https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

SUBSTANTIVE MORAL THEORY*

By Philip Pettit

I. Introduction

Philosophy serves two functions in relation to the moral thinking that all of us practice, philosophers and nonphilosophers alike. First of all, it provides a metaethical commentary on the nature of moral thought, as the methodology or the philosophy of science provides a commentary on the nature of scientific thought. Second, it builds on the common presumptions deployed in people's moral thinking about moral issues, looking for a substantive moral theory that they might support. Metaethics receives a lot of attention among philosophers, but substantive moral theory receives less. Philosophers pursue substantive theory, to be sure, but they often fail to distinguish this pursuit from metaethics. My essay seeks to repair this neglect.

The body of the essay is divided into three sections. In Section II, I give a general characterization of metaethics and show why there is room and reason to seek out a substantive moral theory. In Section III, I offer an overview of some well-established styles of substantive theory. In the final section, I sketch out a theory that I find appealing myself.

II. From Metaethics to Substantive Moral Theory

A. The role of metaethics

True to its name, metaethics consists in higher-level reflection on the nature of moral talk and moral thought. It deals with questions of the following kinds.

- The cognitivism question. Do moral predications have truth-conditions, as cognitivists claim? Do they serve to say how things are and not just play an expressive or related role in people's psychologies? When I say "That's the right option to take," do I characterize the action in some way, or do I just give expression to the fact that I approve of it?

* I was aided in preparing this essay by comments on an earlier draft by Tristram McPherson and by conversations both with him and with Michael Smith. I am grateful for the written suggestions that I received from the editors of this volume and from Jan Narveson and Michael Huemer.

DOI: 10.1017/S0265052508080011
© 2008 Social Philosophy & Policy Foundation. Printed in the USA.

- The realism question. If moral predications are truth-conditional, do they predicate real, bona fide properties, so that their truth-conditions are routinely fulfilled? Does the rightness I ascribe to certain options, if cognitivism is correct, constitute a real property? Or is it as imaginary and unreal as phlogiston proved to be?
- The objectivism question. If the moral properties predicated are bona fide in character, do they answer to our intuitions about the character of moral properties—intuitions to the effect that they are not just properties of our subjective reactions, for example, and not properties that are relative to a variable framework, cultural or otherwise? Does the rightness that belongs to certain options, if realism is correct, consist in a property of the options in themselves, rather than in a relationship to us, and in a property that is accessible in principle from different cultural and other standpoints?

Other, more specific questions come up in metaethics too, depending on the line taken on the cognitivism issue. If we endorse cognitivism, there will be a question about how moral predicates come to designate the properties or alleged properties they assign, and if we adopt expressivism, there will be an issue about how those predicates connect up with the attitudes they express. These are questions in moral semantics, particularly in the foundational semantics that seeks to explain how certain terms, in our mouths, latch onto the items they are said to denote or assign or express.[1] And those semantic questions lead to corresponding questions in moral epistemology, bearing on what sort of knowledge or evidence supports us in our making one or another moral predication.

Metaethics, understood in this way, plays a higher-level role in relation to moral or ethical thinking, similar to the role that methodology or philosophy of science plays in relation to scientific theorizing. Metaethics offers a descriptive, insider-based picture of moral thinking, as such thinking is practiced by ordinary people, and it goes on in a more critical style to articulate, regiment, and perhaps revise that practice. The descriptive account of how people think in the moral domain provides the critical exercise with the data on which it works, and, unsurprisingly, there is a lot more convergence on the data than there is on the critical revisions proposed.

B. The data of metaethics

This convergence appears in the fact that a large number of the following assumptions are taken to be endorsed among the folk (that is, ordinary people) by different metaethical philosophies:

[1] Robert Stalnaker, "Conceptual Truth and Metaphysical Necessity," *Philosophical Studies* 18 (2004).

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

(1) *Domain:* Moral predications predicate thin values of rightness and wrongness, goodness and badness, and thick values and disvalues such as fairness and kindness, cruelty and inequity. Thin values do not have descriptive connotations, but thick values do.
(2) *Status:* Moral predications are matters on which people agree and disagree in apparently the same way that they agree and disagree on matters of belief.
(3) *Independence:* The properties predicated are not properties of our subjective reactions, and are not relative to a variable frame. They are intuitively objective.
(4) *Internalism:* To predicate rightness (wrongness) of an option is to be disposed (not) to choose it, without reliance on any independent desire to do right (avoid wrong).
(5) *Justifiability:* A right option and only a right option is justifiable. There is no ground for resentment or indignation by others, and no ground for guilt on our own part.
(6) *Grounding:* A right option normally has rightness-making properties involving thick values; a wrong option has wrongness-making properties involving thick disvalues.
(7) *Supervenience:* No difference in moral predication is defensible unless there is a nonmoral difference between the items of which the moral property is predicated.
(8) *Universal supervenience:* The nonmoral differences that support differences in moral predication do not privilege particular times, places, or persons in their particularity, only universal, replicable properties.
(9) *Accessible supervenience:* The nonmoral features that are of moral significance are accessible to reflection. People's moral competence will be put in question to the extent that they prove unable to cite the features that shape their moral judgments.
(10) *Paradigms:* There are certain independently fixed paradigms of right and wrong options on which people broadly agree.

Short of being utterly iconoclastic, a metaethical theory will start from presumptions of this kind, or at least from those presumptions that it takes to be embodied in folk moral thought. It will provide a construal of those presumptions under which they provide answers to the questions of cognitivism, realism, and objectivism. And it will provide a more or less limited endorsement of the presumptions, so construed—that is, of the views ascribed to the folk. There will be three levels, then, at which a metaethics may distinguish itself: in determining which of the candidates are genuinely folk presumptions; in construing those presumptions so that they take a particular line on the three questions distinguished; and

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

in endorsing or rejecting the presumptions so construed, whether in whole or in part.

C. Moral functionalism

My preferred metaethics assumes that all or most of the ten candidates listed above are folk presumptions and argues that moral predications are truth-conditional and that the properties they predicate are selected as candidate semantic values (at least their selection is partly determined)[2] by the fact that they satisfy such presumptions. Why will a property deserve to count as the semantic value of a given predicate—the property that the predicate assigns? It will deserve to do so because it has suitable connections with paradigms and nonmoral features, with motivation and justification and action, and with the properties that count on similar grounds as the semantic values of other moral predicates. Reference is fixed under this account by the functional profile that the referent of a moral term or concept is required to fit. For that reason, Frank Jackson and I have described this account as a moral functionalism.[3]

On this account, then, a property will count as the property of justice insofar as (1) it is exemplified in certain procedures, (2) it makes sense of why someone might choose actions that bear it, or might think they are justifiable, and (3) its apparent presence in an action counts in ordinary reasoning as a ground for thinking the action is the right choice. Justice is that property, whatever it may be in itself, that meets those usage specifications; it is individuated functionally by playing the role that they characterize.[4] And as the referent of "justice" will be determined by its place in a network of such connections, so the same will be true of the semantic values of other moral terms as well. Those terms or concepts will have their referents determined simultaneously with the referent of "justice," and the mastery of any one term—the identification of its referent—will go hand in hand with the mastery of others. Semantic competence will be achieved in a package deal.

[2] Philip Pettit, "Descriptivism, Rigidified and Anchored," *Philosophical Studies* 18 (2004): 323–38.

