Philosophy Now

Series Editor: John Shand

This is a fresh and vital series of new introductions to today's most read, discussed and important philosophers. Combining rigorous analysis with authoritative exposition, each book gives clear and comprehensive access to the ideas of those philosophers who have made a truly fundamental and original contribution to the subject. Together the volumes comprise a remarkable gallery of the thinkers who have been at the forefront of philosophical ideas.

Published

Thomas Kuhn
Alexander Bird

John Searle
Nick Fotion

Charles Taylor
Ruth Abbey

Robert Nozick
A. R. Lacey

Robert Nozick

A. R. Lacey

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

VOLUME 18 NUMBER 2 SUMMER 2001

Moral Epistemology

CONTENTS

CAMBRIDGE
UNIVERSITY PRESS

TWO SOURCES OF MORALITY*

By Philip Pettit

I. Introduction

This essay emerges from consideration of a question in the epistemology of ethics or morality. This is not the common claim-centered question as to how moral claims are confirmed and whether their mode of confirmation gives us grounds to be confident about the prospects for ethical discourse. Instead, I am concerned with the less frequently posed concept-centered question of where in human experience moral terms or concepts are grounded—that is, where in experience the moral becomes salient to us. This question was central to moral epistemology in the form it took among thinkers such as Locke, Hume, and Kant, and it remains of the first importance today.[1]

The question calls for a naturalistic genealogy of moral terms and concepts. I assume that we are not possessed of an irreducibly moral sense whereby irreducibly moral properties might be revealed to us. I also assume that we human beings have no nonnaturalistic faculties of perception and cognition, and that in any case there is nothing nonnaturalistic that such faculties might register. The question, then, is what it is about naturalistic experience—what it is about experience of the kind that raises no particular worries for a scientific view of the world—that can occasion moral conceptualization and create an opportunity for the useful deployment of moral terms.

The analysis that one offers of moral terms or concepts will sometimes point to a genealogy in this sense. If one offers a noncognitivist

* In writing this essay, I benefited from exchanges with Allan Gibbard, Oswald Hanfling, Brad Hooker, Victoria McGeer, Susan Mendus, Michael Ridge, Michael Smith, and R. Jay Wallace, and from comments from this volume's contributors and editors. I was particularly influenced by conversations with Stephen Darwall and Geoffrey Sayre-McCord, and I owe a very considerable debt to each. The essay was also improved by helpful discussions when it was presented at the annual meeting of the British Society for Ethical Theory in July 2000, and at a university seminar at the University of Nebraska, Lincoln, in September 2000.

[1] For a recent influential discussion of the concept-centered problem, see Christine M. Korsgaard, *The Sources of Normativity* (New York: Cambridge University Press, 1996); and Christine M. Korsgaard, "Self-Constitution in the Ethics of Plato and Kant," *Journal of Ethics* 3, no. 1 (1999): 1–29. My approach is rather different from hers, as I trace ethical conceptualization to a more social, and less reflective, origin; in this respect it is closer to the approach found in Gerald J. Postema, "Morality in the First Person Plural," *Law and Philosophy* 14, no. 1 (1995): 35–64. For another different approach, one that involves many congenial themes despite supporting noncognitivism, see Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990).

analysis in which moral evaluation is conceived of as a species of attitudinal expression or projection, for example, then that will point to a distinctive story as to what it is about experience that occasions such evaluation.[2] Often, however, the analysis of moral terms will leave the genealogical question undischarged. Suppose that one postulates a network analysis of moral terms under which the meanings of these different terms are established by the linkages that we recognize both among the terms themselves and between these terms and terms of a nonmoral kind; I favor this sort of approach myself.[3] Such a network analysis presupposes that there are some points at which the networked concepts make contact with experience; some moral concepts—some of the terms implicated in the web of moral discourse—must get pinned down for us in experience, however tentatively, by more direct semantic means. But while the analysis will recognize the need for such points of contact, it may say little on how they are established; it may leave the genealogical question effectively unanswered.

While I favor a cognitivist network analysis of moral terms, I try to pursue the genealogical question here in abstraction from issues of analysis. I do often speak in this essay in a way that presupposes that moral concepts serve to ascribe properties, so that, as cognitivists say, the evaluations in which they figure are true or false. I presuppose such cognitivism, however, only at points where those who think that evaluations are merely expressions of noncognitive attitude will be able to recast the argument in a manner that is congenial to their views. For example, in Section VI, I will introduce an account of how sentiments give rise to moral conceptualization. While this account is cast in terms favorable to cognitivism, it will not be difficult to see how a corresponding noncognitivist story would go.

The remainder of this essay is organized into six sections. In Section II, I argue that *intentional subjects*—subjects who form and act on beliefs and desires—need not have any normative concepts whatsoever. In Section III, I argue that creatures who are *discursive* as well as intentional—that is, creatures who express their intentional states through the use of common, voluntary signs—do have to possess certain normative concepts: namely, those associated with reasoning. In Section IV, I point out that the normative concepts required for discourse fall short of being

[2] For a sketch of the cognitivist position I defend, see Frank Jackson and Philip Pettit, "Moral Functionalism and Moral Motivation," *Philosophical Quarterly* 45, no. 178 (1995): 20–40; and Philip Pettit, "Embracing Objectivity in Ethics," in Brian Leiter, ed., *Objectivity in Law and Morals* (Cambridge: Cambridge University Press, 2001). For a critique of expressivism and an indirect argument for cognitivism, see Frank Jackson and Philip Pettit, "A Question for Expressivism," *Analysis* 58, no. 4 (1998): 239–51.

[3] For examples of this sort of approach, see S. L. Hurley, *Natural Reasons* (New York: Oxford University Press, 1989); Jackson and Pettit, "Moral Functionalism and Moral Motivation"; and Ralph Wedgwood, "Conceptual Role Semantics for Moral Terms," *Philosophical Review*, forthcoming.

moral concepts proper. And then Sections V and VI focus on two ways in which moral considerations can become salient for discursive creatures. The first involves people privileging discourse as a form of interaction, and the second involves them extending such discourse to the realm of sentiment. Privileging discourse and using discourse to express sentiment are the two distinct sources of moral conceptualization that are signaled in the title of this essay, and I argue in a brief concluding section—Section VII—that the different concepts they provide represent rival possibilities in the construction of moral theory.

II. Intentional Subjects May Lack Normative Concepts

Human beings are intentional creatures who form and act on beliefs and desires in a more or less rational way. Is the experience of intentional creatures bound to make moral considerations salient, and bound therefore to provide an occasion for the formation of corresponding concepts? No, it is not. I argue in this section that, on the contrary, being an intentional subject is consistent with completely lacking normative concepts.

Any intentional subject has to be designed so that it performs well in representational and related respects, and being an intentional subject amounts to nothing more and nothing less than satisfying such design specifications.[4] This is, at any rate, what I shall assume. The intentional subject will represent things as they appear within the limits of its perceptual and cognitive organization, updating these representations appropriately in the light of new inputs. And it will act in ways that further its desires—presumptively, desires that reflect its overall needs and purposes—in the light of these representations or beliefs. The intentional subject may not perform these tasks perfectly, of course, but it will have to do them well—and perhaps do them well as a result of a certain history or organization—within what are thought to be feasible limits and favorable circumstances.

The rational achievements that the intentional subject has to display, according to this account, are quite impressive. Nevertheless, it turns out that even a relatively simple mechanical or organic system can be intentional in this sense. Without itself having any sort of control over the process, an animal or robot might be constituted so as to adjust to incoming information in a relatively faithful manner, forming beliefs that determine, in the presence of certain standing or situationally variable desires, what the creature does. A suitably preset design could function so as to update the creature's beliefs in response to changes in perceptual input,

[4] On design specifications, see Daniel C. Dennett, *The Intentional Stance* (Cambridge, MA: MIT Press, 1987); Philip Pettit, *The Common Mind: An Essay on Psychology, Society, and Politics* (New York: Oxford University Press, 1993), chap. 1; and Peter Railton, "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action," in Garrett Cullity and Berys Gaut, eds., *Ethics and Practical Reason* (Oxford: Oxford University Press, 1997).

and these beliefs could then interact with the creature's desires in a similarly preset way so as to determine the creature's actions. Such a creature might not fulfill the criteria for being an intentional subject at all times, but at least might do so within intuitively feasible limits and under intuitively favorable circumstances.

