist ought to pursue this same alternative to reduction, i.e., she ought to develop her conceptions of normative concepts like *normative reason* by situating them in relation to other such concepts in their shared lattice of ideas.

The word favoured by Strawson is 'connective' rather than 'constructive'. I've opted for 'constructive' because it's more evocative of substantive theorizing.[250] In any case, the two aren't synonyms, since, unlike connectiveness, constructiveness doesn't discriminate between reductive and non-reductive conceptions, so long as they attempt to explain. For that matter, causal explanation (that is, explanation that depicts *explanans* and *explanandum* as wholly distinct entities) qualifies as constructive. Connective explanation is the more discriminating category, being both essentially a matter of philosophical analysis, and essentially non-reductive.

To see what it entails, it's worth quoting Strawson at length. He begins with an account of the model against which his connective conception is contrasted:

> the most general implication of the name ['analysis'] seems to be that of the resolution of something complex into elements and the exhibition of the ways in which the elements are related in the complex… If we took this notion completely seriously for the case of conceptual analysis—analysis of ideas—we should conclude that our task was to find ideas that were completely simple, that were free from internal conceptual complexity; and then to

---

[248] See Davidson's "Externalisms", page 13.

[249] *Donald Davidson's Triangulation Argument*, page 35.

[250] The thought had occurred to me that the term 'constructive' is clumsy in the context of metanormative philosophy, where it's liable to be confused for 'constructivism', one of the discipline's major players. But the precise meaning of 'metanormative constructivism' is still, I think, up for grabs, and the case might be made that Verheggen's sense of 'constructive' better conveys what thinkers like Rawls and Korsgaard are up to than the image of a rational procedure or of a practical point of view (cf. Street's "What Is Constructivism in Ethics and Metaethics?").

demonstrate how the more or less complex ideas that are of interest to philosophers could be assembled by a kind of logical or conceptual construction out of these simple elements. The aim would be to get a clear grasp of complex meanings by reducing them, without remainder, to simple meanings.[251]

I dwell on this passage because, as a founder of the analytic tradition, Moore contributed to the development and propagation of this variety of analysis, the reductive variety. In a moment (subsection 4.2.2), I'll represent Moore's predilection for austerity as an unfortunate consequence of his view that all ideas are either simples or complexes, i.e., arrangements of simples, with analysis confined to the latter. Recognizing the *sui generis*-ity of normativity, he assigns it to the former category, and promptly gives up the project of elucidating it. This is because he supposes that all analysis is reductive, so that philosophy must stop dead the moment it encounters anything irreducible.

Strawson continues:

there is a certain form of words which the analytical philosopher hates to hear and which his opponent in argument, also an analytical philosopher, delights to pronounce: viz. the words, 'Your analysis is circular'. This means, of course, that included in the elements of his analysis, though perhaps covertly included and only to be revealed by further steps of some kind, is the very concept which the philosopher is claiming to analyse.[252]

Strawson then asks "[n]ow why should this formula be felt to be so damaging?" I take my question—why should primitivism be presumed to entail austerity?—to ask the same thing, or anyway, something very near to it. The answer in both cases is that reductive analysis is (or rather, was, though the echoes continue to resound) popularly taken to be what Moore took it to be, namely, the sole method of philosophical elucidation. The result is the thought that, in the context of philosophy, a failure to reduce is equivalent to a failure to explain. In order to defeat the presupposition that philosophy can only reduce or say nothing, we must identify an alternative. Thus, Strawson present his connective model:

---

[251] *Analysis and Metaphysics*, pages 17-18.
[252] *Ibid.*, pages 18-19.

[l]et us abandon the notion of perfect simplicity in concepts; let us abandon even the notion that analysis must always be in the direction of greater simplicity. Let us imagine, instead, the model of an elaborate network, a system, of connected items, concepts, such that the function of each item, each concept, could, from the philosophical point of view, be properly understood only by grasping its connections with the others, its place in the system—perhaps better still, the picture of a set of interlocking systems of such a kind. If this becomes our model, then there will be no reason to be worried if, in the process of tracing connections from one point to another of the network, we find ourselves returning to, or passing through, our starting-point. We might find, for example, that we could not fully elucidate the concept of knowledge without reference to the concept of sense perception; and that we could not explain all the features of the concept of sense perception without reference to the concept of knowledge. But this might be an unworrying and unsurprising fact. So the general charge of circularity would lose its sting, for we might have moved in a wide, revealing, and illuminating circle.[253]

This image of wide circles, in contrast to narrow ones, is useful. Arrange primitivisms according to the size of the circles they draw, from the vanishingly small circles of austere conceptions on one end, to the enormous circles charted by the great philosophical systematizers on the other; a threshold, past which the immanence constraint raises no complaint, appears somewhere in the middle. Moore's circle is maximally narrow—'good is good' draws a zero-step circle, one shrunken to a point. Scanlon's 'lead back to the same idea' claim, the one I quoted in section 1.1, can be read (uncharitably, as I will soon argue) as painting his circle as one-step, one that passes from the idea of *normative reason* through the *counts-in-favor-of* relation, and back again. One step is more than none, and to this extent there's progress. Still, we have yet to pass the threshold to the right side of the immanence constraint. As Strawson says, "[s]ome circles are too small".[254] In those cases, the charge 'your analysis is circular' is damning after all.

Meanwhile, Myers' Davidson-inspired metanormative conception is stretched across an account of conative content derived from the triangulation argument and Davidson's notion of anomalous causation. It touches on issues in normative ontology, action theory, normative psychology, the theory of agency, and first-order moral theory. This makes his circle quite a lot

---

[253] *Ibid.*, pages 19-20.
[254] *Ibid.*, page 20.

wider. At this stage, we can no longer count steps, but clearly this last circle offers enough to be properly classified as constructive. Plausibly, it offers enough to satisfy the immanence constraint. This is no accident—Myers indicates that his ambition is to "make the circle here as broad and as welcoming as possible".[255]

Strawson's treatment of connective analysis doesn't touch on the potential for explanation via harmony. This comes through more clearly in Rawls' suggestion that, ultimately, a philosophical view is justified by the mutual support of its many parts. This is what licenses the transmission of explanatory power across the view *qua* system. If one piece, e.g., a moral conception, succeeds in its milieu, and if its elaboration requires a specific commitment elsewhere, like in normative psychology, then this gives us a reason to accept that commitment. Of course, it's a question of judgment how far this explanatory power extends. But then, we're already in the business of issuing local judgments, so why not global judgments as well?

4.2 MOORE'S AUSTERE PRIMITIVISM

All this while, I've classified Moore's metaethics as a primitivism. This generates a wrinkle, which I will gloss over: following Scanlon, I've depicted primitiveness as a feature of ideas. But the subject of Moore's primitivism, goodness, isn't an idea; it's a property.[256] I take for granted that some conversion scheme can be implemented to bring these two—idea and property—in line. I recognize though that the passage from properties to ideas is likely a long one.

The pressing question is whether or not Moore's conception is, as I've maintained, austere. My case turns on his comments in §6 of the *Principia Ethica*, the ones that state that goodness is

---

[255] "Replies to Kirk Ludwig and Paul Hurley," page 259. Verheggen reports this same circle-widening ambition (see *Donald Davidson's Triangulation Argument*, pages 34-35).

[256] Even this is a simplification. At times—e.g., *Principia Ethica*, §5—Moore's focus would seem to be the term 'good' rather than the property goodness.

unanalyzable. There's a second wrinkle here, one that must be addressed: if we disregard §6 and a few other passages, the case for classifying Moore's primitivism as austere rather than constructive is significantly weakened. This is due to a tension in the *Principia Ethica*: many of its ideas suggest avenues for the elucidation of goodness, avenues that can't be reconciled with the claims made in §6. I've opted to interpret the text so that it foregrounds its unanalyzability claim, and to interpret that claim so that it denies that any elucidation is possible. I take myself to be permitted to do this because the book's legacy for the most part revolves around its claim that goodness can never be an object of meaningful inquiry. Though it presents a number of other important ideas, like *value-pluralism* and *organic unity*, these are secondary to its endorsement of austerity.

All the same, it's worth keeping the possibility of a constructive Moore in view. As I explain in subsection 4.2.2, if Moore's conception is austere, its commitment to that austerity is unstable. This instability presents the inheritors of Moore's thinking with a choice: unadulterated austerity or constructive primitivism. The latter option involves a more substantive revision of Moore's position. Still, it's available to be pursued should we wish to salvage what we can from Moore after compensating for his failure by the immanence constraint.