[3] See Frank Jackson and Philip Pettit, "Moral Functionalism and Moral Motivation," *Philosophical Quarterly* 45 (1995): 20–40; reprinted in Frank Jackson, Philip Pettit, and Michael Smith, eds., *Mind, Morality, and Explanation* (Oxford: Oxford University Press, 2004). For further background, see Frank Jackson, "Critical Notice of Susan Hurley's *Natural Reasons*," *Australasian Journal of Philosophy* 70 (1992): 475–87; Frank Jackson, *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford: Oxford University Press, 1998); and Philip Pettit, "Embracing Objectivity in Ethics," in *Objectivity in Law and Morals,* ed. Brian Leiter (Cambridge: Cambridge University Press, 2001), 234–86.

[4] Or, to gesture at a variant, it is the higher-order property of there being some property present that meets those specifications; it may be the role property, as it is sometimes called, rather than the realizer property that plays the role. On this distinction, see Frank Jackson and Philip Pettit, "Functionalism and Broad Content," *Mind* 97 (1988): 381–400; reprinted in Jackson, Pettit, and Smith, eds., *Mind, Morality, and Explanation,* 95–118.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

Moral functionalism, so characterized, endorses cognitivism and suggests a distinctive semantics for moral predicates. Moral predications are truth-conditional, under this account, and the predicates they deploy pick out certain properties as their semantic values insofar as those properties satisfy the usage specifications given by the features listed, or some subset. Depending on exactly which subset is valorized, of course, the moral functionalism embraced may take one or another specific form; in that sense, it directs us to a family of metaethical views rather than any particular position. And as moral functionalism gives us a moral semantics, so too it will provide us with a moral epistemology. According to the implied epistemology, we will know when to make one predication or another, depending on the usage specifications that appear to be fulfilled.

Nonetheless, moral functionalism does not yet commit us on the issues of realism and objectivism. According to the account that I find appealing, there will be suitable, naturalistic properties available, however disjunctive in character, that fit the required roles. And in that case moral properties will be real and objective features of the world. Moral functionalism might equally be squared, however, with a nonnaturalistic story about the realizers of the roles. And it might even be maintained consistently with an error theory about the realizers: a theory according to which moral thought is deeply misconceived, there being no properties available to fit the roles that it projects for them.[5]

D. The folk's schematic moral theory

The idea in moral functionalism is that we master the different moral predicates by coming to understand, at one and the same stroke, what is required for each of them to have application—what is required for the property a predicate ascribes to be present. To achieve a mastery or understanding of moral language, then, will be to have recognized how those properties that answer to the moral terms relate to one another and how they relate en bloc to evidence and action. And to take moral language seriously, treating it as a way of representing how things are, will be to have endorsed a theory—a formal or schematic theory—as to the existence of properties that satisfy such a network of connections. According to moral functionalism, there is no engaged moral talk or thought—no serious, literal usage of moral terms and concepts—without a commitment to this modicum of moral theory.[6]

[5] For an example of an error theory, see J. L. Mackie, *Ethics* (Harmondsworth: Penguin, 1977).

[6] The approach represents an analytical functionalism in the moral area that parallels the analytical functionalism that is often defended in the area of mind. It identifies different moral properties, at least in part, by the ways in which they are connected up in the practice, and in the implicit thought, of participants. Notice that the connections signaled involve the way people treat properties and the sorts of practices they follow. The connections are different, but not so different, from the causal connections that figure in functionalism within the philosophy of mind. In particular, they are not themselves normative connections.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

Is moral functionalism unusual or even unique in postulating a folk moral theory of this schematic kind? I do not think so. There are two salient ways in which a metaethics might break with the family of moral-functionalist doctrines. One is by holding that moral predications are not truth-conditional but expressive; and the other is by holding that they are truth-conditional but that the reference of moral terms or concepts is fixed on a purely direct basis, not by satisfaction—not even in part by satisfaction—of suitable usage-specifications. Under either of these variations, however, it appears that the metaethics in question will still have to countenance the existence of a folk moral theory.

Under the first variation, where the truth-conditional character of moral predication is denied, the folk will be said to regard moral predications as purely expressive and to see moral agreement and disagreement as different in character from agreement or disagreement in matters of belief. But this expressivist story does not relieve the folk of having to accept some or all of the other presumptions listed above (in subsection II.B). Presumably, it will remain the case that in order for people to become competent in the non-assertoric practice that moral thinking allegedly involves, they will still have to take guidance from such presumptions. They will have to recognize the role of thick in relation to thin values; they will have to see the motivational connection and the justificatory role of moral predication; and they will have to employ such predications with sensitivity to requirements like supervenience and universalizability. If they commit themselves seriously to moral talk and thought, then they will still have to understand and adjust to this set of presumptions. They will have to endorse the theory that these presumptions constitute but now in a different role from the role the theory plays under a moral-functionalist approach.

Under the second variation from a broadly moral-functionalist approach, moral terms and concepts can gain their semantic values on the basis of exposure to suitable candidates, without a requirement that those candidates should satisfy any presumptions that the folk make about them. Suppose the folk use a moral term "t." Assuming that error theory is false, the idea is that that term will pick up a property in the world as its referent on the basis of something about the property that may in no way register with users of the term. What selects it as the referent of the term may be the fact that the property, unbeknownst to speakers, is at the causal origin of a use of the term that they take as authoritative, or the fact that it uniquely co-varies with the general use of the term, or the fact that among the properties that co-vary with that use, it is a uniquely eligible referent, on some independent criterion of eligibility. The folk's presumptions about the property that they assign with the term "t" may be quite false, then; the reference is fixed by factors that may escape the notice of speakers, and thus the beliefs that the speakers form about the referent properties may be quite misconceived.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

This direct theory of reference for moral terms and concepts is hardly a plausible theory for all the moral predicates there are. But if it is a theory only of some predicates, then it will leave moral functionalism, or something like moral functionalism, in place for the remaining predicates. Thus, it will leave room for a version of the schematic moral theory that moral functionalists ascribe to the folk.

Nonetheless, even if the theory is said to apply to all moral predicates, it will still have to keep a role for such a folk theory. If the referents of moral terms are fixed in a manner that puts them potentially beyond our ken, leaving us prey to endorsing nothing but falsehoods about them, then that opens the way for a radical, incredible scepticism. We will each be held to engage unknown properties in our assertions about different options or policies or persons and to ascribe those properties without any reliable basis—indeed, without any basis at all—for determining whether they actually apply. Making moral predications will be like throwing darts in the dark. In order to guard against this outlandish scenario, those who opt for the direct theory of reference for moral terms will have to tell a story according to which there are reliable guidelines and grounds available for the correct or presumptively correct application of such terms. But where are those guidelines to come from, if not from a body of judgments and presuppositions that are ascribed to the folk and treated as fallible but reliable? And where are we to find such folk commitments if not in the sorts of presumptions listed above? This approach will also have to countenance a folk moral theory, though a theory that serves primarily in an epistemological rather than a semantic role.

E. From schematic to substantive moral theory

If we hold a schematic moral theory, and if that theory guides us in an analysis of moral issues, then reflection on where it guides us—reflection on the pattern we are disposed to follow in our judgments—may enable us to articulate what I call a substantive moral theory. I will argue this line, on the assumption that moral functionalism holds, and that it directs us to genuine properties, not illusions. This is for reasons of convenience only, however, since there will be a parallel line to argue under any metaethics. The folk will be ascribed a schematic moral theory under every approach, as we have seen, and in each case there will be a base for seeking out a substantive theory to which it points.