The beliefs and desires of a simple subject like this will be constraint-bound or norm-bound so far as they have to follow the pattern described. The beliefs will meet certain entry and exit conditions; for example, they will tend to enter the scene as the subject is presented with evidence supporting them, and will tend to exit when the subject is presented with evidence opposing them. Furthermore, both the beliefs and the desires of the subject will meet a range of performance conditions. The desire that p will tend to interact with the belief that if p then q so as either to generate the desire that q or to weaken the desire that p. The belief that p will tend to interact with the belief that if p then q so as either to generate the belief that q or to weaken one of the other beliefs. The desire that p will tend to interact with the belief that one can bring about p by X-ing so as to generate the action of X-ing. And so on.

Even though simple intentional subjects will have to satisfy such constraints or norms of rationality—specifically, of inferential rationality—they will not do so with any awareness of the demands that those norms support. In forming a new belief in a manner dictated by perception, they will not register that the perception makes this belief the right one to uphold. In forming a desire dictated by amendments to their beliefs or by spontaneous inclination, they will not register that these beliefs or the goal to which they are spontaneously inclined make that desire more or less appropriate. In moving to action on the basis of their beliefs and desires, they will not register that the actions they perform are the things they are to do given the ends that attract them and the opportunities available. Normative notions of what it is right to believe, what it is appropriate to desire, or what it is correct to do will have no place in their psychology. Simple inferential subjects may conform to the demands of relevant norms, but they will not conceptualize them as demands with which they ought to conform.

Thus, for the creatures envisaged, perceiving things a certain way will be nothing short of believing that they are that way. Likewise, being inclined to make things a certain way will be nothing short of acting, where possible, to make them that way—nothing short of forming an unqualified desire or disposition to act in that manner. For subjects like us human beings, perceiving things a certain way—that is, having a fallible perception of them as being that way—gives us an inferential reason to believe that they are that way, though a reason that we can override. And for subjects like us, being inclined to make things a certain way gives us an inferential reason to desire without qualification to make them that way, though again we can ignore or inhibit this reason. Perception and

inclination exist for us as states that we conceptualize as such, giving them presumptive but not necessarily conclusive authority in the determination of our final beliefs and our unqualified desires.

For the simple intentional creatures imagined here, however, perception and inclination will lack such visibility and have a drastically different role. In these creatures, perception and inclination will operate invisibly and ineluctably in the determination of how the world presents itself and how it elicits action. The creatures who are moved by certain perceptions and inclinations will be unable to see those states as we do, as presumptive but not compelling indicators of what is the case and what is to be done. Looking through their perceptions, simple intentional creatures will see the world as it is according to the beliefs that they unthinkingly form; looking through their inclinations, they will be moved to act according to the desires that they unthinkingly conceive. Such creatures will be the captive audiences of their perceptual representations and the captive executors of the ends to which they are inclined. They will be the slaves of perception and inclination.

The upshot of this discussion of simple inferential creatures is that we can hardly tell a story as to how the moral becomes salient for creatures like us merely on the basis that we are intentional subjects. There is nothing in intentionality as such that would explain why intentional subjects see things in a manner that might give rise to the introduction of normative terms and concepts. If we are to provide a naturalistic genealogy of such terms and concepts, then we must start from a richer image of human beings.

III. Discursive Subjects Will Have Normative, Inferential Concepts

A. From intentionality to discourse

We human beings are not just intentional subjects who form and act on beliefs and desires in a more or less rational manner. One of the most striking features of our species is that we are also conversational or discursive creatures.[5] Like many nonhuman animals, we form beliefs and desires; unlike nonhuman animals, however, we also give voluntary or intentional expression to the ways that things present themselves as being in the light of our beliefs and desires. We have not just the ability to believe that p, but the ability to assert that p: we can use a "voluntary sign," in Locke's phrase, to represent how things present themselves as being, given this belief.[6] Similarly, we have not just the ability to desire

[5] On this point, see Terrence W. Deacon, *The Symbolic Species: The Co-Evolution of Language and the Human Brain* (New York: W. W. Norton, 1997).

[6] John Locke, *An Essay Concerning Human Understanding*, ed. P. H. Nidditch (Oxford: Oxford University Press, 1975), bk. III, chap. 2.

that q, but the ability to assert that the prospect that q is, say, attractive: we can thus also use a voluntary sign to represent how things present themselves as being, given this desire. We can express our beliefs in regular, content-specifying sentences, and we can express our desires in sentences that ascribe attractiveness or something similar to the items desired.

That we are discursive creatures means that we have access to voluntary signs of these kinds and that we employ these signs with a view to noncoercively and noncollusively reaching a common mind on certain matters. We use voluntary signs not just to bully, impress, or threaten one another—though we may obviously do these things, too—but to discourse or reason together about certain theoretical or practical issues. Our using signs for this sort of purpose is intimately associated with our nature as human beings. Indeed, one traditional view is that the discursive use of signs is constitutively required for human beings to be able to think.[7]

In order to practice discourse, we must have access to both the concept of a sign and the associated concepts of meaning and representation. If we are to use signs as signs and do so intentionally or voluntarily—that is, out of a desire to represent or misrepresent how things are, believing that the signs we use will achieve this representational function—then we must be able to conceive of the signs as having a meaning. I do nothing here to explain how naturalistic creatures can get to have the required concept of meaning.[8] I instead assume that the concept is accessible to creatures of our kind—our naturalistic kind—and I try on the basis of this assumption to develop a story as to how we come to think in moral terms. My reliance on this assumption means that from a naturalistic point of view, there is an important debt in this essay that remains undischarged.

Without explaining how we access the concept of meaning as such, however, it will be useful to say something in explanation of how we generally manage, as discursive creatures, to give our words and sentences the same meanings, and indeed, how we do so in such a way that we know we each mean the same things, know that we know this, and so on. I suggest that we achieve commonality of meaning, at least in the general run of cases, by putting three regulative assumptions to work in our dealings with one another. The assumptions are: first, that the world of common experience makes available candidates for the meanings of our sentences and words; second, that in any speech community we each intend for our sentences and words to have the same presumptively salient meanings; and third, that ordinarily we each have the ability to use

[7] For a qualified defense of this sort of view, see Pettit, *The Common Mind*. For further elaboration of its implications, see Philip Pettit and Michael Smith, "Freedom in Belief and Desire," *Journal of Philosophy* 93, no. 6 (1996): 429–49; and Philip Pettit, *A Theory of Freedom: From the Psychology to the Politics of Agency* (Cambridge: Polity Press, 2001).

[8] For an attempt to deal with some of the issues involved in this problem, see Pettit, *The Common Mind*.

our sentences and words according to such intended meanings. These assumptions of common experience, common intent, and common competence can be seen as guiding a practice such that if it is successful, then we will indeed be using our words and sentences to say corresponding things.

So far as things go smoothly, the assumptions will lead us to ascribe to others those beliefs and desires that we would ourselves avow if we were to use the words and sentences that others employ. Things will not always go smoothly, however. Someone may use a given, apparently belief-expressing sentence such as "p" in a situation where others say, or would say, "Not p." What are we to make of such a case?

As a matter of fact, what we make of it is something that the regulative assumptions reflect well. We balk at the sort of divergence in question and assume that something more or less contingent is amiss. We see whether the divergence may be explained away by the fact that one of the parties is expressing a desire rather than a belief, by the fact that the sentence "p" is vague at a relevant margin, by the fact that it is covertly indexical in the fashion of a sentence using "I" or "now," or by some other relatively simple fact. If any such easy explanation applies, then that means that our three regulative assumptions are unchallenged, and that the default setting under which different parties' words mean the same thing has been preserved. If, however, we fail to explain away the divergence through these simpler methods, we must search for a deeper explanation that will save the assumptions and enable us to preserve the default setting.