4.2.1 Moorean austerity

As I said in subsection 2.1.3, austerity places the strictest limit that can be managed on discourse. I gave it this schematic formulation: an austere conception of α is constituted by the three claims 'there's a thing', 'it's α', and 'that's all we can say about it'. Strawson provides a clue as to where we can find proponents of austerity. We need only venture to the early days of analytic philosophy, when the reductive method of analysis, the one that gave the discipline its name, was given a monopoly on philosophical explanation. Such a wide-reaching reductionism must accept either infinitism or foundationalism; the first analytic philosophers opted for the latter. This is what led

them to endorse the idea of a *simple*, the putative terminus of reduction. Bertrand Russell gave it the name 'logical atom', the "last residue in analysis";[257] the early Wittgenstein called it 'object'.[258] But our concern is normativity—in this context, Moore's conception of goodness as presented in §6 of the *Principia Ethica* is the archetypical austere view. "Good", he says, "is one of those innumerable objects of thought which are themselves incapable of definition, because they are the ultimate terms by reference to which whatever *is* capable of definition must be defined."[259] This effectively characterizes 'good' as an atomic predicate.[260]

Moore's bewitchment by the methods of early analytic philosophy explains how he made the mistake of espousing metanormative austerity.[261] While a part of the trouble is due to a non-reductive impulse overshooting its mark—it often happens that, in her pursuit of a crisp articulation, the non-reductionist overstates her non-reductionism and thereby veers toward austerity—the larger problem is Moore's belief in simples and in the narrow conception of *analysis* as reduction. He infers from the thought that only simples resist analysis to the thought that, if goodness is irreducible, it must be a simple. And since there's nothing to say about simples—apart from the feeble, discourse-ending 'it's there'—there's nothing to say about goodness.

As a non-reductionist, all Moore needs is a prohibition on analysis that ventures beyond the normative domain. But what his austere primitivism delivers him to is a prohibition on analysis

---

[257] *The Philosophy of Logical Atomism*, page 3.

[258] "Objects are simple… Every statement about complexes can be resolved into a statement about their constituents … Objects make up the substance of the world. That is why they cannot be composite" (*Tractatus Logico-Philosophicus*, 2.02-2.021).

[259] *Principia Ethica*, §10; see also §§7-8. The *Principia Ethica* was published fifteen years prior to Russell's lectures on logical atomism (and nearly two decades prior to the publication of the *Tractatus*). But Russell likely already subscribed to the idea at the turn of the century. I suspect it was in circulation in Cambridge before the *Principia Ethica* was published. But as to whether Moore or Russell is the source, I can't guess.

[260] Or it does provided that we're permitted to move from the property goodness to the corresponding semantic entity 'good' in the way Moore likes to do.

[261] Of course, that he was bewitched does nothing to exonerate him, especially given that he's one of the progenitors of those very methods. Incidentally, Kant proved that the relevant idea of *simple* is unintelligible in his arguments against Hume's empiricism a century earlier. (Granted, few at the time seem to have understood Kant's argument, but then, it may well have been Moore who set that trend.)

simpliciter, even within the normative domain. This is why he doesn't recognize the possibility of

analyzing normative complexes so as to uncover their relations to other normative complexes. It's

what prevents him from, say, exploiting his consequentialism in order to characterize the axiatic in

terms of the deontic, as in '*goodness* is *that which ought to be maximized*'. On his view, the relation of

analysis running between the two is asymmetric—the axiatic grounds analysis of the deontic, but not

vice versa, because the axiatic is simple, and the deontic complex.[262] There's no prospect of

elaborating on goodness by inquiring into its relation to the rational animal for this same reason.

Moore says that goodness is both simple and intrinsic, simple in that it has no essential parts,

and intrinsic in that it stands in no essential relations.[263] This has the look of ascribing substantive

features to goodness, but in fact it ascribes nothing. As they're used here, the terms 'simple' and

'intrinsic' are only markers for the absence of features. That is, the sentence 'goodness is simple'

doesn't identify a real property of goodness, since simplicity is a privation, and not a genuine

property. We can guard against the temptation to reify simplicity and intrinsicality by translating

sentences that predicate them into sentences that negate a predication. Take, for example, the claim

'austere primitivism fails to meet the immanence constraint'. The predicate—'fails to meet the

immanence constraint'—doesn't assign the feature *doesn't-meet-the-constraint*. It signals the absence of

the feature of meeting it. The 'not' is properly placed outside, rather than inside, the assignment of

the feature, as in 'not *meets-the-constraint*', which produces 'it isn't the case that austere primitivism

---

[262] This primacy assigned to the axiatic is arbitrary. Moore's open-question argument ought to block analysis of the deontic in just the same way it purportedly blocks analysis of the axiatic (as he later acknowledges in *Ethics*). But Moore had to relax his austerity *somewhere*, since otherwise he'd be left without any means of accounting for the fact that axiatic entities prescribe behaviour, that is, that they give rise to reasons.

[263] I've represented Moore's conception on the scheme of a conception of α according to which α is simple and intrinsic, rather than of a conception of *intrinsic* α according to which intrinsic α is simple. I believe Moore prefers this mode of representation (see "The Conception of Intrinsic Value"). As I read him, he denies that instrumental goods are genuine goods, except in the derivative sense of pointing to the goods that they might bring about.

An idea is intrinsic if its instantiations have no essentials relations. The notion of an intrinsic property is more discriminating than that of a monadic property, since a property might have an essential relation without being analyzed as such. For instance, the largeness of a large cat is extrinsic in that it relies on facts about the ordinary size of cats (her largeness would, as it were, disappear, if she were the only thing in the universe).

meets the immanence constraint'. The same applies to 'austere ideas are simple' and 'austere ideas are intrinsic'—'simple', as Moore uses that term, can be translated to 'not essentially complex', and 'intrinsic' can be translated to 'not essentially related'. Consequently, the claim 'austere ideas are simple' doesn't map 'simple' onto 'austere ideas'. It expresses the negation of 'austere ideas are essentially complex'. The terms with which contemporary metaethicists catalogue the other features of Moore's conception of goodness—'unanalyzability', 'irreducibility', 'non-naturalness', etc.—are more transparently privative, in that they wear their negations on their surfaces.

Incidentally, 'non-natural' is a suggestive word, and it does a fine job of tracing the outlines of a number of important metanormative positions. Still, there are contexts in which it's misleading. Often, '*sui generis*' conveys the idea better, especially in Moore's case, since the crux of his view is that goodness is "simply different from anything else".[264] At any rate, on Moore's usage, to characterize a thing as 'non-natural' isn't to project it into some occult or spectral realm, beyond nature. Even if there were such a realm, it would be excluded by the relevant sense of 'non-natural' in just the same way the physical realm is excluded. Moore's 'naturalistic fallacy' behaves this same way—it isn't confined to naturalistic reduction. The idea is that every identity theory, even one that attempts to reduce goodness to a supernatural entity, is defeated by it.[265] I dwell on this because, if there were an ontological dichotomy according to which all entities are distributed either into the natural domain or the supernatural domain, the negation of the one ('α is non-natural') would be the affirmation of the other ('α is supernatural'). But Moore's sense of 'non-natural' refers to a true privation, one that can't be redeployed as an affirmation.

---

[264] *Principia Ethica*, §10. Note, though, that in calling a thing 'natural' we typically mean that it participates in the causal nexus. This suggests that non-naturalness is a matter of exclusion from that nexus. Many understand the term in something like this way. But causality isn't the exclusive dominion of the posits of the hard sciences; normative entities can have causal powers too (as Davidson demonstrates in "Mental Events"). So, *sui generis*-ity doesn't entail causal impotence, and as such, it can't do all the work that non-naturalness is made to do.

[265] Korsgaard makes a related point in "Realism and Constructivism in Twentieth-Century Moral Philosophy", page 306.

To say of goodness that it's indefinable is only to close an avenue for metanormative thought. Remarkably (and, I think, provocatively, even irritatingly), Moore relishes the resulting obscurity. This can be seen particularly clearly in the passage I quoted in section 1.2: "[i]f I am asked "What is good?" my answer is that good is good, and that is the end of the matter. Or if I am asked "how is good to be defined?" my answer is that it cannot be defined, and that is all I have to say about it."[266] He is unmoved by, or anyway unyielding to, the efforts of the rational animal to discern goodness. On his conception, her questions are misplaced, and so appropriately deflected. Really, the tautology 'α is α', employed as it is here—as a mock answer, a rhetorical device with which to disengage from a philosophical exchange—might be used as a slogan for austerity.

Moore's adherence to the reductive conception of *analysis*, with its framework of simples and complexes, strongly suggests that his primitivism is austere. But it's the §6 passage, and in particular its 'that is the end of the matter' and 'that is all I have to say about it', that confirms it. Still, in order to complete the argument, we must introduce a last piece: Moore needs a way to replace 'α' in the schematic claim 'α is α'. That is, he needs a way to specify his primitive. The issue is best framed epistemologically, though its upshot is metaphysical—if Moore is, by his own lights, warranted in saying of his primitive that it's goodness, it must be (or anyway, it's reasonable to suppose) that he's spotted something in it, a detail that reveals that this is what it is. And this detail must, we might suppose, provide material for an additional claim to be made about it, with the consequence that Moore's primitivism isn't austere after all.