Suppose, then, that presumptions like those given in the earlier list serve for me as semantic markers or indicators of what is right or wrong, good or bad, just or unjust, and so on. And suppose that working with these presumptively common guidelines, I argue for the rightness or wrongness, the goodness or badness, the justice or injustice, of one or

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

another action or policy. I will do so under the assumption that while there are trade-offs to be made and fine lines to be drawn in the course of the exercise—it will involve the search for a reflective equilibrium between the presumptions I maintain and my intuitions about the cases adjudicated[7]—still there is a correct and an incorrect way of extrapolating to other cases, and a way of extrapolating on which I may expect a degree of convergence with others.[8]

Assuming that the extrapolative, judgmental dispositions that the schematic presumptions underpin will guide me fairly determinately, there is an immediate prospect that I can work my way toward a substantive moral theory. I will be able to explore this possibility, it appears, using a familiar philosophical method of analysis: the method of cases. Equipped with the judgmental dispositions, I can envisage any number of different moral issues and I can investigate where the schematic presumptions would lead me across those possible, imagined cases. With those judgments of possible cases more or less secure, I can then look at the moral configuration they assume, in order to see whether there is any substantive commonality in the extension of rightness and wrongness, or goodness and badness, or justice and injustice, and so on. The challenge will be to find a commonality that is more specific and more salient than the pattern given by the schematic presumptions, as I interpret and apply them.

According to any moral-functionalist metaethics, or at least one that rejects wholesale error, the schematic presumptions associated with moral terms or concepts will serve like a net with which I can trawl the different cases and issues, real and imagined, that I consider. That net will sift out some options as instances of the right actions or policies to adopt in given cases, for example, and leave others as instances of actions or policies that are wrong. The idea in pursuing a substantive moral theory is to see whether there is a salient pattern to be found in such instances of rightness or wrongness over and beyond the pattern that consists in the fact that they are caught or not caught by the net. That they are or are not caught by the net is presumably due to their inherent character, and substantive moral theory is built on the hope that this character can be

[7] On reflective equilibrium, see John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971).

[8] If I take myself to be guided by common markers of value and disvalue, after all, then I will have to assume that that guidance is not peculiar to me. The common meanings established for terms like "just" and "good" and "right" hold out the prospect of convergence in judgments of value, and so long as that possibility is live, I will find a point in talking with others, exchanging comments on what differences and continuities register with us, how different presumptions are weighted in relation to one another, how they are amended in the course of reflective equilibration, and so on. I may come to believe, of course, that the meanings available leave issues of value indeterminate, and I may come to treat those issues, like various issues of baldness and color and the like, as vague at certain margins. But I will presumably retreat to this sort of quietism only as a last resort; it amounts to throwing in the conversational towel.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

identified, using something like the method of reflection on cases, by those who make moral judgments.[9]

There may not prove to be any interesting characterization available at the level of substantive theory, of course, but that is not a reason for despairing at the outset of the enterprise. At least, it is not a reason for despairing, under the assumption that cases can be morally adjudicated on the basis of how they are described, and that adjudication does not require causal exposure to the case assessed, or even to a vivid simulacrum on stage, in a film, or in a novel. This assumption, which is implied in the claim that I described as accessible supervenience, is validated in the way we expect someone to be able to explain a disparity in the moral judgment made over two cases by pointing to a descriptive difference between the cases—in particular, by pointing to it as a factor that justifies the disparity, not merely as a conjectural source of the disparity. Suppose some individuals report a moral judgment about one or another issue: a considered judgment, not just a dogged but not necessarily endorsed intuition. If they are unable to justify the judgment by reference to some descriptive characterization of the case, then we will doubt their skill at making moral judgments. And if they are unwilling to acknowledge the need for such descriptive justification of the judgment, then we will doubt their very understanding of what moral judgment involves.[10]

[9] For an argument that there must be an inherent character there to explain the semantic competence of moral speakers, in particular their ability to catch on to the reference of a term like "right," see Frank Jackson, Philip Pettit, and Michael Smith, "Ethical Particularism and Patterns," in *Moral Particularism,* ed. Brad Hooker and Margaret Little (Oxford: Clarendon Press, 2000). This argument challenges the radical version of particularism, according to which there need be no independent pattern of any kind in the extension of a term like "right." A more moderate form of particularism would defend the aesthetic model of moral judgment mentioned below in note 10; this model holds that there need be no pattern in the extension of "right" that is accessible, at least in principle, to reflection on the part of competent speakers. See Jonathan Dancy, *Ethics without Principles* (Oxford: Oxford University Press, 2004).

[10] This assumption is rejected under what may be described as the aesthetic model of moral judgment. When I make an aesthetic judgment on a painting, I must do so on the grounds of how it impacts perceptually and causally on me. It would be absurd to have a painting described and then, on the basis of that description alone, to pass aesthetic judgment—to say, for example, that the painting must be very beautiful, or elegant, or whatever. See Philip Pettit, "The Possibility of Aesthetic Realism," in *Pleasure, Preference, and Value: Studies in Philosophical Aesthetics,* ed. Eva Schaper (Cambridge: Cambridge University Press, 1983), 17–38. The aesthetic model of moral judgment would suggest that I can only form a moral judgment about an action on the basis of how it impacts experientially on me. There may be a substantive pattern to identify across the aesthetic judgments we defend (or across moral judgments, if they are conceived as similar to aesthetic ones), but it will be there to be identified only on the basis of empirical research; there will be no possibility of relying on reflection in order to bring it to light. Those who passed a positive judgment on the drip paintings of Jackson Pollock, and a negative judgment on various imitations, were in no position to identify the substantive pattern that physicists have since claimed to unearth in Pollock's paintings—a pattern of self-similarity at different grains of analysis. See Richard P. Taylor, "Order in Pollock's Chaos," *Scientific American* (December 2002): 117–21.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

F. Looking for substantive moral theories

With the case made for why substantive moral theory, like metaethics, is a quarry that philosophers may pursue, we are in a position to look for some examples of such theorizing in the tradition, and indeed among contemporary figures. I pursue that task in the next section.

The theory I ascribe to some thinkers may be conceived by them as an exercise in metaethics, not as a substantive moral theory. Consider utilitarianism, Kantianism, contractualism, or virtue ethics. These are sometimes taken as metaethical theses to the effect that insofar as they bear on rightness, the schematic presumptions in our earlier list, whether or not the folk realize it, can really be boiled down to a single, favored thesis. The claims made under such approaches would be that the concept of the right is just the concept of that which maximizes happiness, or that which can be coherently willed as a universal law, or that which is immune to reasonable complaint, or that which the virtuous would choose.

It is not really plausible, however, to cast the claims in this way. For in supporting the claims, defenders generally appeal to background assumptions about the rightness property, thereby displaying an adherence to a distinct schematic theory. They try to show that if we identify instances of the right on the basis of such schematic assumptions, then we will find that the favored account applies across those judgments; the account identifies a substantive pattern in the extension of "right" and "wrong," as those words are used under the guidance of the schematic assumptions.

One reason for the failure to recognize various theories as substantive moral theories, not as contributions to metaethics, may derive from an uncertainty about the point of such theory. So what purpose is a substantive theory—say, a substantive theory of the right—likely to serve?