The ideal, multilateral way of preserving the default setting would be for the parties to air the considerations that lead them on their different ways—or to seek out further relevant considerations—and thereby find themselves able to agree that one of them, as they will say after the fact, did not fully understand the terms, was underinformed, or was misled as a result of certain limitations, obstacles, or blind spots; more on these in a moment. A somewhat less satisfactory, but still multilateral, way of saving the default setting would be for the parties to agree that they have such different beliefs on collateral matters that there is no practical prospect of resolving the issue at hand, even if it is resoluble in principle. And short of achieving a multilateral resolution of the divergence, the parties might still preserve the default setting, each doing so in its own way, by unilaterally postulating that the other party is unwittingly affected by destabilizing factors and so is not fully competent on this particular occasion or on this particular topic.

What do we do if we find that even this sort of unilateral explanation of the divergence is not available? At that point, and at that point only, we will have to accept that the default setting is mistaken. In such a case, the parties have different meanings for one or more of the words in question, or perhaps the words do not have any proper meaning at all—perhaps we were wrong to suppose that there was something real for the words to ascribe.[9]

As envisaged in the story just told, the regulation provided by our three assumptions does not proceed on the basis of prior agreement as to what constitute the limitations or obstacles or blind spots, or the collateral sources of discrepancy, that are likely to destabilize discursive performance. Such destabilizing factors will be identified in the process of mutual regulation, as the parties to the enterprise gradually converge on lists of factors that can serve in any area to disturb discursive competence. The factors that the parties ought to identify as destabilizing, given the enterprise at hand, are those factors such that by indicting them as destabilizers the parties can maximize noncoercive and noncollusive convergence between one another. We can think of the collaboration that the regulative assumptions sponsor as a search procedure whereby destabilizers are identified—rightly or, as it may be, wrongly—and the assumptions are vindicated.[10]

This point is worth illustrating. Consider the ways in which you and I might coordinate in the use of a color predicate such as "red." Assuming common experience, common intent, and common competence, we will expect convergence between us in the use of this predicate. If there is divergence that we cannot explain away through use of an aforementioned "simple" explanation—say, that one of us is using "red" in the indexical sense of "looks red to me"—then we must search around for a factor that is disrupting my color-recognition competence or yours. We can think of the factors actually recognized as destabilizers of such competence—wearing colored glasses, operating in sodium lighting, being partially color-blind—as factors that are identified in the course of precisely this sort of search.

The practice just described can be summed up in a flowchart that each of us can be represented as implementing; see Figure 1 below.

I noted above that the three regulative assumptions guide a practice that, if successful, ensures that we have common meanings in mind. This point is relatively straightforward. The assumptions will successfully drive the practice, so far as we manage in cases of discrepancy to identify the destabilizers whose presence the assumptions predict; we will have to do this by way of multilateral resolution in many cases, though sometimes we will only be able to engage in unilateral explanation of one another's failures. If the assumptions are successful in this way, then the practice will make it the case that what a sign denotes is that particular element that will reliably prompt the employment of the sign in those favorable circumstances where destabilizers are not at work. What "red" refers to,

[9] On this claim, however, see ibid., chap. 2.

[10] For more on this approach, see Philip Pettit, "A Theory of Normal and Ideal Conditions," *Philosophical Studies* 96, no. 1 (1999): 21–44.

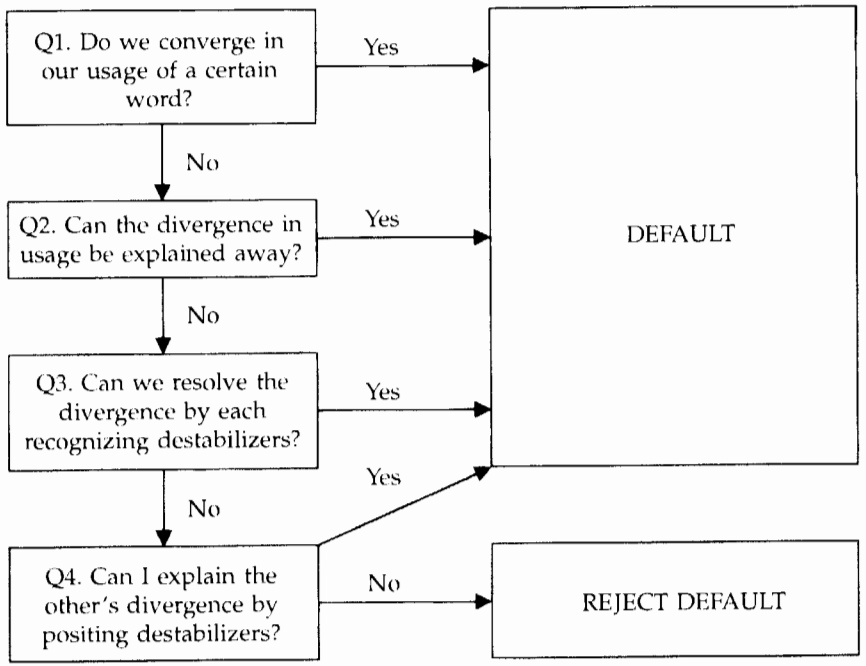

FIGURE 1

for example, is the property that makes things look red, and reliably prompts people to say "red," in circumstances where factors of the kind illustrated above are absent—that is, circumstances that are favorable for the perception of red.

Does the fact that we are discursive creatures in the sense envisaged here mean that our experience will make moral matters salient and will call for the introduction of moral terms? I go on to argue now that as discursive creatures, we must have experiences of certain responses as being normatively required, and normatively required in a way that is salient to us all. As I explain in Section IV, however, I think that these experiences leave us short of moral concepts proper.

B. The normative becomes salient

We saw above that beliefs and desires are constraint-bound or norm-bound in the sense that we only ascribe them to a subject so far as we find—at least in intuitively favorable circumstances and within feasible limits—that the states in that subject meet various design specifications (i.e., various entry, exit, and performance conditions). Given that beliefs and desires are bound to norms of inferential rationality in this way,

however, the sentences in which any one of us voluntarily expresses how things present themselves in the light of our beliefs and desires must engage in a more intimate way with corresponding norms.

Consider some of the requirements of inferential rationality, conclusive or defeasible, to which intentional subjects must generally conform, however blindly. To revert to our earlier examples, the fact that perception provides evidence that p rationally requires an intentional subject to form the belief that p. The fact that a subject believes that p and that if p then q rationally requires the subject to believe that q. The fact that a subject desires that p and believes that he or she can bring p about by X-ing rationally requires the subject to X. And so on. If the beliefs and desires of intentional subjects have to satisfy these sorts of inferential norms, then the same is going to be true of the sentences in which discursive subjects like us express these beliefs and desires; otherwise, the use of these sentences will not amount to serious and sincere expression. Just as the belief that p has to be sensitive to evidence supporting or refuting p, for example, the serious and sincere assertoric use of the sentence "p" will have to be equally sensitive to such evidence. Furthermore, just as the desire that p has to be sensitive to the things the agent believes and the things toward which the agent is inclined, the serious and sincere expression of the desire that p through use of a sentence such as "The prospect of p is attractive" will also have to be sensitive to those matters—presumably, to those matters as they themselves are capable of being expressed in sentential usage.

We can imagine beliefs and desires evolving on an involuntary, mechanical basis so as to keep in step with the required demands of rationality. In such cases, a subject's beliefs and desires will be sensitive to the demands of inferential norms without this sensitivity being controlled by beliefs as to what these demands are. But we cannot envisage our human performance with voluntary signs as displaying sensitivity to norms in the same blind fashion. While beliefs and desires may materialize and mutate within intentional creatures in a norm-sensitive manner to which the creatures' beliefs are irrelevant, nothing of the kind can be true of the ways in which linguistic representations unfold.