Reasonable though this expectation may be, Moore doesn't feel any pressure to meet it. This is because he's enlisted the traditional ally of austerity, radical intuitionism.[267] Armed with such

---

[266] *Principia Ethica*, §6.

[267] As I mentioned in section 2.2, Moore doesn't explicitly endorse radical intuitionism. I've attributed it to him because it's entailed by his metaphysical conception. There's certainly room to resist this attribution; the textual evidence is, I grant, quite flimsy. Still, I think radical intuitionism makes the best sense of Moore's claims, so that the attribution is, in the end, charitable rather than uncharitable.

a view, he needn't bother to offer anything like bases for his claims. That he has a primitive and that it's goodness, these are things he simply knows. When I first introduced the idea of *radical intuitionism*, it was to explain how austere primitivism might meet the thin formulation of the immanence constraint. But it also serves this second purpose, namely, establishing an epistemic context in which warrant can be given without thereby providing any material for inquiry. The first and second of the three claims that constitute austere primitivism—the existential claim 'there is something', and the individuating claim 'it's α'—are made within this context. Thus, they don't supply the starting-points for discourse that ordinary existential and individuating claims supply. The last claim—the austerity-announcing claim, 'that's it'—is a bit of metacommentary, which indicates that the first two claims are to be interpreted as belonging to this unusual context.

Austere primitivism doesn't require a fourth claim to establish its allegiance to radical intuitionism. In a matter of speaking, that allegiance is contained within the third claim, the 'that's it'. The claim 'α is known by radical intuitionism' can be taken in two senses, neither of which trespasses on austerity—either it's straightforwardly privative, so that it can be translated into 'α isn't known via any intervening entity or novel description', or it can be deflated to one or both of the existential and individuating claims. Recall that radical intuition isn't a faculty; it isn't a *means* of contact. It simply states that contact has been made. And if we have no resources for 'I know that α is goodness because β', then we're left with simply 'I know that α is goodness', which is just a pleonasm of 'α is goodness'.

4.2.2 Unstable austerity

A true austere conception stops here. It has all that it needs in its three claims—the first establishes that it's a positive ontological conception, the second establishes its subject matter, and the last forecloses further discourse. But in practice, few austere primitivists limit themselves to just these.

In spite of Moore's professed unhelpfulness, in spite of his 'that is all I have to say about it', he has a

lot to tell us about goodness. For one thing, his conception connects a bundle of deontic notions—

duty, *rightness*, and *ought*—to goodness via consequentialism. The early Moore drafts this connection

as analytic:

> [i]t is asked 'What is a man's duty under these circumstances?' or 'Is it right to act in this
> way?' or 'What ought we to aim at securing?' But all these questions are capable of further
> analysis; a correct answer to any of them involves both judgments of what is good in itself
> and causal judgments… Such a judgment can only mean that the course of action in
> question is *the* best thing to do; that, by acting so, every good that *can* be secured will have
> been secured.[268]

The later Moore abandons this view of the connection, though he maintains that the connection is

so close as to be detected by intuition: "[i]t seems to me quite self-evident that it must always be our

duty to do what will produce the best effects".[269]

And for a second thing, he tethers the normative domain to the physical domain via

supervenience. He articulates the idea of *supervenience* in the course of his investigation of the

necessity claims engendered by the idea of *intrinsicality*. He says that

> [a] kind of value is intrinsic if and only if, it is impossible that x and y should have different
> values of the kind, unless they differ in intrinsic nature; and in the equivalent statement: A
> kind of value is intrinsic if and only if, when anything possesses it, that same thing or
> anything exactly like it would necessarily or must always, under all circumstances, possess it
> in exactly the same degree.[270]

That is, any two objects, x and y, can't have different axiatic properties without having different non-

axiatic properties, and, if an object, x, is valuable, any second object, y, that has the same non-axiatic

profile as x must be valuable in the same way and to the same extent. Once generalized for all

normative entities, this is normative supervenience.

---

[268] *Ibid.*, §17.
[269] *Ethics*, page 121 (I assume that the term 'self-evident' marks an intuition).
[270] "The Conception of Intrinsic Value", page 290. Moore didn't use the term 'supervenience', but he has a strong claim to having originated its contemporary philosophical sense.

Consequentialism establishes the authority of goodness with respect to rational conation; supervenience binds it to the material world. And then there are the details of his ethics: the significance he accords to friendship and aesthetics,[271] his value-pluralism,[272] and his doctrine of organic unities.[273] The details are at times sparse, but they're enough to offer *some* nourishment to the rational animal. This presents a problem: how can these things cohere with Moore's austere primitivism? That is, how can Moore confine himself with 'good is good', together with 'that is the end of the matter' and 'that is all I have to say about it', and then proceed to instruct us so extensively on the subject of goodness? Isn't it the case that, if we have an austere primitivism, we can't have a theory?

It's difficult to envision Moore's response. My best guess is that he'd invoke his distinction between 'good' and 'the good':

> I suppose it must be granted that good is an adjective. Well, the good, that which is good, must therefore be the substantive to which the adjective good will apply: it must be the whole of that to which the adjective will apply, and the adjective must *always* truly apply to it. But if it is that to which the adjective will apply, it must be something different from that adjective itself; and the whole of that something different, whatever it is, will be our definition of *the* good.[274]

The idea, I suppose, is that goodness (or perhaps what's intended is the term 'good' itself), is an austere primitive (or it refers to an austere primitive), "the whole" of good things (or perhaps what's intended is the term 'the good' itself) is not (or it doesn't refer to one).[275] Moore might hold of each of the things I've mentioned—consequentialism, supervenience, the features of his ethics—that they

---

[271] "By far the most valuable things, which we know or can imagine, are certain states of consciousness, which may be roughly described as the pleasures of human intercourse and the enjoyment of beautiful objects" (*Principia Ethica.,* §113).
[272] "I have maintained that very many things are good and evil in themselves, and that neither class of things possesses any other property which is both common to all its members and peculiar to them" (*ibid.*, preface).
[273] *Ibid.*, §22.
[274] *Ibid.*, §9.
[275] Forgive the parentheses. I mentioned before that Moore oscillates between talk of the property goodness and the term 'good'. Things get especially confusing in §9.

belong to his account of the whole of good things (equivalently, his analysis of 'the good'), and not to goodness as such. This would leave his commitment to austerity undisturbed.

But this can't work. The principles that unify good things into a whole, i.e., that tie the several good things together in virtue of their common goodness, must emanate from goodness itself. If those things are unified in some other way, then their aggregate isn't properly called 'the good'. This is true by Moore's own stipulation: "the adjective ['good'] must *always* truly apply to [the good]"—the only way to secure this '*always*' is via direct connection to goodness. But the principles that emanate from goodness just *are* the principles of goodness. And since Moore's austere primitivism can't admit any such principles, it follows that it can't admit any unification of 'the good' either. This result is confirmed by simple common sense. It's not as though an apparently unanalyzable term 'α' can be made analyzable simply by inserting the definite article to make 'the α'. If we can understand 'the α', it's via 'α'. And, more to the point, the extension of 'α', or of any other term, is given by its intension.

In any case, because Moore's conception includes theory, its commitment to austerity is unstable. In the end, we can't provide a wholly determinate answer to the question 'is this conception austere?', not given the entirety of the view as advanced throughout the *Principia Ethica* and elsewhere. The absence of a determinate answer reveals a tension within his conception, one that a revised Moorean position should resolve. But there are two paths to a resolution, and it isn't clear which one the faithful Moorean should choose. She must either affirm the austere pronouncements of §6, and abandon all of the rest, or else weaken those pronouncements so as to rescue Moorean theory. If the latter, the faithful Moorean is a constructive primitivist; if the former, she's an austere primitivist. I've all this while called Moore austere primitivist because it strikes me that that reading does less violence to his view.

Generalizing from Moore's case, we can say that an austere conception is unstable if it exhibits a tension the resolution of which might just as easily result in a stable constructive primitivism as not. I suspect most candidate austere primitivisms are unstable in this sense, if just because it takes tremendous discipline to confine one's self to just three claims. The addition of just one claim results in instability, since this is enough to contradict the austerity claim, 'that's it'. But typically, the contradiction is owed to the presence of a bundle of claims, namely, a theory. Austerity entails theory-pessimism, i.e., if α is austere, then there can be no theory of α. Normative thought is rife with suspicion for theory, but theory-pessimism originating in austerity is rare. Moore's contemporary, H.A. Prichard, acknowledges and embraces the theory-pessimistic consequences of his own austere moral primitivism.[276] But he may be the only one.

Anyway, the more important boundary is the one between unstable austere primitivism and constructive primitivism. Because instability is a serious flaw, considerable evidence must be gathered in order to justify attributing it to a conception. That is, charity prohibits foisting it onto a conception in the absence of proper cause. I've said that I have such cause in Moore's case due to his remarks in §6. The emphasis given to these remarks strikes me as so central to Moore's thinking that it would do greater damage to his conception to ignore them or reinterpret them, so that the more charitable move is to attribute to him the mistake of having devised an unstable austere primitivism.