Such a theory may be comprehensive or restricted; it may claim to identify a common character in all instances of rightness or just a character common to instances of rightness in one or another area. In either case, its use will be twofold. In relation to moral analysis, it will provide a basis on which judgments about particular cases can be facilitated or supported. And in relation to schematic theory, it will provide some vindication of the presumption that the marks of rightness identify a more or less clearly patterned and learnable extension—an extension such that it is possible for those who are exposed to some proper subset of instances to extrapolate to further cases.

III. Five Substantive Moral Theories

There are at least five substantive theories of morality in the grand tradition. They focus on some central demands that morality intuitively imposes—some allegedly salient patterns in the distribution of the rightness property—where these demands may include both strict imperatives

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

and looser ideals. The theories identify the central constraints of morality with requirements, respectively, of nature, self-interest, benevolence, reason, and justifiability.

A. Nature

The identification of moral requirements with requirements of nature takes a number of different forms and is common in classical and medieval thought. The Aristotelian equation of the demands of morality with what is required by nature if we human beings are to flourish and enjoy eudaimonia offers one version of the idea; this approach retains a place among neo-Aristotelians like Philippa Foot and Rosalind Hursthouse.[11] Another version is the Stoic conception of laws of human nature that are available to our reason from the intimations of instinct, laws such that by being faithful to them we bring ourselves into line with the order of the cosmos. The medieval synthesis that was achieved by Thomas Aquinas built on both of these foundations, identifying the prospect of human fulfillment with obedience to laws whereby the natural instincts for self-preservation, procreation of the species, life in society, and knowledge of god could be reconciled and satisfied.

The metaphysics of naturally based requirements gave way, toward the end of the medieval period, to a metaphysics of natural rights, under pressures that are still much debated.[12] Natural rights are supposed to place demands, not on their holders, but on those who deal with their holders. Absent an agreement to be treated otherwise, such rights are meant to require that in dealing with one another we treat certain areas of choice, usually related to life, liberty, and property, as sacred; we allow each other more or less total discretion within those boundaries.

The moral metaphysics associated with natural flourishing, natural law, and natural rights is not likely to appeal on a broad front in contemporary discussion. There is no consensus about what human flourishing involves, or what natural instinct intimates, so that neither base identifies a determinate set of requirements. And while many people talk the language of natural rights, this metaphysics also fails to deliver the sorts of posits that are likely to satisfy us. If talk of rights is reduced to other terms, as it frequently is, then it does not give us distinctive requirements; it becomes a mere *facon de parler.* And if it is not reduced, so that natural rights are taken to make primitive, unanalyzable claims, then they cease to be obviously plausible entities to posit; they begin to look meta-

[11] See Philippa Foot, *Natural Goodness* (Oxford: Oxford University Press, 2001); and Rosalind Hursthouse, *On Virtue Ethics* (Oxford: Oxford University Press, 1999).

[12] See Annabel Brett, *Liberty, Right, and Nature* (Cambridge: Cambridge University Press, 1998); and Richard Tuck, *Natural Rights Theories: Their Origin and Development* (Cambridge: Cambridge University Press, 1979).

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

physically dubious: "simple nonsense," as Jeremy Bentham thought, even "nonsense upon stilts."[13]

B. Self-interest

In the history of ethical and political thought, the moral metaphysics that supplanted that of nature was one of self-interest. Sixteenth-century thinkers such as Hugo Grotius and Thomas Hobbes continued to use the language of natural law, but, especially in the work of Hobbes, natural laws came to be more or less synonymous with maxims of self-interest, in particular, self-preservation—in Hobbes's words, "conclusions or theorems" that people could work out by reason "concerning what conduceth to the conservation and defence of themselves."[14]

There is more determinate content to the idea of what is required for our self-interest and self-preservation than there is to the idea of what is required by our nature. But how could such requirements ever hope to count as moral requirements? In particular, how could they justify what is done in their name, especially when this may impose a risk of disadvantage on others?

The answer that Hobbes gave, and the answer given by contemporary Hobbesians like David Gauthier, is that if the requirements of self-interest are adjusted for life with others, they begin to approximate intuitively moral demands.[15] The idea is that self-interest will give each individual a reason to help establish certain principles whereby the common life of members of the community can be organized to maximum mutual advantage and, even more important, a reason to abide by the principles accepted.

But still, there is a problem. Won't people want to exploit others by free-riding on their efforts, whenever this is in their self-interest? Indeed, won't this be required when such exploitation maximizes personal advantage by giving the agent enough benefit to outweigh the damage that others suffer? It seems quite counterintuitive to identify moral requirements with imperatives that allow, even mandate, such exploitation. And that problem is exacerbated by the fact that the pursuit or implementation of such prescriptions will often be self-defeating for the collectivity; if each individual acts on these prescriptions, then overall self-interest will not be well served.[16]

Hobbes's answer to this problem, which is scarcely persuasive, is that free-riding will not be in people's self-interest, given the likelihood of

[13] Jeremy Bentham, "Anarchical Fallacies," in *The Works of Jeremy Bentham,* ed. J. Bowring (Edinburgh, 1843).

[14] Thomas Hobbes, *Leviathan,* ed. E. Curley (Indianapolis, IN: Hackett, 1994), 15.41.

[15] David Gauthier, *Morals by Agreement* (Oxford: Oxford University Press, 1986). On the Hobbesian approach more generally, see J. S. Kraus, *The Limits of Hobbesian Contractarianism* (Cambridge: Cambridge University Press, 1993).

[16] Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), part 1.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

detection. Someone "that breaketh his covenant," being readily detectable, "declareth that he may with reason do so," and "can in reason expect no other means of safety than what can be had from his own single power."[17] Gauthier's answer, which is only a little more satisfying, is that the disposition to abide with an agreement made will generally be in people's self-interest. While this may be correct, it doesn't really address the problem: people are presumably able to act against their dispositions, especially when self-interest requires it.[18]

C. Benevolence

Despairing of the resources of self-interest, or at least of self-interest taken on its own, the eighteenth-century British philosopher Francis Hutcheson introduced the idea that moral requirements are just the requirements of benevolence.[19] Hutcheson argued that people like others to be happy and do not like them to be unhappy, and that this benevolence leads people to perform and approve of suitable actions, and to eschew and disapprove of others. Where Hutcheson took this to be an innate desire, David Hume argued that it could be explained by reference to the sympathy we feel when others suffer what we think would be painful, or enjoy what we think would be pleasant.[20] Exercised within the general point of view in which we put aside our own particular concerns, sympathy, according to Hume, makes sense of the benevolence that Hutcheson postulated.[21]

What we are offered here is essentially a utilitarian account of the requirements of morality, as indeed Hutcheson already advertised in 1725: "that action is best, which procures the greatest happiness for the greatest

[17] Hobbes, *Leviathan,* 15.5. See K. Hoekstra, "Hobbes and the Foole," *Political Theory* 25 (1997): 620–54.