The reason for this, at base, is that how these representations unfold will have to reflect our beliefs as to what they mean; this goes back to the undischarged assumption that we have access to the concept of meaning. These beliefs amount to nothing more or less than beliefs as to when we ought and ought not employ the signs. To know the meaning of a sign as such, after all, is precisely to be able to tell, with greater or lesser exactitude, when it is right to use the sign and when it is not. If our sentences unfold under the control of such beliefs as to what they mean, then this is to say that our employment of the representations is controlled by the relevant norms instead of merely being sensitive to them; it is guided by our beliefs as to the demands of the norms.

Consider the inferential norms mentioned earlier in this section, putting aside complications to do with their defeasibility. According to the first norm, the fact that perception provides evidence in support of p rationally requires an intentional subject to form the corresponding belief. Whereas merely intentional subjects will blindly conform to this requirement and form the belief that p, discursive subjects will go further by endorsing the corresponding voluntary representation under the control of a belief that the perception makes it right to say that p. According to the second norm, the fact that an intentional subject believes that p and that if p then q rationally requires the subject to believe that q. Merely intentional subjects will again blindly conform to this requirement and form the belief, whereas discursive subjects will endorse the corresponding voluntary representation under the control of a belief that the fact that p and that if p then q, as they are taken to be, make it right to say that q. Finally, according to the third norm, the fact that a subject desires that p and believes that he or she can bring p about by X-ing rationally requires the subject to X. Analogously with the other cases, merely intentional subjects will blindly conform to this requirement and adopt the required action, while discursive subjects will adopt that action under the control of a belief that the attractiveness of the prospect of p, combined with the linkage between X-ing and this prospect, makes X-ing the thing to do.

This discussion shows, then, that when we practice voluntary signing or expression, we must become aware of certain normative requirements—namely, the demands of inferential rationality on our performance. As well as being able to register what inferential norms demand, of course, we must be able to regulate our responses as the norms require; that is, we must be able to regulate suitably both our utterances and the beliefs, desires, and actions associated with them. Otherwise, we could not successfully practice discourse; there would be no point in anyone's talking to us. We must be able to think that this or that is what we ought to say and ought to believe, desire, or do, and we must be able to keep ourselves in line with such normative thoughts, saying what we think we ought to say, and believing, desiring, and doing what we think we ought to believe, desire, and do. In short, we must be able to ratiocinate or reason, or at any rate we must be able to do so under favorable circumstances and within feasible constraints, however they are identified.[11]

Let me offer one last word in qualification of this picture. Consistently with being governed by reason in this way, the signs that I utter may be deployed in a relatively habitual and unmonitored manner. All that is required is that even if habit or inertia rules the things I say in expression of my beliefs or desires, or the things I do on the basis of these representations, I will intentionally review and amend the things that I say or do

[11] On this point, see Victoria McGeer and Philip Pettit, "The Self-Regulating Mind," forthcoming in *Language and Communication*.

in the event of any inferential irrationality becoming apparent; in relations of discourse it may be assumed that such irrationalities can always become apparent. The governance exercised by reason may be virtual rather than active, in other words. It may be that habit explains many of the things that we say in discourse, with reason remaining on standby, ready to intervene as soon as there is a challenge to the inferential rationality of those utterances.[12]

IV. Discursive Subjects May Lack Moral Concepts

Let us grant that discursive subjects will have to be aware of the demands made in different cases by shared norms of inferential rationality. Let us also grant that to this extent discursive subjects must be in possession of normative concepts. What we now have to notice, however, is that having access in this way to normative concepts does not involve having access to moral concepts.

What sorts of normative considerations will discursive subjects be able to access in the formation of their beliefs and desires, and in the actions they choose to perform? The facts that p and that if p then q, as they see them, may give them reason to believe that q. The fact that some prospect r is attractive and that r only if s may give them reason to desire that s. The fact that the r-prospect is attractive together with the fact that they can bring about r by X-ing may give them reason to perform the action of X-ing. Given that facts of this sort are admitted—and conceptualized in common terms—discursive subjects will generally be able to invoke common norms of inferential rationality to justify the moves they make to others. They may occasionally diverge from one another in such moves, but divergence of that kind will always cause them to balk and to try to explain or resolve the difference between them; the differences will not be regarded with insouciance, as if they were fault-free.[13] That the moves illustrated are subject to the rule of reason—in this case, the rule of inferential reason—means precisely that divergences in those moves must always raise the possibility that someone is at fault in the move they make.

It is important to note that all of the normative considerations available in this way to discursive subjects involve belief. They bear on what the subject should come to believe, period; or on what the subject, given an existing preference, should come to desire—or come to do—in virtue of a change of belief. Conspicuously, however, they do not include considerations that might support the formation or revision of any particular basic

[12] For discussion of this notion of virtual control, see Philip Pettit, "The Virtual Reality of *Homo Economicus*," *The Monist* 78, no. 3 (1995): 308–29.

[13] I borrow this notion of fault-freedom from Huw Price, *Facts and the Function of Truth* (Oxford: Blackwell, 1988).

desire: that is, any desire of the kind that is not explained by the agent's preexisting preferences together with his or her beliefs as to how those preferences are best satisfied.

This should not be surprising. While divergences among discursive parties with respect to the beliefs they form cannot be regarded as fault-free—at least not if the beliefs are determinate and nonindexical—divergences among them with respect to their basic desires certainly can be regarded in this way. As noted above, discursive subjects differ from intentional subjects in that discursive subjects express their beliefs and desires voluntarily. Given the function of belief in representing the ways things are, beliefs cannot rationally diverge, and therefore discursive subjects' expressions of belief cannot rationally diverge either. Because desire has a directive rather than a representational function, however, desires can rationally diverge, and discursive subjects can rationally support divergent expressions of desire. As a result of this difference between expressions of belief and desire, discursive subjects will be under pressure to provide reasons for holding beliefs that diverge from those of others, but will be under no pressure to provide reasons for holding desires that diverge from those of others.

For all that the story so far suggests, the divergence that is possible in desire may be so radical that there are cases in which we cannot see the desire evinced in another's actions and words as an intelligible desire to embrace. In such a case, not only may we not ourselves feel the desire in question, but we may not even be able to see how we could be moved by it if we were in the agent's position. The desire might be as silly as the basic, unintelligible inclination, in Elizabeth Anscombe's famous example, to possess a saucer of mud.[14]

The upshot of this analysis of desire is that while discursive creatures can express desires in a way that provides an inferential grounding for the formation of other desires and the performance of certain actions, they may lack a capacity to do anything about resolving differences between those whose basic desires diverge. This is because discursive creatures may lack the capacity to provide reasons in support of such desires. Discursive creatures may try to resolve practical as well as theoretical predicaments in a discursive way. However, they will resolve shared predicaments only so far as the relevant parties happen to have convergent desires, and they will resolve individual predicaments only so far as they take the desires of the individual in question—even the basic desire for a saucer of mud—as an unchallengable given. The considerations that express the desires of discursive creatures may be action-supporting in a distinctive way, but they will not themselves be reason-supported; there will be no question of justifying the desires these considerations express by invoking something that makes these desires the right ones for an agent to have.

[14] G. E. M. Anscombe, *Intention* (Oxford: Blackwell, 1957).

I sum up this result by saying that the discursive creatures envisaged so far will have normative, inferential concepts available to them, but they will not have access to moral concepts proper. A moral or ethical consideration is meant to be able to provide an inferential reason, conclusive or defeasible, for forming a certain desire and performing a corresponding action: it does this in the same way that the normative consideration that something is attractive—the sort of consideration that expresses a desire—may provide a similar inferential reason. However, a moral consideration is also supposed to be a sort of consideration such that divergence between certain parties in regard to whether it is endorsed—divergence on whether it is believed—is not a matter of indifference; it is not something that is necessarily fault-free. A moral consideration is meant to be not just an action-supporting consideration, but also a reason-supported one.

Moral considerations will satisfy this dual condition of being reason-supported as well as action-supporting as long as one of two things is assumed by all those invoking them: either that they are agent-neutral considerations that everyone has some reason to endorse, or that they are agent-relative considerations that anyone in the agent's position would have some reason to endorse.[15] Examples of agent-neutral considerations include "This is the kind thing to do," "This is the only just response," and "This is for the overall good." In contrast, examples of agent-relative considerations include "This will bring me pleasure," "This will help my children," and "This will be for the good of my friends."