Of course, philosophy is by default in the business of explanation (on the widest sense of that term). The presence of a text already establishes the ambition to explain. Thus, for any given subject matter that a philosopher discusses, he must say very loudly, and repeatedly, that he doesn't

---

[276] See his "Does Moral Philosophy Rest on a Mistake?". That said, Prichard operates with a peculiar conception of *moral philosophy*, and this leaves some room to doubt whether he's really as theory-pessimistic as his austere primitivism demands. It may be that his view is better read as including a version of moral particularism (cf. Prichard's "What Is the Basis of Moral Obligation?"), so that even *his* austere primitivism is, in the end, unstable.

intend to explain any feature of it if he's to neutralize the expectation that he will. Here again, I take

Moore's 'good is good' to be loud enough to accomplish that.

But all of this shows that finding one's way to the opposite side of the boundary is

comparatively easy. There's a strong presumption of innocence—unless we find the smoking gun in

his hand, we must classify Scanlon's conception as constructive. So, in the next section, I focus on

the passage that, on its surface, most resembles Moore's §6, and show that there's nothing

incriminating in it.

4.3 SCANLON'S CONSTRUCTIVE PRIMITIVISM

I opened section 1.1 with this passage:

> I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason
> for something seems to me to lead back to the same idea: a consideration that counts in
> favor of it. "Counts in favor how?" one might ask. "By providing a reason for it" seems to
> be the only answer.[277]

This may be read as echoing Moore, with the idea of *normative reason* in place of goodness. It

announces Scanlon's non-reductionism in much the same way §6 announces Moore's non-

reductionism. The language is much less aggressive, but this is, we might suppose, only a question

of style, and not of content. All the same, I deny that Scanlon's passage compels the same

conclusion that I reached for Moore. The passage does show that Scanlon is a primitivist, but it's

compatible with either an austere or a constructive reading of that primitivism. And because

Scanlon's conception advances a number of theories, ordinary interpretative charity demands that

we categorize him as a constructive primitivist. This gives rise to an obvious question: why do these

two passages, otherwise so similar, produce these different results? This difference is particularly

striking given that I countenance unstable austerity. If this middle category is available, why not fit

---

[277] *What We Owe to Each Other*, page 17.

Scanlon under it alongside Moore? But I don't sense any instability in Scanlon. The quoted passage is best read as articulating innocent non-reductionism, rather than austerity.

Before spelling out the correct interpretation of the passage, I'll acknowledge the extent of the overlap with Moore. It's clear enough that Scanlon doesn't espouse theory-pessimism. He's the architect of a contractualist moral theory, one that's continuous with his metanormative conception in the sense that it represents moral wrongness as constructed from reasons to reject certain principles for the regulation of behaviour. His conceptions of *normative reason* and of *moral wrongness* couldn't cohere in this way if the former entailed theory-pessimism. If he's a near-brute primitivist, it must be via instability, as with Moore.

And it isn't just contractualism that threatens to destabilize a would-be Scanlonian austere primitivism. There are also the features of his conception of *normative reason* that I catalogued in subsection 1.2.1: normative reasons stand in counts-in-favor-of relations; the opposite *relatum*, that of which normative reasons count in favour, is some attitude or action of some agent; that agent figures, whether under a general or particular description, within the relation as well; etc. These features of his conception are more remarkable in this context since, in light of them, Scanlon needn't refuse the question 'what is the nature of a normative reason?' altogether. He has materials with which to furnish a response.

If we squint a bit, we could say that something like this is true of Moore too, since his consequentialism expands on the relational architecture of goodness to some degree. As it happens, Scanlon develops a buck-passing account of value, which reverses Moore's preferred relation of intra-normative reduction by defining the axiatic in terms of the deontic.[278] The two views are in this sense opposites; still, they occupy, as it were, the same spot in their respective conceptions. And

---

[278] As presented in chapter 2 of *What We Owe to Each Other*. See especially *ibid.*, page 97.

Scanlon has a rather elaborate apparatus for normative supervenience.[279] To this extent, we have symmetry between the two philosophers. But for the most part the similarities between them are confined to those places in which Moore engages in positive theorizing, so that all that they can show is that, if Scanlon's conception were austere, it would be at least as unstable as Moore's. They can't be marshalled as evidence that Scanlon's conception is austere. The two also agree about normative irreducibility, which is an ingredient of austere primitivism, but it doesn't settle anything by itself, since it's equally compatible with constructive primitivism.

I mention this in order to narrow our focus. The appropriate way to settle this issue is by looking to each philosopher's statement of his non-reductionism. It's there that the most conspicuous traces of austerity, if there are any, will appear. In the paragraph immediately after Scanlon declares that he's a primitivist at the beginning of chapter 1 of *What We Owe to Each Other*, he writes: "[t]he idea of a reason does not seem to me to be a problematic one that stands in need of explanation."[280] I would guess that, among those opening lines, this is the sentence that is most likely to be misread. It looks like the express repudiation of the thick formulation of the immanence constraint, but it isn't. As I said a moment ago, by default, philosophy explains. The first chapter of *What We Owe to Each Other* is written about the idea of *normative reason*; the presumption must be that it seeks to explain that idea. If we're to relinquish this presumption, it can only be because we're explicitly and repeatedly instructed to do so. We might read this sentence as issuing that instruction, but what of the rest of the chapter? We're better off reading the sentence such that it's designed to serve a different purpose.

How, then, should the sentence be interpreted? I think the most natural reading is clearly innocuous. Again, the sentence is 'the idea of a reason does not seem to me to be a problematic one

---

[279] As presented in lecture 2 of *Being Realistic about Reasons*.
[280] *What We Owe to Each Other*, page 17.

that stands in need of explanation'. The suggestion isn't that only problematic ideas need explanation. Rather, what the sentence says is that the idea isn't problematic and, as such, it doesn't require the special sort of explanation required for problematic ideas.[281] One way problematic ideas call for special explanation is by embodying a misunderstanding of the world. Those ideas call for debunking explanations. The idea of *magic*, for instance, misunderstands the causal laws that govern the physical domain.[282] Since we can't explain how we came to have that idea by finding some stretch of the world it captures, we must avail ourselves instead of a bit of anthropology, say, which explains the mistake. Alternatively, we can explain a problematic idea by reframing its application— perhaps our idea of *magic* should be understood in terms of the experience of awe and wonder, rather than in the (failed) cognition of causal laws. Both debunking explanations and reframing explanations are special forms of explanation. When Scanlon denies that the idea of *normative reason* stands in need of explanation, he only means to deny that it requires anything like these forms of explanation. The sentence leaves open the possibility of some less distortional form of explanation.

In fact, I think all of Scanlon's apparently austerity-espousing claims are better read as repudiations of reductive and revisionary proposals. Take the first sentence of chapter 1: "I will take the idea of a reason as primitive." Scanlon and I mean the same thing by 'primitive'; as such, all that this sentence conveys is that the idea of *normative reason* is irreducible (and perhaps also that some entities instantiate it). The next sentence—"[a]ny attempt to explain what it is to be a reason for something seems to me to lead back to the same idea"—is a restatement of that non-reductionism.

---

[281] I think we can be certain this is what he means. Consider, as evidence, this passage, which appears a few pages earlier: "[a]s I will argue in Chapter 1, I do not believe that we should regard the idea of a reason as mysterious, or as one that needs, or can be given, a philosophical explanation in terms of some other, more basic notion. In particular, the idea of a reason should not be thought to present metaphysical or epistemological difficulties that render it suspect" (*ibid.*, page 3).

[282] Cf. Scanlon on witches and spirits at page 21 of *Being Realistic about Reasons*.

It indicates that he subscribes to the view that non-reductionism is to be understood in terms of circular explanation (as opposed to, say, ontological primacy or fundamentality).

I anticipate an objection here: suppose that what Scanlon intends in this second sentence is 'seems to me to lead *immediately* back to the same idea'. This claim is *prima facie* cause for alarm, since its thrust would seem to be that talk about the nature of normative reasons is exhausted rather quickly, so that the circle Scanlon carves for the idea of *normative reason* is narrow, too narrow to allow for anything like a profitable investigation. The next couple sentences—""Counts in favor how?" one might ask. "By providing a reason for it" seems to be the only answer"—might be read as suggesting that an 'immediately' is implied, since, if 'by providing a reason for it' really is the *only* answer that can be given, it would seem to follow that inquiry into normative reasons is returned *immediately* to its starting-place.