[18] Gauthier argues, in later work, that rationality can oblige people to take a temporally extended course of action (say, promising at one time and then actually delivering on the promise later) even where promising-and-not-delivering has higher expected utility than promising-and-delivering. See David Gauthier, "Resolute Choice and Rational Deliberation," *Noûs* 31 (1997): 1–25. See also Edward McClennen, *Rationality and Dynamic Choice* (Cambridge: Cambridge University Press, 1990). But this is an unorthodox and highly controversial view, since it introduces an artificial way of limiting the options that agents can rationally consider; it rules out the option of reconsidering whether to deliver on the promise, even when that possibility has become salient. It may, of course, be rational to have a policy of not in general reconsidering one's options—a policy of not reconsidering except when the red lights go on. See Michael Bratman, *Faces of Intention: Selected Essays on Intention and Agency* (Cambridge: Cambridge University Press, 1999). But this does not mean that rationality rules out reconsideration, period.

[19] Gilbert Harman, *Explaining Value and Other Essays in Moral Philosophy* (Oxford: Oxford University Press, 2000), chap. 11.

[20] David Hume, *An Inquiry Concerning the Principles of Morals*, ed. J. B. Schneewind (Indianapolis, IN: Hackett, 1983); David Hume, *A Treatise of Human Nature,* 2d ed., ed. P. H. Nidditch (Oxford: Oxford University Press, 1978).

[21] 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.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

numbers; and that, worst, which, in like manner, occasions misery."[22] The account was later given full formal dress in the work of Jeremy Bentham and other utilitarians. It identifies the requirements of morality, quite simply, with the requirements associated with maximizing the expectation of happiness, whether among human beings or among sentient creatures more generally. It retains a prominent place in contemporary moral philosophy, figuring in the work of Peter Singer and many others.[23]

The value of benevolence is only one among a number of different values—thick values—in our folk moral theory, and it is counterintuitive to construe all moral requirements as dictates of benevolence. Equally, since benevolence is so demanding, it is counterintuitive to take everything that benevolence might dictate as a requirement of morality. Utilitarianism would not substantiate our intuitive conception of moral requirements so much as replace it with a novel conception; utilitarians, as they were seen to be in the nineteenth century, are philosophical radicals.

D. Reason

Reacting against the focus on sentiment and sympathy in Hume's work, Immanuel Kant argued that moral demands were not requirements of benevolence but rather requirements of reason, under his particular conception of reason. He gave rise thereby to one of the major strands in contemporary moral theory. This strand is faithfully maintained in the work of many contemporary Kantians, such as Barbara Herman and Christine Korsgaard, but it is also represented in non-Kantian form by the work of other, very different thinkers, such as R. M. Hare.[24]

The Kantian claim is that the very idea of acting on reason imposes requirements of a kind that can plausibly be identified with the requirements of morality.[25] There are many variations on the argument for this idea, and many versions of each of those variations, but one central line of thought is this: If I am in causal charge of what I choose, as I naturally assume in any choice, then I must choose in accordance with a universal law. Causation, Kant assumes, is essentially lawlike. It follows that if I fail to choose in accordance with a universal law, then it is not I who is in causal charge of what is done but an alien, heteronomous force within me. Choosing in accordance with a universal law would mean choosing in accordance with a maxim that can be coherently imposed as a universal

[22] Quoted in Harman, *Explaining Value,* 186.

[23] Peter Singer, *Practical Ethics,* 2d ed. (Cambridge: Cambridge University Press, 1993).

[24] R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford: Oxford University Press, 1981); Barbara Herman, *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press, 1993); Christine Korsgaard, *The Sources of Normativity* (New York: Cambridge University Press, 1996).

[25] Immanuel Kant, *Groundwork of the Metaphysics of Morals* (1785), trans. Mary Gregor (Cambridge: Cambridge University Press, 1998).

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

law. Thus, if I am to vindicate the natural assumption that I am in causal charge of what I choose, then I must choose in accordance with a maxim that I can will as a universal law. This is a constraint that binds any agent of reason unconditionally; it is a categorical imperative, as Kant calls it.

Kant's view is that the requirements of morality just are the requirements imposed by the categorical imperative of choosing in accordance with coherently universalizable maxims. Let the maxim by which I act be universalizable and I will be acting morally; let it be nonuniversalizable and I will be acting immorally. Imagine that I steal someone else's property, acting on the maxim of a thief. The Kantian idea is that I cannot coherently will this maxim of stealing what belongs to another as a universal law because if everyone acted on that maxim, then there would be no such thing as property and no such thing as stealing.

The Kantian proposal is hard to defend, at least in this version, since the issue of whether I can will a maxim as a universal law turns on how that maxim is described. I may not be able to endorse the maxim of stealing as a universal law, but I can endorse many more-specific versions of that maxim as universal laws. For example, there is no reason why I should not be able to endorse, as a universal law, the maxim of stealing another's property in very specifically characterized circumstances, such as those I happen currently to be in. I might endorse as a universal law the maxim of stealing a book from someone who no longer needs it urgently, on the eve of an examination that requires me to have read that book.

R. M. Hare's "prescriptivism" is an attempt to build on roughly this Kantian idea, avoiding the sort of problem I have just raised.[26] Saying that an option is right, on Hare's approach, is just prescribing it universally, so that an action will be right just in case it is universally prescribable. But it is not enough for universalizability in this sense that universally prescribing the option is a coherent strategy. What is also required is that I be able to think that this is the option I would prescribe for me now, no matter what position I occupied and no matter what impact I suffered. Universally prescribing the option must be a psychologically feasible strategy, as we might say, not just a coherent one. That is why Hare's test of universalizability is supposed to avoid the indeterminacy problem.

Hare assumes that in prescribing an option for myself, without universalizing, I will always take into account the desires that I expect the option to satisfy: I will prescribe the option on the grounds that it promises satisfaction of those desires. He argues that when I prescribe the option universally—that is, not just for me now, but for anyone in a relevantly similar situation—I will prescribe it on the grounds that it best provides for the overall satisfaction of the desires of affected parties. Thus, on this approach, there is an even more specific commonality to right options over and beyond that of being universally prescribable. This is that the

[26] See Hare, *Moral Thinking*.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

options maximize expected desire-satisfaction—that is, that they have the character prized by the preference-based form of utilitarianism. Hare offers substantive theories at two levels; at one level, universal prescribability is the common feature of right options; at a second level, utilitarian desire-satisfaction plays that role.

Hare's approach suffers, for this reason, from all the problems associated with utilitarianism, but it is also subject to a more damaging objection.[27] The argument for the determinacy of universal prescribability, and thus for Hare's utilitarianism, depends crucially on the assumption that in prescribing an option for myself, without universalization, I will always do so on the grounds that it promises maximal desire-satisfaction. And that assumption is very questionable.

It may well be true that whenever I prescribe an option for myself, I do so out of a corresponding desire, but that does not mean that I prescribe and choose it on the grounds that it will satisfy such a desire. It does not follow, then, that in prescribing something universally I will prescribe it on the grounds that it does as well as possible by desire-satisfaction overall. Socrates prescribed drinking the hemlock for himself on the grounds that this is what respect for the law required, not on the grounds that doing so would satisfy his desires, in particular his desire to show respect for the law. Thus, he would have prescribed the choice of the hemlock for anyone in any situation where those grounds remained valid. And he would have prescribed it in that universal way, regardless of whether or not it happened to increase the net balance of desire-satisfaction.