It is not particularly contentious to say that divergence with respect to moral considerations is not treated as a fault-free matter of indifference among those who invoke them. The role played by moral considerations in supporting an appraisal of what people do, and the associated role they play in enabling people to justify their desires and actions to one another, require that divergence over such considerations suggest fault. In order to admit the point, we do not have to think that every moral divergence lends itself to mutually satisfactory resolution; as noncognitivists usually say, there may be some values with respect to which people will differ in a fault-free way. All that has to be allowed is that the considerations that people invoke in a moral key generally invite discussion, and that such discussion generally holds out a prospect of leading to noncoercive, noncollusive agreement. In other words, all that has to be accepted is that divergences over moral considerations generally allow of discursive resolution.

[15] An agent-neutral consideration is expressed in general terms, and can be fully understood independently of who the agent is for whom it is a consideration. An agent-relative consideration is expressed in indexical terms—terms like "me" or "my" or "mine"—and cannot be fully understood independently of who the agent in question is. If you overhear me say that I am moved by a concern for the general happiness, you can know all there is to know about which consideration is in question without knowing who is speaking. If you overhear me say that I am moved by a concern for the welfare of my children, you do not know all there is to know about which consideration is in question without knowing who is speaking, and in particular without knowing which children are to be favored.

The argument of this section has shown that considered in the abstract, discursive subjects are capable of having divergent desires without this divergence offending against any norms that they recognize; they may lack the moral concepts that would raise questions about it. The upshot is that we have to look beyond our status as discursive creatures in order to explain how the morally normative, as opposed to the merely inferentially normative, becomes salient to us; I do this in the next two sections. Before going on to those discussions, however, it is worth mentioning that the previous section does not leave us quite as badly off as might be thought. While the discussion of discursive subjects does not explain how moral terms or concepts materialize, it does take us well past the point at which the discussion of intentionality left us.

I argued in Section II that simple intentional subjects would have no awareness, as such, of their perceptions and inclinations. For such subjects, perceiving things a certain way would be nothing short of believing them to be that way; analogously, being inclined to make things a certain way would be nothing short of being decisively disposed—desiring, without qualification—to make them that way. Given these observations, it appeared that there would be no call in such creatures for an awareness of perception as providing a reason for belief, or of inclination as providing a reason for (unqualified) desire; perception and inclination would operate in these creatures with an invisible, ineluctable force, and would determine their beliefs and their desires.

The thing to notice about the discursive image of human beings, as distinct from the merely intentional image, is that just as the discursive image is not consistent with perception maintaining a silent, irresistible role in the determination of people's beliefs, it is also not consistent with inclination remaining in such a position with respect to people's desires. The discursive image requires that humans be able to become aware of inclinations as such, other people's as well as their own.

Assume that people use signs that express how the world presents itself in the light of their perceptions, saying on the basis of perception that this or that is how things are. It is inevitable that sooner or later others will differ from them, finding different claims perceptually compelling. This means that people are bound to become aware that their perceptions are not incontrovertible evidence that things are this or that way. Making sense of divergence with others under the regulative assumptions of discourse will force people to recognize their perceptions—and the perceptions of others—for what they really are: representations of how things are that may go wide of the mark. This recognition will not give people a perception of perception—it will not bring an extra experiential faculty into play—but it will force them to notice and conceptualize the perception that has always been there. The "glass" of perception will fog up, as it were, and cease to exist as a purely diaphanous, unnoticed medium.

As it is with perception, so it is with inclination. Discourse in which people give voice to their perceptions makes clear, according to the account just presented, that how things are in the world is not an immediate given; it is given only through perception. Likewise, discourse in which people give voice to their inclinations, speaking of what is attractive and what is unattractive, will make equally clear that the way the world inclines them to act is not an unquestioned given either; it is a product of what they will have to see as inclination. They will previously have looked through their inclinations, unaware of them as such, and seen the world as a single field for action. After engaging in discourse, they will be sensitized to the fact that the world presents itself as a different sort of field to different discursive participants. For people to make sense of this, they will have to recognize that the way the world elicits action from each of them is a function of the inclinations that each of them brings to the world, not just a reflection of what is available in common to all.

As I argued earlier in this section, there will be no pressure on people to avoid divergence in their basic desires, though there will be pressure to avoid divergence in ultimate beliefs. For all that we need suppose up to this point in our story, people may be generally reconciled to competition and conflict in the realm of desire; they may resolve practical predicaments only so far as they are led by agreed matters of belief, and their existing inclinations, to a common desire. The mere fact that people use signs to express how the world presents itself in light of their inclinations will be enough on its own to make inclination visible. This will reveal to people that the world is not an action-generating environment with which they are in immediate contact any more than it is a belief-generating environment with which they are in immediate contact. It will force them to see themselves and one another not just as centers of individually variable representation, but also as centers of individually variable inclination.

This observation establishes a negative element in the genealogy of moral concepts. It shows that discursive creatures, unlike merely intentional ones, are not slaves of unnoticed inclinations in a way that rules out the appearance or relevance of moral considerations. We must now turn from this negative argument, however, to an attempt to explain how, more positively, discursive creatures can get to be sensitized to moral considerations that are capable of regulating inclination and desire. I argue that there are two forms of experience that will serve discursive creatures in this respect.

V. Privileging Discourse Will Give Rise to Certain Moral Concepts

So far as people enter discourse with one another, so I have argued, they will have to become aware as such of their own inclinations and the

rival inclinations of others. How can people take a further step and come to access action-guiding considerations that are capable of directing inclination in the distinctively moral or ethical manner? What experiences available to them in the discursive scenario sketched above could give rise to terms and concepts with the dual capacity, distinctive of the moral, to support action and be supported by reason? In this section, I describe one more or less obvious way in which certain experiences can perform this task, and in the next I move on to another.

Under the picture presented in Section IV, people are inevitably going to be involved at various times in discourse with one another. While many conversations serve nondiscursive purposes, conversation will often be devoted to the resolution of discursive questions as to how a certain theoretical puzzle or practical predicament should be resolved. Some of the questions may concern what one or another party to the conversation should think or do, while others may focus on what the group as a whole—be it a group of two, three, or more—should think or do collectively. When people take part in such a conversation, they will put their heads together and reason with one another about how the puzzle or predicament should be resolved.

This sort of discourse embodies the recognition of rational, inferential norms that describe when it is appropriate, in view of how things are perceptually presented, to endorse a given consideration; when it is appropriate to endorse two or more considerations at the same time; when it is appropriate to move from the endorsement of one or more considerations to the endorsement of a further consideration; and when it is appropriate to move from the endorsement of certain considerations to the performance of a particular action. We could not treat another person as a discursive partner if he or she were not generally able to recognize and respond to what such norms require. If he or she failed to internalize inferential discursive norms in this way, then we might as well be talking to the dog or the wall.[16]

Given that discursive conversation of this kind has to be governed by rational, *inferential* norms of discourse, it follows that those taking part in it will have to conform also to corresponding *interactive* norms of discourse. If I am reasoning together with you about something, then the discursive conversation in which we are involved requires by its very nature that all the overtures we make be consistent with the joint endeavor's being governed by rational, inferential norms in the manner envisaged. Thus, if I put forward a consideration that I know to be false, I am not behaving in line with this endeavor. Likewise, I am not behaving in line with the endeavor if I insincerely voice an intention as to what I will do in the future, or if I make a coercive threat to the effect that I will punish you unless you agree to something. Reasoning together, by its

[16] On this point, see Pettit and Smith, "Freedom in Belief and Desire."

very nature, is inconsistent with such modes of conduct among the parties involved. If any of them behave in these ways, then whatever transpires between them cannot be described as an instance of reasoning in common about a discursive question.