But even here, there's room for an innocent reading. We can, for instance, understand 'only' and 'immediately' as indicating that the idea of *normative reason* must figure somewhere in our answer without supposing that it must be the whole of that answer. As we've seen, the non-reductionist about α holds that α is, in a manner of speaking, present throughout all inquiry into its nature. Even if our inquiry leads us through the development of an elaborate theory, with many steps, the idea of α must accompany us through each of these steps. This is just what it means for the idea to be irreducible. But the omnipresence of α in α-discourse doesn't commit us to the view that all we can do is repeat "that's what α is". Informative theorizing about α remains possible.

I've already mentioned that these sentences open a chapter about reasons. The relationship Scanlon draws between *provide-a-reason-for* and *counts-in-favour-of* is present all throughout that chapter, at least in the sense that it closes the explanatory circle for the idea of *normative reason* and thereby prevents any reductive analysis. But obviously he doesn't spend the chapter simply passing from the

one to the other again and again. He treats a wealth of subjects in those sixty pages. This shows his explanatory circle to be wide.

Relatedly, when, in the fifth sentence of the chapter, he says "I will presuppose the idea of a reason", we must hear this use of 'presuppose' so that it coheres with the elaboration he gives that idea throughout the rest of that chapter. It must be that what Scanlon presupposes is confined to one aspect of the idea. This way, he can set it to one side, and devote his attention to the remaining aspects. What he presupposes is just that the idea doesn't warrant elimination. That is, he's performing the ontological move I described in subsection 2.1.1, the move that's distinctive of that peculiarly ontological discursive space in which the realist is confronted with the reductionist and the nihilist. Scanlon's 'presuppose' does the work of 'it's there', or of 'backed into this corner, all that I can do is assert that the world is such that it has the idea of *normative reason* among its constituents'. As we've seen, this is a legitimate move to make in that space—it signals that a view is primitivist, but it doesn't yet settle the question of whether or not that view is austere. It's only when it's made outside of that context that we have cause to complain. But when Scanlon leaves that context a few paragraphs later, it's in order to begin his investigation of the relationship between reason and rationality, of the relationship between reason and desire, etc. In the course of those investigations, he doesn't fall back on 'it's there'. He defends novel claims with arguments.

This is the key point of divergence between Scanlon and Moore. I've made quite a lot of Moore's 'that is the end of the matter' and 'that is all I have to say about it'. Of course, they too could be read as innocent moves made in a narrow ontological context. But if we read them that way, we'd miss their connection to his intuitionism. We'd also miss a central feature of his conception. We can see this particularly clearly in the fact that, if we read Moore such that his remarks in §6 are innocent because confined to a special context, we'd need to make radical

adjustments to our interpretation of his conception. Meanwhile, if we read Scanlon the way I recommend, we needn't adjust our interpretation of his view very far, perhaps not at all.

I've focused on the beginning of chapter 1 of *What We Owe to Each Other*. The first lecture of *Being Realistic about Reasons* is also relevant, but if anything, the case for constructive primitivism is stronger there. It has its share of cumbersome bits, e.g., his use of the term 'fundamental', and his response to Korsgaard.[283] All the same, it's clear that his ambitions in that lecture (and in that book) are better framed as non-reductionist than austere. So, in the absence of any obvious evidence of austerity, we aren't entitled to foist the mistake of instability onto Scanlon. His primitivism is consistent with his theories. Therefore, it's a constructive primitivism.

4.4 CONTENT AND QUIETISM

Throughout this dissertation, I've touched on three desiderata for metanormative conceptions: constructiveness, satisfaction of the thick formulation of the immanence constraint, and joint satisfaction of the immanence and transcendence constraints. They're nested in that the last can't be met unless the second is met, and the second can't be met unless the first is met. It can, and often does, happen that a conception gets as far as the first desideratum, but no further. Because Scanlon's conception is constructive, it meets the thick formulation on its negative face. But this still leaves its positive face and the transcendence constraint.

The case might be made that the positive face is too demanding, especially in its insistence on a first-order theory of practical reasons. Still, the thick formulation requires an explanation of the possibility of practical thought, and this, we might reasonably suppose, must involve a theory of practical content if nothing else. It's conspicuous, then, that non-natural normative realists for the

---

[283] I've already addressed both of these. See subsection 2.1.1 for the former, and section 3.3 for the latter.

most part don't offer any. Nagel and David Enoch have the good grace to admit that this is the case;[284] Scanlon doesn't even broach the subject. Why not?

One answer is that there simply isn't much hope of contriving a workable non-naturalist theory of content. If a concept is non-natural, its instantiations are excluded from the causal nexus. This precludes our telling an externalist story about the acquisition of non-natural concepts.[285] Internalist stories are less attractive, but even if we opt to settle for one, it isn't clear how it would go in this case. For one thing, any theory that seeks to derive normative content from internal states, like neural states or states of consciousness, must fail by the transcendence constraint. And anyway, internalism has no way to address the aspect problem (as we saw in subsection 2.3.1). A utilitarian, say, might seek to ground the acquisition of the concept of *bad* in the experience of pain, but since experiences are private, she has no way to disambiguate the acquisition of *bad* from the acquisition of, say, *bad-for-me*.

There's a third option: refuse to grant the legitimacy of the thick formulation. I'll call this 'practical quietism'. Now, 'quietism' is a volatile term; we must be mindful of the associations it kicks up. The quietism I have in mind is distinct from the metaphilosophy associated with (and perhaps espoused by) the late Wittgenstein, what's come to be known as the therapeutic approach, the bid to "[give] philosophy peace".[286] The therapeutic approach is remarkably restrictive—as

---

[284] Nagel says he has no "theory of thought" at page 95 of *The View from Nowhere*. At page 107, he says, referring to Kripke's Wittgenstein-inspired rule-following problem, "I have no idea how to deal with it." Enoch develops the beginnings of a non-natural theory of content but emphasizes that his treatment is only preliminary (*Taking Morality Seriously*, pages 177-184).

[285] As Enoch acknowledges at *ibid.*, page 177. Non-naturalism doesn't necessarily represent normative entities as causally inert, but our concern in this section is Scanlon's non-naturalism, which does represent them in something like this way. (Or anyway, he subscribes to the view that claims about causality fall within the dominion of the scientific domain; see *Being Realistic about Reasons*, page 20).

[286] *Philosophical Investigations*, §133.

McDowell indicates, it "does not include putting forward distinctively philosophical theses."[287] This is how it interprets the Wittgensteinian slogan that philosophy "leaves everything as it is."[288]

Practical quietism likely does have a Wittgensteinian pedigree. Still, it isn't as severe as the therapeutic approach, not in its purest form. For one thing, it doesn't condemn positive philosophy across the board. It's an instance of the kind of softened quietism articulated by McDowell, the kind that permits, e.g., philosophical "reflections about the requirements of justice or the proper shape of a political community."[289] In this less categorical form, quietism only takes issue with philosophy that engages in foundational projects, i.e., philosophy that asks "something of the form, "How is such and such possible?""[290]

But this is precisely what the thick formulation does—it asks how practical thought is possible. Thus, the practical quietist must dismiss that formulation as the product of a confusion. She likely takes the reasoning with which I defended the immanence constraint to be confused as well. I would guess that the charge could be formulated as follows: my case revolves around an idea of *the possibility of practical thought*; in order to think this idea, I'd have to step outside of thought itself, and assess its nature; there's no stepping outside of thought itself; therefore, the idea is unintelligible and any case that revolves around it is confused. (That said, the true quietist might not permit herself this line of argument, since it helps itself to something like idea of *boundaries of thought*, which is also, by her lights, unintelligible.)

Does anything suggest that Scanlon is a practical quietist? He doesn't formally declare his adherence to any Wittgensteinian programme,[291] but he does indicate that he's unimpressed by many of the metaphysical questions that are pursued in contemporary metanormative thought. He says:

---

[287] "Wittgensteinian "Quietism"", page 365.
[288] *Philosophical Investigations*, §124.
[289] "Wittgensteinian "Quietism"", page 367.
[290] *Ibid.*, page 371.
[291] As far as I can tell, he doesn't mention Wittgenstein at all.

> [w]hat drives me to look for a characterization of the subject matter of judgments of right and wrong… is not a concern about the metaphysical reality of moral facts. If we could characterize the method of reasoning through which we arrive at judgments of right and wrong, and could explain why there is good reason to give judgments arrived at in this way the kind of importance that moral judgments are normally thought to have, then we would, I believe, have given a sufficient answer to the question of the subject matter of right and wrong as well. No interesting question would remain about the ontology of morals—for example, about the metaphysical status of moral facts.[292]

One question he means to rule out is whether the truth of a normative claim involves a robust ontological commitment, that is, whether analysis of that claim reveals it to quantify over ontologically hefty entities. He's especially keen to deny that we must determine what stretch of the world these entities occupy.[293] He doesn't go so far as to call these questions confused, but he does deny that a reasonably comprehensive metanormative conception must address them.