E. Justifiability

A fifth and distinct conception of the requirements of morality was provided by Adam Smith, a friend and admirer of Hume.[28] Taking up the Humean idea of sympathy, he developed it on different lines.[29] We sympathize with the actions and reactions of others, Smith said, but only insofar as we would display more or less similar responses in their circumstances; if others underreact or overreact, by the criteria we employ in an impartial viewpoint, then we disapprove. Given the notion of impartial sympathy, Smith tells a distinctive and persuasive story about how we come to identify certain requirements for ourselves; these, by his

[27] For a critique, see Philip Pettit, "Universalizability without Utilitarianism," *Mind* 96 (1987): 74–82; and Philip Pettit and Michael Smith, "Backgrounding Desire," *Philosophical Review* 99 (1990): 565–92; reprinted in Jackson, Pettit, and Smith, eds., *Mind, Morality, and Explanation*, 131–53. See also Philip Pettit, "Preference, Deliberation, and Satisfaction," in *Preferences and Well-Being*, ed. Serena Olsaretti (Cambridge: Cambridge University Press, 2006), 131–53.

[28] Adam Smith, *The Theory of Moral Sentiments* (1759), ed. D. D. Raphael and A. L. McFie (Indianapolis, IN: Liberty Classics, 1982).

[29] Here I follow Harman, *Explaining Value*, chap. 11.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

account, are the requirements of morality. The elements in the story are as follows:

(1) We sympathize with others just so far as their actions and reactions seem fitting to their circumstances, being responses we can imagine ourselves having;
(2) we learn that others sympathize with us in turn, forming positive or negative appraisals, in the same impartial and critical manner;
(3) we learn to put ourselves in the shoes of others, and to see our own actions and reactions as they would see them from their impartial, critical point of view;
(4) we are naturally motivated to seek the approval of others, and to avoid their disapproval; and so
(5) we come to be motivated to act and react in a manner deserving of more or less impartial sympathy.

According to this complex story, then, the requirements of morality should be identified with what is required, not for the sympathy that the ignorant or biased may bestow, but for the impartial sympathy of others. The requirements of morality are those requirements that we have to satisfy if, on the negative side, we are to avoid the reasonably based complaints of others and if, on the positive side, we are to win their reasonably based plaudits. Where Hutcheson and Hume had equated the requirements of morality with the requirements of benevolence, then, Smith was led from a similar starting point to take a rather different view. He identifies the requirements of morality with the conditions whose realization would put us beyond the complaint of others, and within the ambit of their approval. We might describe these, in a generic phrase, as requirements of justifiability.

The requirements to which Smith directs us have great appeal. They are not obscure, like natural law and natural rights. They are not counter-intuitively egocentric like the requirements of self-interest, nor counter-intuitively altruistic like the requirements of benevolence. And they do not suffer, on the face of it, from the indeterminacy problem associated with Kant. The justifiability feature is a very promising way of unifying at least a good portion of those options that count intuitively—that is, count by our presupposed, schematic theory—as right.

Two contemporary theories build on broadly the same base as Adam Smith's, at least as I see them. One is the contractualist approach of T. M. Scanlon.[30] Right actions, for Scanlon, are those that are not wrong; and wrong actions are those against which others can raise complaint on the basis of principles that no one could reasonably reject as principles for the

[30] T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998).

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

general regulation of social behavior. While Scanlon does not cite Smith as a predecessor, the approach represents something close to a negative version of the justifiability line. The requirements of morality are identified as those whose fulfillment makes for immunity to reasonable complaints, though no mention is made of requirements whose fulfillment would serve to attract corresponding plaudits. Scanlon differs from Smith in giving us a quasi-contractual rather than a genealogical story as to why certain complaints are reasonable: they are backed by principles that no one could reasonably reject as principles of social cooperation. But Scanlon's approach and Smith's still clearly belong in the same family.

A second contemporary theory that shares a common base with Adam Smith's is that of Michael Smith.[31] Where the linking idea between Adam Smith and Scanlon is that of immunity to complaint, the linking idea here is that of the impartial spectator. Adam Smith's impartial spectator is the internalized other, as we have seen; Michael Smith's is the idealized self. The demands of morality, on this approach, belong among those requirements that I would advise my actual self to satisfy in any situation, were I, the advising self, to be beyond criticism—were I to be fully informed, for example, to have fully coherent and unified attitudes, and to be thinking in a fully logical way. These are requirements, therefore, such that satisfying them ensures that I am justified; I pass muster with the impartial spectator of my ideal advising self.[32]

There is one salient problem that arises with Adam Smith's theory, and it is inherited by the contemporary variants I have mentioned. The problem is that we are given no reason to think that the demands of justifiability will have the stability that we naturally expect in any requirements that count as moral. Nothing that Adam Smith says provides a ground for expecting stability across cultures, for example, in the standards to which impartial sympathy would hold us. Such an expectation would have to be premised on an empirical claim that the psychological mechanisms identified operate in the same way, and to at least a related effect, across different contexts.

Scanlon is also exposed to a problem of stability. For what sorts of considerations count as good reasons for rejecting a principle or making a complaint? Scanlon emphasizes that they must be "personal" or agent-relative reasons that vary from individual to individual.[33] For all that he shows, however, the sorts of personal reasons that count as good reasons

[31] Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994).

[32] These are requirements of reason on two counts, and in that respect connect also with the Kantian approach. First, they are requirements associated with the perfect, informed operation of reason within the advising self. Second, they are requirements that the advisee self, as a creature of reason, will want to satisfy. As between the self that does satisfy them and the self that doesn't, there is no question about which is the more rational.

[33] See Scanlon, *What We Owe to Each Other;* and Michael Ridge, "Saving Scanlon," *Journal of Political Philosophy* 9 (2001): 472–81.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

for rejecting a principle or making a complaint may vary from culture to culture.

Michael Smith's proposal also faces a problem of stability, this time across individuals as well as across cultures. Why should people be expected to converge on the requirements that their idealized selves would impose, especially if they start from very different standpoints? Why shouldn't it be the case that my idealized self would recommend that I take lines of action that your idealized self would reject? Why shouldn't our idealized selves represent impartial spectators who work with different terms of appraisal?

In response to this problem, Michael Smith makes two points, one bearing on the possibility of convergence in what our idealized selves recommend, the other bearing on the necessity of such convergence. The possibility claim is that although our actual selves may have quite different tastes and inclinations on certain fronts, our idealized selves might still agree that any person with the one set of tastes and inclinations should do such and such and any person with the other should do so and so. The necessity claim, which is much more important for the problem raised, is that anything that an idealized self would recommend for a person of a particular mentality, in particular circumstances, has to be something that any idealized self, no matter whose, would recommend for that person; otherwise, it wouldn't pass muster as a rationally compelling recommendation.

These claims, however, are far from conclusive. Someone with a particular goal or allegiance—say, an allegiance to a particular religion—may just not be ready to regard that feature as a contingent circumstance of positioning such that there is a good, position-free question about how a person, so positioned, should act. And, taking this view, such an agent would naturally refuse to treat as idealized any counterpart who, not recognizing the true faith or religion, lacked that allegiance. Asked to defend this line, the interlocutor might argue that an idealized counterpart who failed to recognize that particular faith or religion would just not count as fully informed.

IV. A New Departure

In this final section, I sketch a sixth substantive theory that equates the central requirements of morality with the preconditions for the practice, surely more or less inescapable among human beings, of relating to one another in a co-reasoning, conversational mode. This approach breaks from those considered in Section III insofar as the obligations that it unifies on this basis are artificial in the sense in which Hume thought that obligations of justice were artificial. The approach presupposes a disposition on people's part to generate a certain institution or practice and

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

identifies the relevant requirements by reference to that institution or practice. The best way to present the approach in the brief compass available is to reduce it to a number of propositions, without expanding in elaboration of these.[34] My aim is not to give a full-dress defense of the theory but rather to show that it is available, and how it fits in relation to the other forms of substantive moral theory that I have surveyed.