Not only will people have to conform to interactive norms of discursive conversation—that is, the norms that derive from the more basic inferential norms. People who take part in discursive conversation must also be aware of what these interactive norms require, at least on a case-by-case basis, and they must prove generally responsive to these requirements. Being able to take part in such conversation is a skill that people learn, not something that comes by way of automatism and reflex. In learning this skill, it is essential that people are enabled to recognize and reject the sorts of deceptive, manipulative, and coercive moves that are inconsistent with the exercise of reasoning with others about the resolution of a theoretical or practical issue. People will be free to deceive or not to deceive, to manipulate or not to manipulate, to coerce or not to coerce, but they must be able to see the deceptive, manipulative, or coercive option as being inconsistent with maintaining a discursive posture toward others. And of course, they must be able to respond to this perception and prove themselves capable of treating others in the manner of a fellow reasoner.

If, however, people are to be able to see certain types of options as being inconsistent with the discursive posture—the posture of mutual address, as Stephen Darwall calls it[17]—then they must be in a position to introduce terms with ethical force in order to describe the options in question. In what follows, these terms will be represented by the letter "O" (for "objectionable"). To the extent that discourse and address are practices in which people engage, the consideration that an option is O will have the profile, action-supporting and reason-supported, of a moral consideration.

The consideration that an option is O will serve to direct or support action to the extent that the consideration gives expression to an unquestioned, willing involvement in the discursive practice that makes the O-concept available in the first place. To think, situating oneself within discourse, that a certain sort of action is O will be to register a reason, conclusive or defeasible, for not performing that action. It will be like thinking, as one stands behind a certain inclination or desire, that things inconsistent with the satisfaction of that desire are unattractive.

However, the consideration that an option is O will also be capable of being supported by reason, unlike, say, the consideration that things inconsistent with possessing a saucer of mud are unattractive. There is nothing that a person can do to provide support for the basic desire for a saucer of mud. Nothing about that desire, so we have been supposing,

[17] Stephen Darwall, "Reciprocal Recognition: The Second-Person Standpoint in Moral Thought and Theory" (manuscript).

makes it something that it is in any sense right—right, by common lights—to embrace. Things are very different, however, when it comes to the desire for the sort of discursive involvement that is presupposed when one treats O-considerations as action-guiding. Such involvement is bound to recommend itself to people generally, given their discursive nature. Therefore, every individual will be in a position to justify their treating the fact that an option is O as a reason for not performing the corresponding action; every person can ground such treatment in the fact—however they express it—that such treatment is required for maintaining the discursive or addressive stance.

O-type predicates may be dependent on discourse and address in either of two ways, both of which are consistent with this observation. Someone who says that an option is O may be speaking from a neutral or outside perspective, in which case he or she is saying that the option is inconsistent with address—that is, the option is ruled out by the discursive practice of mutual address. Alternatively, one who says that an option is O might be speaking from within the perspective of mutual address, thus taking that perspective as given and putting its relevance beyond current question. In this case, one is saying that the option is ruled out, period, though perhaps only ruled out defeasibly rather than conclusively. The first construal would make O-type terms *relational* in character; the second would make them *perspectival* or *locational*.

I assume that O-type terms and concepts are going to be address-dependent in the locational way. This construal is more economical than is the relational, since it does not require people to have the explicit concept of discourse or address; it requires only that people be involved willingly and unquestioningly in discursive practice. In virtue of occupying the discursive stance, people will see certain types of actions as O, but they may not notice the relationship with discourse in virtue of which the actions present themselves as O. The discursive stance will cast a shadow on relevant action-types, as it were, and make them saliently available for conceptualization as O. Furthermore, it will do this without people necessarily being aware of the role that the stance plays in casting this shadow.

There is another advantage, apart from this economy, in taking O-type terms to be locational rather than relational in character. The locational construal makes sense of why people are spontaneously prepared to see actions performed outside the context of discursive address as O; it expands the domain in which O-type terms are naturally and nontendentiously applied. Taking up the discursive stance, I may think that my doing something to another person would be O, even if at the time I had no interest in addressing the person affected; in thinking this, I am spontaneously and understandably privileging the discursive stance I currently occupy. And in the same way that I may think this sort of thing about actions that I might myself perform outside of a context of address, so I may think it of actions that others are currently performing outside of such a context. Having no hesitation about privileging the context of address that I occupy myself, I may be quite happy to denounce those actions as O.

The elements of a genealogy of moral concepts are now in place. Discourse involves something that we may hope is naturalistically unobjectionable: the exchange of common signs in voluntary expression of intentional states. Those within the discursive stance will have to see the use of these signs as being subject to rational, inferential norms, as I argued in Section IV. And, as I have just argued in this section, these people will have to see activities that engage with the discursive exchange of such signs—activities that may or may not be consistent with practicing discourse and address—as being subject to interactive norms and corresponding moral considerations.

I think that this account should prove reasonably persuasive for anyone who accepts the discursive image sketched in Section III. However, even if the account is mistaken in some respects, it should still serve to render ethical conceptualization intelligible. The fact that people could become sensitized to moral considerations in the manner described here at least establishes that there is nothing inherently mysterious, even under broadly naturalistic assumptions, about their having access to moral terms and concepts. Let the account be entirely hypothetical in character, and it will still serve to render the appearance of ethical terms and concepts intelligible.[18] Donald Davidson has claimed to render our semantic competence with respect to an indefinite number of sentences intelligible by showing that it could have come about—though actually it certainly did not—through our mastering a certain sort of theory for the language in question.[19] We might claim that the account provided above, even if it is not true to the actual evolution of moral concepts, can render moral conceptualization intelligible in the same way.

Rather than end this discussion of address-dependent concepts on a weak note, however, I would like to emphasize how much discursive address distinguishes and enhances human life and how natural it is that it should provide the entrée to ethical talk and thought. It is possible for one person to exercise a substantive influence over another in a discursive exchange by proving to be better-informed, clearer-headed, more persuasive, or whatever. The influence thereby exercised is quite different from deceptive, manipulative, or coercive influence, or from any sort of influence found elsewhere in the sentient world. It is of a kind that must be welcomed by the person influenced, so far as he or she is genuinely concerned with the discursive issue under discussion. It is also of a kind

[18] It would be entirely hypothetical under the story told in Robert B. Brandom, *Making It Explicit* (Cambridge, MA: Harvard University Press, 1994).

[19] Donald Davidson, *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984).

that in no way detracts from the fitness of the influenced person to be held responsible for whatever he or she does; it does not undermine or diminish his or her autonomy as an agent.[20] If discursive exchange is the locus at which ethical terms and concepts first gain life, it is as fitting a locus as we could ever imagine.

VI. Discursifying Sentiment Will Give Rise to Other Moral Concepts

In the account just provided, people take the practice of discourse as given and from the privileged perspective that it supplies they distinguish activities that are dissonant with discursive address from activities that are consonant with it, indicting the former as O. In other words, people privilege discourse and let the discursive stance partition activities into those that are O and those that are not.

Can we see how people might develop ethical terms and concepts apart from the address-dependent ones that this privileging of discourse will explain? Is there any other source of ethical conceptualization that we can identify on the basis of the account provided so far? I mentioned earlier that a person may see the attractiveness of an option as an inferential reason to choose it without that consideration being supported in the manner required for ethical force—that is, without there being any reason, accessible to the agent, for why he or she should endorse that consideration as a reason for action. The question now is whether people can be expected to develop concepts whereby some such considerations can be invested with reason-supported as well as action-supporting status, on the model of address-dependent considerations.

I think that we can expect people to develop such concepts. They will be given reason enough to do so by the fact that, having entered ethical space—having come to see the possibility of regulating inclination on the basis of mutually endorsed considerations—they are bound to be attracted by the possibility of the richer coordination of their activities that further ethical concepts would hold out. Moreover, as there is reason for people to try to develop further ethical concepts, so there is also going to be a resource available whereby they can achieve this goal.

For any option that an agent might see as attractive, people will be in a position to support the action-supporting reason in question—the presence of the inclination—so far as they converge on standards for when it is appropriate to form and act on the inclination. This means that people will be in a position to endorse the action-supporting reason provided by any inclination to the extent that their assessments converge with respect to a property that the inclination can be generally supposed to track. The possibility of finding standards for when embarrassment or fear is appropriate, for example, is just the possibility of finding a property—that of being truly embarrassing or truly frightening, as we may say—that feelings of embarrassment or fear are supposed to track.