We might think that all of this should be attributed to a comparatively innocent Rawlsianism, rather than to practical quietism. After all, it recalls the strategy of collapsing the boundary between normative and metanormative thought that I described in connection with Rawls in subsection 3.4.2. But I don't think Rawls' strategy can be taken as far as Scanlon needs it to go. As I interpret it, the strategy only collects normative and metanormative views together so that they're assessed as a whole, rather than independently. This makes normative considerations relevant to our choice among metanormative views, but it doesn't reduce any metanormative questions to normative questions. Granted, Rawls says that "if we can find an accurate account of our moral conceptions, then questions of meaning and justification may prove much easier to answer. Indeed some of them may no longer be real questions at all."[294] But this last sentence falls well short of quietism. What Rawls is prescribing is an open-textured, exploratory approach, such that, for instance, if we encounter an apparently intractable problem in the ordinary language analysis of moral claims, we

---

[292] *What We Owe to Each Other*, page 2.
[293] *Ibid.* See also *Being Realistic about Reasons*, page 24.
[294] *A Theory of Justice.*, page 45.

ought to turn to first-order moral theory, the development of which might reveal a path to the solution to our original problem. There's nothing here to suggest that we're entitled to ignore that problem (or to condemn it as confused) in advance. More importantly, it doesn't *assume* that first-order theorizing will solve or dissipate each of our second-order problems. It only invites us to see how far such theorizing can get us.

That said, it does happen that apparently second-order work is immediately completed by first-order theory. Issues in normative semantics can be settled in this way. Taken in this spirit, I agree with this passage:

> [i]n order to show that questions of right and wrong have correct answers, it is enough to show that we have good grounds for taking certain conclusions that actions are right or are wrong to be correct, understood as conclusions about morality, and that we therefore have good grounds for giving these conclusions the particular importance that we normally attach to moral judgments.[295]

If we have an acceptable moral theory—e.g., one that specifies our reasons to do what's right and avoid what's wrong—then we have no further need of a theory of moral truth. The semantics of moral claims can be read off of the moral theory.[296] And, if we can specify patterns of moral reasoning that transmit truth values, we rule out (or if not that, embarrass) moral non-cognitivism.[297] But all of this arises downstream from the question of how we fix moral content. We can't interpret our moral theory until we know what 'moral' means. And for this, we need foundational inquiry into how moral content is fixed.

There are issues in normative epistemology that produce a similar result. The epistemological conception Scanlon favours is Rawlsian reflective equilibrium (as described in

---

[295] *What We Owe to Each Other*, pages 2-3.
[296] This is trivially true if we represent moral theory as the articulation of principles for the distribution of moral predicates like 'is right' and 'is wrong'. But this would be to understand theory in terms of its semantic consequences. It's more illuminating to represent moral theory in the way Scanlon does, i.e., in terms of reasons and reasoning.
[297] Of course, quasi-realism survives. Perhaps the only views that are altogether defeated in this way are the positions in ordinary language philosophy that Rawls had to dislodge in order to rescue moral and political philosophy.

subsection 2.2.2).[298] Because reflective equilibrium doubles as a method for theory-construction,

epistemological conceptions produced in this way amount to first-order theories. This suggests a

procedure for addressing skeptical worries raised about normative matters: translate them into first-

order concerns, and then engage them on those terms. So, for instance, skepticism about the duty

to honour a promise can be assuaged by appealing to, say, the fact that we have a reason to want to

be governed by a principle that prohibits disappointing legitimately acquired expectations.[299]

Meanwhile, the kind of thoroughgoing moral skepticism that takes the shape of a 'why be moral?'

question can be defeated by appealing to our reasons to live alongside others on terms they can't

reasonably reject.[300] We might also use this procedure to engage skepticisms that are more

conventionally located at the second-order. Scanlon says that "my judging that something is a

reason does not make it so. Whether my judgment has this effect or not is a first-order normative

question, the answer to which seems obviously to be that it does not."[301] That is, crude

subjectivism, according to which I have a reason if I judge that I do, is straightforwardly defeated by

the recognition that I might judge that I have a reason to throw myself down the stairs in

circumstances where I obviously have no reason to do so.

This procedure of assimilating skeptical conceptions into the first-order is remarkably

powerful. But not every skeptical challenge can be met in this way. Some impugn first-order theory

itself. Remarkably, Scanlon characterizes this wider skepticism as unmotivated:

> there are central cases in which judgments about reasons seem clearly true. If we should
> reject these judgments, this has to be on the basis of substantive grounds for thinking them
> mistaken; not on the basis of questions about how we could be in touch with such facts at
> all. General doubts of the latter kind would be relevant only if normative conclusions could
> have the significance they claim only if the facts they purport to represent had some special

---

[298] Scanlon describes a decision procedure at pages 65-66 of *What We Owe to Each Other*. He confirms that this is reflective equilibrium and that it's his chosen epistemological conception in lecture 4 of *Being Realistic about Reasons*.
[299] This is Scanlon's principle of fidelity (see *What We Owe to Each Other*, page 304).
[300] See *ibid.*, pages 153-155.
[301] *Being Realistic about Reasons*, page 93.

metaphysical character that would make them inaccessible to us. I see no … reasons to believe this in the case of conclusions about practical reasons…[302]

I don't see what sustains the criterion according to which deep skepticism is only viable if normativity exhibits a special metaphysical character. But even if we grant it, nothing prevents the skeptic from insisting that normative facts have this character (it isn't inconceivable that normative transcendence might be taken to have exactly this consequence).[303]

Reflective equilibrium can't defeat this variety of skepticism on its own. We can test principles against considered judgments all we like; unless we have some assurance that our considered judgments are largely on the right track, we might only be entrenching our mistakes rather than coming any nearer to the truth. Scanlon argues that reflective equilibrium already supplies that assurance, since a judgment doesn't qualify as considered unless it "seems to me to be clearly true *when I am thinking about the matter under good conditions for arriving at judgments of the kind in question.*"[304] On his view, this suffices for a presumption of truth. I don't share his confidence. Even if we bracket the fact that good conditions for judgment are no guarantee against error, we're still vulnerable to the skeptic who asks how we know which conditions are the right ones.

So, *contra* Scanlon, both normative semantics and normative epistemology exhibit a two-tiered structure: local, or surface-level, concerns can be subsumed under first-order theory, but there remain deeper concerns that demand a second-order response. As it happens, in both cases these deeper concerns can be addressed in a single move. To see this, we need only observe that normative skepticism relies on a theory of practical content of its own—the skeptical proposal that,

---

[302] *Ibid.*, page 86. In the note at that page, he says: "[h]ere I am in agreement with Ronald Dworkin that only internal skepticism is worth worrying about."

[303] Perhaps the idea is that the skeptic mustn't simply advance her view in a vacuum—she must identify substantive considerations that back it up. But the skeptic does identify a substantive consideration, namely, the fact that the appearances are preserved by rival explanations (e.g., the view that practical thought is only the product of evolutionary pressures and confabulation). If Scanlon is only willing to accept first-order considerations, he's begging the question. (Scanlon is unlike Moore in that his primitivism is non-austere. Still, he does from time to time call to mind Moore's infamous 'here is one hand' argument.)

[304] *Ibid.*, page 82. See also *What We Owe to Each Other*, page 71.

for all we know, we never apply our concepts correctly is only intelligible if we have concepts to misapply. Thus, deep normative skepticism is defeated if the acquisition of normative concepts is such that it necessarily includes instances of successful application.

Davidson's triangulation argument demonstrates that this is the case. Fixing the contents of our normative terms requires both discourse with a fellow agent and contact with a worldly normative entity. That entity is what supplies the apex of the triangle, which must be causally efficacious if it's to contribute its share to the content-fixing process, so that in the moment that, say, the goodness of the cat's comfort tethers the term 'goodness' to goodness, it causes the belief that her comfort is good. And, since the object of the belief causes that belief specifically in virtue of its bearing the aspect under which it's believed, this secures the knowledge that the cat's comfort is good. It follows that we can't simultaneously hold that we have normative concepts and that our applications of those concepts may radically and systemically misrepresent normative reality.[305]

Note, though, that the Davidsonian can't dispense with reflective equilibrium. The triangulation argument only demonstrates that our beliefs, normative and otherwise, are largely correct, in the sense that no one's whole set of beliefs can be radically or systemically in error. But some false beliefs remain, and because we can't discern which beliefs are which, we need a method that leverages the absence of systemic error in order to eliminate more prosaic error. Several arguments could be marshalled to demonstrate that reflective equilibrium is the method the Davidsonian needs, but surveying them here would take us too far afield. The basic idea is that the

---

[305] Davidson presents several variations on this argument, though his target is always external world skepticism, rather than deep normative skepticism. As before, I rely on Verheggen's formulation (see chapter 4 of *Donald Davidson's Triangulation Argument*), and Myers' extension of Davidson's arguments to metanormative thought.