A. Human beings are not just rational animals; they are reasoning creatures as well

To be rational is to be a locus of evidence-responsive beliefs and desires, and to be disposed to act in the way that is required according to those beliefs, for the satisfaction of those desires. While various other animals may be rational without ever reasoning, we human beings are not among them: not only are we rational, or at least close to rational, we also reason. Reasoning involves intentionally asking oneself questions such that by answering them, one may hope to guard against certain failures of rationality in the evolution of one's beliefs and desires; it consists in the sort of activity we naturally ascribe to Rodin's figure of *Le Penseur* but never to the dog or the cat. The questions typically considered bear on how far a certain body of evidence supports a proposition, whether two propositions are consistent, and whether making a certain proposition true will be a means of realizing another. By asking oneself questions of this kind about propositions, and forming meta-propositional beliefs in response (beliefs about the relationships and properties of relevant propositions), one may hope to put extra checks on the ordinary process in which intentional attitudes materialize and mutate. If one prompts oneself to recognize that "p" and "q" are inconsistent, for example, that ought to correct or inhibit the belief that it is the case, or indeed the intention to make it the case, that p and q.

B. Human beings co-reason with one another, as well as reasoning with themselves

Having access to language, human beings are able, uniquely among animals, to reason with one another as well as with themselves. They call on one another as advisers or informants or critics, and they try to guard against any irrationality in what they accept in common or attempt to achieve together. People who converse and co-reason in this way do for

[34] For the central, motivating idea, see Philip Pettit and Michael Smith, "The Truth in Deontology," in *Reason and Value: Themes from the Moral Philosophy of Joseph Raz,* ed. R. J. Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith (Oxford: Oxford University Press, 2004). And for a fuller account, see Philip Pettit, "Joining the Dots," in *Common Minds: Themes from the Philosophy of Philip Pettit,* ed. Michael Smith, H. G. Brennan, R. E. Goodin, and F. C. Jackson (Oxford: Oxford University Press, 2007).

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

one another what they try to do for themselves in personal reasoning. They lend one another their reason, as we might say, each serving in the forum of the other's mind as an *amicus curiae,* a friend of the court. The service aspect of the activity comes out in the fact that although they may influence one another significantly, neither will impose on the other. The information or advice or guidance each provides is given on a take-it-or-leave-it basis; it is of the essence of co-reasoning that no one forces another's hand.[35]

C. Co-reasoning requires the eschewal of force and coercion and related activities

Any social activity, whether it be checkers or chess or co-reasoning, puts constraints in place that all participants must honor, on pain of no longer taking part in that activity; these are the constitutive rules or requirements of the practice. As the counterpart of personal reasoning, co-reasoning rules out a variety of intuitively objectionable activities in this manner. These are activities such that one cannot pursue them and claim at the same time to be reasoning with others, as those persons might reason with themselves. They include overtly forcing people to do something by removing one or more options; overtly coercing people to do something by penalizing and thereby changing one or more of their options; and overtly purporting to have contracted or changed the options in this manner. But they also include the covert activity of pretending to co-reason while actually undermining the practice, whether through misinforming others, or triggering some background manipulation of their ability to choose, or relying on the intimidating effect of one's power over them, or watching over and riding herd on their performance—surveying how it goes, while remaining disposed to resort to a corrective intervention in the event that co-reasoning does not lead them in the desired direction.

D. These are presumptively moral requirements of respect

The relationship established in co-reasoning allows one person to influence the choice of another but does not allow anyone to undermine another's power of choice; in that sense, its requirements enforce a regime of mutual respect. In deliberative decision-making, I have to assume with respect to each of the options among which I am deciding that it is within

[35] Co-reasoning need not involve any degree of rigor or formality. It may often be wholly implicit, as when you draw my attention to something and let me fill in the missing lesson. And it may be quite rhetorical in form. It may involve a story or a parable, in which you invite me to see things another way; it may introduce ironic or sarcastic or mocking comment; and it may employ metaphor and image, and all the colors of persuasive overture. Such rhetoric may be needed in order to knock me out of my complacency, let me see how stupid my point of view is, and make your standpoint seem truly habitable.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

my power; in choosing among three options, x, y, and z, I must be able to think with respect to each option: "I can do that."[36] If such can-do assumptions were not true, or if I did not think they were true, there would be no point in deliberation. Co-reasoning is an inherently respectful form of interaction insofar as it leaves both the truth and the thinkability of can-do assumptions in place. It does not undermine the truth of any such assumption by reducing my capacity to choose, contracting my range of options, or changing an option in the choice-set, say by the imposition of a penalty.[37] And it does not force me to reject any true assumption, misleading me about how things stand. Co-reasoning may make me aware of an extra option, as when you inform me that someone (perhaps you yourself) would be willing to offer a refusable reward for the choice of a particular option, x. But this does not directly affect the truth or thinkability of the can-do assumption with respect to option x or y or z; it merely adds the extra option, x+, of doing x and taking the reward.[38]

E. Co-reasoners purport to satisfy the requirements of respect

One can aspire to reason with others only on the basis of a shared understanding of the practice and an acceptance of the demands it imposes. Thus, to present oneself as co-reasoning is to purport to endorse those demands as constraints on how one is going to behave toward those with whom one means to reason. It is to renounce any resort to force or coercion and any reliance on resources of duplicity or intimidation or correction, and to present oneself as embracing a relationship of respect for one's partners. Those who purport to co-reason with others but then fail in some way to practice respect are hoist by their own petard. They can be indicted with failing to live up to the expectations that their presentation invited others to form; they can be condemned by criteria they are on record as endorsing.

F. Would-be co-reasoners also purport to satisfy the requirements of respect

Might a person subscribe to the requirements only when embarking on an episode of co-reasoning and not more generally? Not plausibly. I can

[36] We may sometimes describe an option in a way that reaches out to a desired consequence, as when we think of it as hitting the target rather than firing the gun, but each option has to be accessible also as something we can just do.

[37] For relevant arguments, see John Broome, *Weighing Goods* (Oxford: Blackwell, 1991); and Philip Pettit, "Decision Theory and Folk Psychology," in *Essays in the Foundations of Decision Theory*, ed. Michael Bacharach and Susan Hurley (Oxford: Blackwell, 1991); reprinted in Philip Pettit, *Rules, Reasons, and Norms* (Oxford: Oxford University Press, 2002).