[20] On this point, see Pettit, *A Theory of Freedom*.

This observation points us toward a search procedure whereby people might aspire to single out inclinations that can generate ethical considerations. People might work with the assumption that a given sentiment, such as embarrassment or fear, has the status of representing some property commonly available to all—the truly embarrassing or the truly frightening—and see how far this assumption proves workable. The assumption will prove workable if people generally find it possible to converge in their responses involving that sentiment and, where they diverge, to resolve or explain the divergence in a manner that is consistent with the sentiment's having representational status. The assumption will prove workable with fear, for example, so far as people tend to call the same scenarios "frightening" and find themselves able, in cases of divergence, to agree on which of the discrepant responses should be discounted.

The search procedure envisaged here can be more precisely characterized in abstract terms. Suppose that people tend to experience a sentiment S according to a relatively common pattern—that is, facing certain stimuli under certain conditions routinely leads people to experience S. People in such circumstances might wish to test S for its capacity to be supported by reason and to serve in an ethical role. They can do this by introducing a predicate "D" (for "desirable") as an expression of S in the same way that "frightening" is an expression of fear, and then seeing if they can subject "D" effectively to the three regulative assumptions involved in common discursive practice. They can try, that is, to determine whether they can "discursify" the predicate "D," enabling it to function in a way that parallels regular descriptive predicates. The question to be tested in the exercise will be whether people, while taking "D" as an expression of the sentiment S, can still successfully regulate their discursive practice by the assumptions (1) that there is an agent-neutral or agent-relative property, D, waiting to be tracked by "D"; (2) that each individual using "D" intends to track the same, presumptively salient property—whatever that is—in his or her use of "D"; and (3) that each individual is generally competent in his or her tracking attempts.

Let us think about how the process might go. People will use "D" in expression of S, intending thereby to mark the presence of a property that provides a reason for ascribing D and holding S. So far as people manage to converge in their usage of "D," all will go well. And so far as they diverge, all will still go well if they can sustain the regulative assumptions. They will be able to sustain these assumptions to the extent that they can come to agree, case by case, on the sorts of factors that destabilize discursive performance and give rise to divergence. In other words, they will be able to sustain these assumptions to the extent that they have a shared means of identifying certain conditions—call them C—that are

free of such destabilizing influences and are favorable for detecting the presence of the property D, and hence for endorsing the sentiment S.

There will be a fact of the matter, under this picture, as to whether something is D or not, and this fact will be independent of whether the subject feels S or not. This is why the picture can make room for thinking that the consideration expressed by "D," which will provide an inferential reason for the agent to act as S prompts, may or may not be deserving of endorsement. To the extent that limitations, obstacles, blind spots, and the like are possible—to the extent that C-conditions may not obtain—it will be possible for a person to feel S when the sentiment is rationally ruled out and not to feel S when it is rationally ruled in. Thus, it will be possible for people to feel S when the object of their sentiment is not D—as they may themselves be able to see—and not to feel S when the object of their sentiment—again, as they may be able to see—is D. The connection between the sentiment S and the property D will not be that something has the property if and only if it produces S in a relevant person. Rather, it will be that something has the property if and only if it produces S in a relevant person under the C-conditions where the person's response is not in any way destabilized.

This story, it should be noticed, is quite consistent with people's thinking of the property D as a feature of the world in itself, or of the world in the way it impacts on the agent, without their registering its connection to a sentiment. Just as O-type terms can be dependent on address without being relational terms that express the relationship of being inconsistent with address, so D-type terms can be dependent on sentiments without being relational terms that express the relationship of occasioning sentiment. As in the case of address-dependent concepts, it is more economical to assume that these sentiment-dependent concepts are perspectival or locational, since this means that the registering and conceptualization of D-properties will not require a prior registering and conceptualization of the S-sentiments or the C-conditions. Those who use the D-type terms need not have distinguished C-conditions as a general type, they need not have identified or named the S-sentiments within themselves, and they need not be aware of the fact that the things they call "D" give rise to S under C-conditions.[21] As was the case with O-type terms, the assumption that D-type terms are locational also has the effect of expanding the potential domain in which these terms can be spontaneously applied. I may think that my doing something would be D, even were I not at the time to feel or endorse the sentiment S, and I may think that the same is true of others' currently acting in that way, even others who do not feel or endorse the sentiment.

[21] See Philip Pettit, "Realism and Response-Dependence," *Mind* 100, no. 4 (1991): 587–626; and Philip Pettit, "Terms, Things, and Response-Dependence," *European Review of Philosophy* 3 (1998): 61–72.

What might be examples of D-type terms? Some will be terms that answer determinately to recognizable sentiments, such as "frightening," "embarrassing," "thrilling," "nauseating," "humiliating," and "inspiring." Others will be similar to these except so far as they mark more explicitly the fact that you can have a sentiment without the relevant predicate being applicable—as, of course, you can be frightened without anything being truly frightening—and vice versa. In this second category are terms such as "admirable," "honorable," "desirable," and the like, as well as negative counterparts such as "abominable," "dishonorable," and "undesirable." Still other terms will focus not on the sentiment involved in the response, and not on the fact that this may or may not be rightly held, but on the sort of action, person, or state of affairs that attracts the positive or negative sentiment in question. In this third category are more descriptively laden terms such as "kind" and "cruel," "fair" and "biased," "courageous" and "cowardly," "prudent" and "imprudent," and so on.

Notice that among these examples of D-type terms we should expect to find terms that register aspects of experience and interaction that are dependent on the practice of discourse itself and the prior development of O-type terms. So far as O-ness becomes salient in the light of discursive practice, and so far as it is more or less bound to be motivating, it is bound to engage with our sentiments; plausibly, it will engage with the phenomenologically distinctive sentiments of resentment and indignation to which Peter Strawson draws attention.[22] To the extent that those actions that present themselves as O—deception, manipulation, coercion, and the like—inherit these features of O-ness, they are likely also to attract autonomous sentiments of disapproval. We should be ready to admit that there will inevitably be interactions of these kinds between address-dependent and sentiment-dependent ethical concepts.

We have seen why terms or concepts that fit with the D-schema should have the dual capacity that is distinctive of the ethical. D-considerations, being expressive of sentiments in the way in which "attractive" is expressive of inclination, will provide inferential reasons, conclusive or defeasible, for acting. D-considerations will also be supported by reasons, so far as there will be reason to endorse them in those cases where they truly obtain. Whereas O-considerations are likely to be agent-neutral in character, presenting reasons that all of us have reason to endorse, D-considerations are going to be agent-relative as often as they are agent-neutral. They will include considerations to do with the children or projects for which the agent has a special sentiment—that is, his or her own—as well as considerations to do with matters in respect of which we may all expect to feel roughly the same sentiments.

[22] Peter Strawson, "Freedom and Resentment," in Gary Watson, ed., *Free Will* (Oxford: Oxford University Press, 1982).

The account given of O-type terms in Section V explained how creatures like us, so far as we practice discourse and privilege the addressive stance, will experience certain activities as being morally objectionable or, of course, unobjectionable. The account just given of D-type terms plays a similar genealogical role. It explains how discursive creatures like us, already familiar with ethical concepts, can privilege certain sentiments—those that we find capable of being "discursified"—and can experience the objects of these sentiments as being morally desirable or undesirable. Under the account, an arbitrary X will be D if and only if X is disposed to produce S in conditions C. What this formula communicates is that concepts of the D-type will be available in canonical or nonparasitic mode only to those creatures who are capable of having the S-response and who discursify this response in the sense of standing by it just in those conditions that prove to be C.