 The skeptic might object that the triangulation argument only shows that we have knowledge in basic cases and that this isn't enough to prevent systemic error. But basic cases only arise within a system of their own, the holistic system of attitudes that enables discourse between triangulators. It follows that it's impossible to get just one thing right. If I have one thing right, then I have many things right.

holism of the mental reveals the normative domain to be characterized by a coherence best mapped through the systematization of judgments in principles.[306]

This leaves us with a kind of division of labour, or relay race, in which Davidson runs the first lap, the one establishing the presumptive veridicality of judgments, and Rawls runs the second lap, the one that establishes that specific judgments are true.[307] The second lap belongs to the first-order, but the first lap remains at the second-order. This is as it should be; normative epistemology is too complex to be tackled at just one of the two levels.

Strictly speaking, we don't need the triangulation argument to deploy this line of reasoning against the skeptic. Any theory of content will do, as long as it has the consequence that our first, meaning-fixing steps must have been right. But Scanlon has no theory of content, so he can't avail himself of this strategy. This is why he must fall back on practical quietism, and maintain that any view that can't be translated into a first-order consideration isn't worth engaging.

Due to this quietism, Scanlon's non-natural realism fails by the thick formulation of the immanence complaint and has no answer to the deep (though still perfectly intelligible) normative skeptic. It also makes something of a mystery of the relationship between normative judgments and conations.[308] But it's vulnerable to a fourth, and more remarkable, line of criticism: it isn't clear that it meets the transcendence constraint. It's here, if anywhere, that non-natural realism ought to excel, since it explicitly disavows the many reductionisms and scientistic naturalisms that would otherwise threaten normative prescriptivity. But increasingly, its adherents represent non-naturalness as entailing a kind of ontological insubstantiality. As Scanlon explains, his ontology

---

[306] This recommends the move from considered judgments to narrow reflective equilibrium. The move from narrow to wide reflective equilibrium is recommended on the grounds that, given triangulation, few of the many different moral traditions we've accumulated throughout history can be radically wrong. Thus, we should expect to find something worth preserving in all but the most disastrous conceptions.

[307] This is Myers' proposal. See note 115.

[308] Scanlon's idea of *judgment-sensitivity* (see note 34) may be taken to establish this relationship, but since he offers no account of judgment-sensitivity itself, this only pushes the problem back one step (see Myers' "Davidson's Meta-normative Naturalism and the Rationality Requirement", pages 31, 33-34).

takes as basic a range of domains, including mathematics, science, and moral and practical reasoning. It holds that statements within all of these domains are capable of truth and falsity, and that the truth values of statements about one domain, insofar as they do not conflict with statements of some other domain, are properly settled by the standards of the domain that they are about… [I]t is tempting to think of a domain as consisting of a realm of objects of a certain kind and their properties. But this would be misleading. The normative domain, for example, is not a distinct realm of objects… [A] domain is better understood in terms of the kind of claims it involves, and hence in terms of concepts that it deals with[.]" [309]

Later, he says the normativity is objective in at least these two ways: it's judgment-independent, i.e., we can misrepresent it,[310] and it's choice-independent, i.e., "the standards for assessing such judgments do not depend on what we, collectively, have done, chosen, or adopted, and would not be different had we done, chosen, or adopted something else."[311]

Neither judgment-independence nor choice-independence suffices to meet the transcendence constraint. What we need is the possibility of a perspective from which to assess the whole of our agential nature, including unchosen aspects of that nature. That said, Scanlon's conception might, for all he's written, meet stricter standards for normative objectivity than the two he's mentioned. The question, then, is whether his ontological conception gives us reason to believe it puts the requisite perspective out of reach. I think it does. Consider this passage:

> We could adopt some way of talking which specified criteria of identity for objects of a certain sort, and truth conditions for sentences containing terms referring to them which allowed for existential generalization from such sentences. According to my view, as long as this way of talking was well defined, internally coherent, and *did not have any presuppositions or implications that might conflict with those of other domains, such as science*, by accepting these statements we would be committed to the existence of things quantified over in the existential statements counted as true in this way of talking.[312]

---

[309] *Being Realistic about Reasons*, page 19.

[310] *Ibid.*, page 93. Scanlon seems reluctant to countenance the possibility of systemic error. This isn't surprising given that he's said that the normative domain is defined in terms of a distinctive kind of claim, that is, in terms of discourse. No pattern of discourse could survive systemic error. But this gets things backward—we're not entitled to tailor our criteria for objectivity to suit our ontological conceptions. Rather, our ontological conceptions ought to be designed to meet independent criteria for objectivity.

[311] *Ibid.*, page 94.

[312] *Ibid.*, page 27. Scanlon says this in the course of articulating an objection to this view. Still, he acknowledges that his domain-ontology has this consequence.

In effect, on his view normative ontology amounts to a pattern of discourse. This isn't to say that entities posited by true claims made within that discourse are fictions—those entities genuinely exist. But he takes the nature of existence itself to be answerable to the distinctive features of each domain: "[w]e make claims expressed by the existential quantifier in many domains, but what is required to justify any existential claim, and what follows from such a claim, varies, depending on the kind of thing that is claimed to exist."[313] This leaves us with no space from which to assess the discourse itself. And, since discourse belongs to agency, Scanlon's normative ontology can't meet the transcendence constraint.

This complaint should impress Scanlon even given his practical quietism. He requires of complaints raised against his ontology that they take this shape: "the charge would be that in order for normative truths to have the significance normally attributed to them, they would have to be true (or justified) in a sense that goes beyond what reasoning internal to the normative domain (i.e. thinking about what reasons we have) could by itself establish."[314] The complaint from the transcendence constraint takes precisely that shape—the metaphysical question it raises emanates from normativity itself, so that it's "domain-specific".[315]

A last thought: any formulation of normative objectivity must involve normative discourse, if just because formulations only arise in discourse. Doesn't this show the transcendence constraint to be impossible to satisfy? And, for that matter, how does the Davidsonian conception I favour do any better than Scanlon's? Objectivity can't be a matter of locating discourse-independent truth-makers, since there's no truth prior to discourse, and no truth-makers prior to truth. For this same reason, we can't determine the apexes of triangles before the triangulators have fixed them.

---

[313] *Ibid.*, page 25. I mentioned at the beginning of this chapter that Scanlon is friendly with quasi-realism. This comes through especially clearly here.
[314] *Ibid.*, pages 28-29.
[315] *Ibid.*, page 25.

But discourse-transcendent objectivity doesn't call for discourse-independent truth-makers or apexes. The triangulators triangulate on a thing. This thing was there prior to triangulation. It had the feature under which the triangulators describe it prior to their describing it in that way. Thus, the thing and its feature transcend discourse.

The difficulty isn't that there are no normative entities to be found prior to normative discourse. The difficulty is that they abound—there's a cacophony of such entities. It isn't a matter of finding our way *to* them, but finding our way *among* them, among *all* of them, the great buzzing throng of entities. We couldn't possibly manage that until we pinned them down in discourse. Granted, they can't tell us what to do until we can think them. As a consequence, pre-discursive normativity is wholly axiatic, and not deontic. But that's enough.[316]

In subsection 1.2.3, I mentioned Moore's isolability test, and suggested this variation: each intrinsic value could obtain even if it were the case that no agency ever did. One might argue that the beauty of a bucolic landscape can't in the end be isolated from us because its beauty is only recognizable when observed from a certain position on the surface of the planet, and it would never occur to us to occupy that position unless we were already the sort of creatures who routinely find ourselves in positions like it. But this argument is too clever for its own good. The beauty is there from the start and whether or not it would occur to us to look for it is neither here nor there. So, aesthetic value is isolatable from us after all; it doesn't owe anything to our practices, habits, ways of life, or whatever it may be.

---

[316] It's enough in that it makes the question 'are our discursive practices good?' intelligible. Asking this question is what enables us to occupy the perspective from which to assess the whole of our agential nature. The question and the perspective must arise within discourse. Still, they have the potential to disrupt our practices. At the very least, they undermine the complacency that would otherwise predominate if we supposed that the way we talk about normativity determines the entirety of the normative domain.
  Scanlon might make the case that it's better to defeat complacency at the first-order than with a bit of metaphysics. For the most part, I'm inclined to agree. But there's an exception: talk of practices and the like encourages its own special variety of parochialism and conservatism. And since such talk is given a metaphysical valence, it must be met with a metaphysical response.

Anscombe, Elizabeth. 1957. *Intention*. Cambridge: Harvard University Press.

———. 1958. "Modern Moral Philosophy." *Philosophy* 33: 1–19.

Aristotle. 2009. *The Nicomachean Ethics*. Translated by David Ross. Oxford: Oxford University Press.

Bentham, Jeremy. 1879. *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press.

Broome, John. 2007. "Does Rationality Consist in Responding Correctly to Reasons?" *Journal of Moral Philosophy* 4 (3): 349–74.

Clarke, Samuel. 1990. "A Discourse of Natural Religion." In *Moral Philosophy from Montaigne to Kant*, edited by J. B. Schneewind. Cambridge: Cambridge University Press.