Just as the earlier genealogy sheds light on the possibility of ethical conceptualization even if it is not accurate in its details, the same applies in this case as well. It is worth mentioning, however, that the genealogy provided in this section has a long history of support and may be regarded for this reason as having particularly telling claims. The genealogy conforms to a narrative of sentiment and training in sentiment that was common currency during the Enlightenment period, being best known today through the work of David Hume.[23] It also fits well with the work of those contemporary metaethical thinkers who remain faithful to the broadly Humean framework.[24]

VII. Conclusion

The epistemology of ethics needs to give some account of how ethical judgments are to be confirmed, but it also has to address the question of how ethical conceptualization can arise from experience. The experience of intentional subjects cannot give rise as such to moral conceptualization,

[23] David Hume, *"Of the Standard of Taste" and Other Essays*, ed. John W. Lenz (Indianapolis, IN: Bobbs-Merrill, 1965); David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2d ed. (Oxford: Oxford University Press, 1978); and David Hume, *An Inquiry Concerning the Principles of Morals*, ed. J. B. Schneewind (Indianapolis, IN: Hackett, 1983). See also Geoffrey Sayre-McCord, "On Why Hume's 'General Point of View' Isn't Ideal—and Shouldn't Be," *Social Philosophy and Policy* 11, no. 1 (1994): 202–28; Stephen Darwall, *The British Moralists and the Internal "Ought"* (Cambridge: Cambridge University Press, 1995); and J. B. Schneewind, *The Invention of Autonomy* (Cambridge: Cambridge University Press, 1998). Writers who emphasize the possibility of reading Hume on cognitivist lines, so that his approach would mirror that taken in this essay, include J. L. Mackie, *Hume's Moral Theory* (London: Routledge, 1980); R. M. Sainsbury, "Projections and Relations," *The Monist* 81, no. 1 (1998): 133–60; and Peter Railton, "Taste and Value," in Roger Crisp and Brad Hooker, eds., *Well-Being and Morality: Essays in Honour of James Griffin* (Oxford: Oxford University Press, 2000).

[24] For a modern development of the Humean approach, see, for example, Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994). Whereas Smith's approach gives prominence to the notion of an ideal point of view, however, it is arguable that Hume relied instead on the notion of a general point of view. For a fine case in support of this interpretation, see Sayre-McCord, "On Why Hume's 'General Point of View' Isn't Ideal."

but the experience of discursive creatures certainly can. The experience of discursive creatures will inevitably occasion the appearance of normative concepts, since these are required for reasoning. Such experience also has the capacity to give rise to properly moral conceptualization. It will do this first by making salient the demands presupposed by the practice of discursive address, and second by fostering the "discursification" of certain sentiments.

These two genealogical stories are both fairly persuasive to my eye, though each needs more detailed development. Distinguishing them is important, because the address-dependent and sentiment-dependent concepts that they respectively underpin are quite different in character. Any comprehensive moral theory has to recognize the place of each sort of concept, and has to tell a story about how the different demands in question should be reconciled. The relative priorities a moral theory assigns to each of these types of concepts will play a large role in determining the overall shape of the theory.

The address-dependent concepts are, broadly speaking, deontological in character. The question to which they relate bears on whether an action is in breach of discursive norms—whether it is *deon*, to use the Greek word, or owing to another—and it applies only to the actions that people take in relation to others. It may be possible to treat such action-centered, negative considerations in a consequentialist rather than a nonconsequentialist way: someone might think that right action requires the maximization of nonobjectionable behavior toward others, for example, even if this means that one must sometimes engage in objectionable behavior oneself. The normal response to such considerations, however, is going to be nonconsequentialist in character. It is going to recommend that people seek to instantiate nonobjectionable properties, even at a cost in the extent to which the properties are realized overall; such an approach is supported in traditional theories of obligations and rights, as well as in more recent versions of contractual theory.[25]

The sentiment-dependent concepts, in contrast, are axiological rather than deontological in nature. They relate to the question of what is for the best—what is *axios* or desirable—which applies in principle to almost any variety of item: not just to actions, for example, but also to people, motives, and situations. One sort of axiological approach that such concepts might prompt is virtue-ethical in character, focusing on the properties that virtuous agents ought to instantiate. However, the axiological approach is just as likely to assume a consequentialist form, directing our attention instead to the properties that people ought to promote, even if promoting them sometimes means not instantiating them in their own behavior, psychology, or relationships.

[25] On obligations and rights, see Ronald Dworkin, *Taking Rights Seriously* (London: Duckworth, 1978). On contractualism, see T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998).

The argument of this essay suggests that address-dependent, essentially deontological concepts have a certain primacy in moral thought, representing the ethical considerations that introduce us—and that may be uniquely capable of introducing us—to the space of ethical reasons. This suggestion is consistent, of course, with the view that issues of what it is right or wrong to do are ultimately determined by what is for the best overall—that is, by consequentialist considerations. I defend such a view myself, believing that however serious the demands associated with the requirements of address—however serious the demands of civility or comity, as we might call them—they will have to be overridden in circumstances where it is clear that honoring them is not for the best.[26]

I believe that the debate between different approaches to ethical theory, consequentialist and nonconsequentialist, should benefit from recognition of the very different currencies provided for ethical thought by address-dependent and sentiment-dependent concepts. It may be that a failure to recognize this divergence in conceptual currency explains attachments on either side of the divide over consequentialism, even if recognition of the divergence does not argue unambiguously in favor of any particular position. Once we see the divergent ways in which our species comes to conceptualize ethical considerations, we may be better placed to assess the different ethical positions to which reflection on these considerations can drive us.

Philosophy, Research School of Social Sciences,
Australian National University

[26] See Philip Pettit, "The Consequentialist Perspective," in Marcia Baron, Philip Pettit, and Michael Slote, *Three Methods of Ethics: A Debate* (Oxford: Blackwell, 1997). See also Philip Pettit and T. M. Scanlon, "Contractualism and Consequentialism," *Theoria* (2000): 228–45.

"BECAUSE I WANT IT"*

By Stephen Darwall

I. Introduction

How can an agent's desire or will give him reasons for acting? Not long ago, this might have seemed a silly question, since it was widely believed that all reasons for acting are based in the agent's desires.[1] The interesting question, it seemed, was not how what an agent wants could give him reasons, but how anything else could. In recent years, however, this earlier orthodoxy has increasingly appeared wrongheaded as a growing number of philosophers have come to stress the action-guiding role of reasons in deliberation from the agent's point of view.[2] What a deliberating agent has in view is rarely his own will or desires as such, even if taking something as a reason is intimately tied to desire.[3] Someone who wants to escape a burning building does not evaluate her options by considering which is likeliest to realize what she wants or wills. She is focused, rather, on her desire's object: getting out alive. The fact that a successful route would realize something she wants is apt to strike her as beside the point or, at best, as a trivial bonus.

This point is sometimes put by saying that desires are in the "background," rather than the foreground, of the practical scene a deliberating agent faces.[4] The metaphor is somewhat misleading, however, since an agent's desires are normally not so much in the background of her deliberative field as outside of it altogether.[5] If we must locate them spatially, a better place might be within or behind the standpoint from which

* An earlier draft of this paper was presented at a conference on practical reason at the University of Rome (La Sapienza). I am indebted to members of the audience, especially to Jonathan Dancy, Carla Bagnoli, Allan Gibbard, and Tito Magri. I also owe much to the other contributors to this volume. Finally, I am indebted to Nishiten Shah for very helpful discussion.

[1] Where 'desire' is understood in the broad sense, as any disposition an agent might have to bring something about. I shall generally follow this usage.

[2] For earlier criticisms of this view, see Thomas Nagel, *The Possibility of Altruism* (Oxford: Clarendon Press, 1970); E. J. Bond, *Reason and Value* (Cambridge: Cambridge University Press, 1983); and Stephen Darwall, *Impartial Reason* (Ithaca, NY: Cornell University Press, 1983). For more recent critiques, see Philip Pettit and Michael Smith, "Backgrounding Desire," *Philosophical Review* 99, no. 4 (1990): 565–92; Warren Quinn, "Putting Rationality in Its Place," in Quinn, *Morality and Action* (Cambridge: Cambridge University Press, 1993), 228–55; and T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), 41–55.

[3] On this latter point, see Scanlon, *What We Owe to Each Other*, 39–41.

[4] Pettit and Smith, "Backgrounding Desire."

[5] I take the term "deliberative field" from Barbara Herman, *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press, 1993), 193–207.