Collins, Arthur W. 1987. *The Nature of Mental Things*. Notre Dame: University of Notre Dame Press.

Cuneo, Terence. 2007. *The Normative Web*. Oxford: Oxford University Press.

Dancy, Jonathan. 1993. *Moral Reasons*. Oxford: Blackwell.

Davidson, Donald. 1980a. "Actions, Reasons, and Causes." In *Essays on Actions and Events*. New York: Oxford University Press.

———. 1980b. "Freedom to Act." In *Essays on Actions and Events*. New York: Oxford University Press.

———. 1980c. "How Is Weakness of the Will Possible?" In *Essays on Actions and Events*. New York: Oxford University Press.

———. 1980d. "Mental Events." In *Essays on Actions and Events*. New York: Oxford University Press.

———. 2001a. "Comments on Karlovy Vary Papers." In *Interpreting Davidson*, edited by Petr Kotatko, Peter Pagin, and Gabriel Segal. Stanford: CSLI.

———. 2001b. "Externalisms." In *Interpreting Davidson*, edited by Petr Kotatko, Peter Pagin, and Gabriel Segal. Stanford: CSLI.

———. 2001c. "Rational Animals." In *Subjective, Intersubjective, Objective*. New York: Oxford University Press.

———. 2001d. "The Emergence of Thought." In *Subjective, Intersubjective, Objective*. New York: Oxford University Press.

———. 2001e. "The Second Person." In *Subjective, Intersubjective, Objective*. New York: Oxford University Press.

———. 2004. "Paradoxes of Irrationality." In *Problems of Rationality*. Oxford: Clarendon Press.

———. 2005. "A Nice Derangement of Epitaphs." In *Truth, Language, and History*. Oxford: Clarendon Press.

Enoch, David. 2011. *Taking Morality Seriously*. Oxford: Oxford University Press.

Fernandez, Patricio A. 2015. "A Realistic Practical Conclusion." *American Philosophical Quarterly* 52 (2): 115–28.

Frankena, William. 1963. *Ethics*. Englewood Cliffs: Prentice Hall.

Harman, Gilbert. 1977. *The Nature of Morality*. New York: Oxford University Press.

Hobbes, Thomas. 1996. *Leviathan*. New York: Oxford University Press.

Hume, David. 1975. *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge. Oxford: Clarendon Press.

Kant, Immanuel. 1970. "Perpetual Peace." In *Political Writings*, edited by H.S. Reiss and H. B. Nisbet. Cambridge: Cambridge University Press.

———. 1996. *The Metaphysics of Morals*. Cambridge: Cambridge University Press.

———. 1997a. *Critique of Practical Reason*. Translated by Mary Gregor. Cambridge: Cambridge University Press.

———. 1997b. *Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor. Cambridge: Cambridge University Press.

———. 1998. *Critique of Pure Reason*. Translated by Paul Guyer and Allen W. Wood. New York: Cambridge University Press.

Keller, Simon. 2007. "Virtue Ethics Is Self-Effacing." *Australasian Journal of Philosophy* 85 (2): 221–31.

Korsgaard, Christine M. 1996a. "Kant's Analysis of Obligation." In *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

———. 1996b. "Skepticism about Practical Reason." In *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

———. 1996c. *The Sources of Normativity*. Cambridge: Cambridge University Press.

———. 2008a. "Realism and Constructivism in Twentieth-Century Moral Philosophy." In *The Constitution of Agency*. Oxford: Oxford University Press.

———. 2008b. "The Normativity of Instrumental Reason." In *The Constitution of Agency*. Oxford: Oxford University Press.

———. 2009. *Self-Constitution*. Oxford: Oxford University Press.

———. 2018. *Fellow Creatures*. Oxford: Oxford University Press.

Lewis, David. 1984. "Putnam's Paradox." *Australasian Journal of Philosophy* 62 (3): 221–36.

Mackie, J. L. 1976. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin Books.

McDowell, John. 1998a. "Functionalism and Anomalous Monism." In *Mind, Value, & Reality*. Cambridge: Harvard University Press.

———. 1998b. "Might There Be External Reasons?" In *Mind, Value, and Reality*. Cambridge, Mass.: Harvard University Press.

———. 2009. "Wittgensteinian 'Quietism.'" *Common Knowledge* 15 (3): 365–72.

Mill, John Stuart. 2001. *Utilitarianism*. Indianapolis: Hackett Publishing Company.

Moore, G. E. 1966. *Ethics*. London: Oxford University Press.

———. 1993a. *Principia Ethica*. Revised edition. Cambridge: Cambridge University Press.

———. 1993b. "The Conception of Intrinsic Value." In *Principia Ethica*, revised edition. Cambridge: Cambridge University Press.

Myers, Robert H. 2020. "Replies to Kirk Ludwig and Paul Hurley." *Dialogue: Canadian Philosophical Review* 59 (2): 255-269.

———. 2021. "Davidson's Meta-Normative Naturalism and the Rationality Requirement." In *Donald Davidson on Action, Mind and Value*. Singapore: Springer.

Myers, Robert, and Claudine Verheggen. 2016. *Donald Davidson's Triangulation Argument*. New York: Routledge.

———. forthcoming. "Realism Rehabilitated." In *Neopragmatism*, edited by Joshua Gert. Oxford: Oxford University Press.

Nagel, Thomas. 1970. *The Possibility of Altruism*. Princeton: Princeton University Press.

———. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

———. 1997. *The Last Word*. Oxford: Oxford University Press.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

———. 2011a. *On What Matters, Vol. I*. Oxford: Oxford University Press.

———. 2011b. *On What Matters, Vol. II*. Oxford: Oxford University Press.

Pettit, Philip, and Michael Smith. 2008. "External Reasons." *Philosophy and Phenomenological Research* 55 (1): 142–79.

Plato. 1997. "Republic." In *Plato: Complete Works*, edited by John M. Cooper, translated by G. M. A. Grube and C. D. C. Reeve. Indianapolis: Hackett Publishing Company.

Prichard, H. A. 2002a. "Does Moral Philosophy Rest on a Mistake?" In *Moral Writings*. Oxford: Clarendon Press.

———. 2002c. "What Is the Basis of Moral Obligation?" In *Moral Writings*. Oxford: Clarendon Press.

Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy & Public Affairs* 13 (2): 134–171.

———. 1986. "Moral Realism." *The Philosophical Review* 95 (2): 163–207.

———. 1989. "Naturalism and Prescriptivity." *Social Philosophy and Policy* 7: 151–74.

Rawls, John. 1993. *Political Liberalism*. New York: Columbia University Press.

———. 1999a. *A Theory of Justice*. Revised edition. Cambridge, Mass.: Harvard University Press.

———. 1999b. "Kantian Constructivism in Moral Theory." In *Collected Papers*. Cambridge: Cambridge University Press.

———. 1999c. "Outline of a Decision Procedure for Ethics." In *Collected Papers*. Cambridge: Cambridge University Press.

Raz, Joseph. 1975. *Practical Reason and Norms*. Oxford: Oxford University Press.

Russell, Bertrand. 1972. *The Philosophy of Logical Atomism*. London and New York: Routledge.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.

———. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.

Sidgwick, Henry. 1874. *The Methods of Ethics*. London: MacMillan and Co.

Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.

Stocker, Michael. 1976. "The Schizophrenia of Modern Ethical Theories." *The Journal of Philosophy* 73 (14): 453–66.

Strawson, P. F. 1994. *Analysis and Metaphysics: An Introduction to Philosophy*. Oxford: Oxford University Press.

Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127 (1): 109–66.

———. 2008. "Constructivism about Reasons." *Oxford Studies in Metaethics* 3: 207–45.

———. 2009. "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters." *Philosophical Issues* 19 (1): 273–98.

———. 2010. "What Is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5 (5): 363–84.

Thompson, Michael. 2008. *Life and Action*. Cambridge: Harvard University Press.

Velleman, David. 1992. "The Guise of the Good." *Nous* 26 (1): 3–26.

Williams, Bernard. 1981. "Internal and External Reasons." In *Moral Luck*. Cambridge: Cambridge University Press.

———. 1985. *Ethics and the Limits of Philosophy*. London: Fontana.

———. 1995. "Internal Reasons and the Obscurity of Blame." In *Making Sense of Humanity*. Cambridge: Cambridge University Press.

Williams, Bernard, and J. J. C. Smart. 1973. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.

Wittgenstein, Ludwig. 1965. "A Lecture on Ethics." *The Philosophical Review* 74 (1): 3–12.

———. 1974. *Tractatus Logico-Philosophicus*. Translated by D. F. Pears and B. F. McGuinness. London and New York: Routledge.

———. 2009. *Philosophical Investigations*. Translated by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte. Oxford: Wiley-Blackwell.

Wright, Crispin. 1996. "Truth in Ethics." In *Truth in Ethics*, edited by B. Hooker. Oxford: Blackwell.