[38] See Pettit and Smith, "The Truth in Deontology." The making of offers may not have this cast outside what I later describe as the circumstances of respect, where everyone is sufficiently well-off to be able to function properly in the local society. Extreme conditions might make it rational for someone to accept the offer of a slave contract, but that offer could hardly be said to be respectful. See Philip Pettit, "Freedom in the Market," *Politics, Philosophy, and Economics* 5 (2006): 131–49.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

hardly fail to honor these requirements in dealing with others and expect to have access to co-reasoning on an opportunistic basis. What assurance could others have of my willingness to respect them in the course of co-reasoning, if I present myself as someone ready to use force or coercion against them? What assurance could they have, indeed, if I present myself as ready in this way to use force and coercion against others who occupy a comparable position? In order to make it possible to resort to co-reasoning as opportunity requires, or inclination prompts—in order to have access to the practice—I must show myself ready to respect those with whom I might wish to co-reason, and any others who are likely to count as comparators. I must purport to relate to my community as someone who is respectful of others, disavowing any attempt to influence them other than by appeal to co-reasoning or to overtures (say, of camaraderie and bantering) that are congruent with co-reasoning.[39]

G. The requirements of respect have a default status in social life

It is possible for someone to opt out of a regime of mutual respect, manifestly displaying an attitude of aggression, or an assumption of occupying superior status. However, there is no need to opt in, at least not in circumstances where resources are sufficiently rich and well distributed to enable everyone to function properly as a member of the society—that is, not in what we might call the circumstances of respect.[40] The default assumption, absent hostility, will be that one is accessible to co-reasoning—at least to the extent that time and other constraints allow for such interaction—and that one eschews any resort to force or coercion or deception or the like in dealing with others. If natural life is lived in the precincts of battle, as Hobbes maintained, then civil life is lived in the precincts of conversation and co-reasoning. By not displaying hostility, one acquiesces in the manifest fact that others assume one is accessible to co-reason and ready to honor the associated demands of respect. One is naturally taken to endorse the requirements of respect as expectations to which one may be held, criteria by which one may be judged.

H. The requirements of respect are defeasible requirements: excuses

That the requirements of respect have this default status, however, does not mean that they are nondefeasible, and that a breach is never excused.

[39] The lesson is akin to that of the master and slave. What sensible slave would be honest in speaking his or her mind with a master? What sensible master would expect the slave to be honest? What sensible slave, indeed, would expect to be expected to be honest? Notice that if a class of masters is saliently distinguished, as under an apartheid regime, then masters may disrespect those in the slave class without disturbing their own reputations as persons who respect other masters; slaves will not count as comparators of the masters. That is why it is important, not just to focus on the instantiation of respect, but also on the promotion, as I describe it later, of respect-enjoyment.

[40] On such functioning requirements, see Amartya Sen, *Commodities and Capabilities* (Amsterdam: North-Holland, 1985).

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

So long as the practice of co-reasoning is in place, one will be strictly obliged by its rules.[41] But often the practice will not clearly be in place, and then a breach of the rules will be excusable. No one can be bound to display respect for those who would do him harm, for example, since those others thereby reject the relevance of the practice. Nor, by a plausible extension, can anyone be bound to display respect for those who would do harm to others who are party to the practice; no one can be bound to tell the truth about a victim's whereabouts to a would-be murderer. Nor can anyone be bound to display respect under such dire distress—such "moral catastrophe"[42]—that the very practice of respect is in jeopardy; the circumstances of respect, as I called them, would not then obtain. Disrespect is excused in cases like these, because the rules of co-reasoning do not clearly apply.

I. The requirements of respect are defeasible requirements: justifications

But are there grounds on which a breach of the requirements of respect in these or in other cases might be positively justified, not just excused? Is there any value, accessible to all, whose promotion might argue that whereas satisfying those requirements is generally for the best, there are cases where it is not and where a breach may be justifiable? Arguably, the very practice of respect points us to something that is valuable in this way: namely, the status that people enjoy insofar as they command the respect of others—if you like, "respect-enjoyment." The demands of such a value might argue for the justification of disrespect in perverse cases where, by showing disrespect to some innocent persons, the extent to which people enjoy respect overall is increased—at the limit, they may be saved from death or destitution. The recognition of such a value or target makes room for a consequentialist perspective on respect. One might argue that while it is generally for the best to practice respect, there may be cases where the promotion of respect-enjoyment (or, indeed, some other such value) justifies a temporary suspension of the practice, even in dealing with an innocent party.[43] The practice of friendship is not jeop-

[41] Thus, one will not be able to argue that one is co-reasoning with another, on the grounds that while one is violating the requirements of respect in some way, that is because this promises to maximize the respect of others for others. One might as well argue that one is playing chess, on the grounds that while one is violating the rules of chess, that is because such violation promises to minimize violations overall. The practice of co-reasoning requires one to instantiate respect toward those with whom one co-reasons, not to promote such respect overall. This may not be for the best, but it will be required for being able to claim to co-reason.

[42] See Robert Nozick, *Anarchy, State, and Utopia* (Oxford: Blackwell, 1974), 30n.

[43] Consequentialism might prescribe the honoring of the requirements of respect under the proviso that this is for the best overall. Or, more plausibly, it might prescribe it under a stricter proviso that imposes constraints on how far agents should check on what is for the best (because such checking might itself be bad for respect-satisfaction). One such proviso would prescribe the honoring of the requirements except when there is independently

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

ardized by the fact, acknowledged on all sides, that the demands of friendship are only compelling provided that they do not put innocent life at risk, or create some similar horror. Neither should the practice of respect be jeopardized by the recognition of a parallel, consequence-based proviso.

J. From ethics to politics

Not only is the emerging point of view consistent with a consequentialist perspective, but it is consistent with a consequentialism that would argue for the formation of a state—specifically, a state that would ensure that the circumstances of respect are realized for all and that there are sufficient guards against the strong riding herd on the weak that respect is widely commanded and enjoyed. Thus, the cause of promoting respect-enjoyment overall may argue for a collective, political arrangement under which the weak are protected and empowered and a regime of mutual respect is shaped so as to be for the best overall.[44]

If the argument encoded in the foregoing ten points is on the right track, then many moral obligations are really obligations under the code of co-reasoning, the ethic of respect. This ethic is grounded in the fact that, uniquely among animals, we human beings are creatures of the word and can relate to one another in the co-reasoning manner that speech makes possible. The constraints of the code are those associated with being, as we say, on speaking terms with one another. They describe the only possible basis for civil relationships in which conversation is the norm—not coercion, and not command. And they make salient the inclusive ideal of extending the enjoyment of respect and status to all. They do not cover the treatment we should offer to animals, of course, nor any of a range of

salient reason to think that this is not for the best overall; conformity to the requirements would become a default option under this proviso, but not an unconditionally compulsory one. See Philip Pettit, "A Consequentialist Perspective on Ethics," in Marcia Baron, Michael Slote, and Philip Pettit, *Three Methods of Ethics: A Debate* (Oxford: Blackwell, 1997). One way of seeing T. M. Scanlon's contractualism is as a theory that endorses, broadly, the requirements of respect in dealings between the respectful, and that argues for a distinctive, nonconsequentialist form of modulation for other cases: the right way to behave in those second-best cases is in accord with the principles for such cases that we might expect to prove reasonably unobjectionable among people respectfully debating with one another. See Philip Pettit, "Can Contract Theory Ground Morality?" in James Dreier, ed., *Contemporary Debates in Moral Theory* (Oxford: Blackwell, 2005).

[44] I take it that the sort of regime envisaged would maximize freedom in the republican sense of nondomination. See Philip Pettit, *Republicanism: A Theory of Freedom and Government* (Oxford: Oxford University Press, 1997). Another approach that might be taken to politics, starting from the practice of respect, is to ask in contractarian style about the sort of global regime that people might be led to adopt as a result of reasoning with one another in a situation of mutual respect. Yet another approach would be to ask about the global sort of regime that would emerge as a result of the different local arrangements that we might expect people to make with one another in co-reasoning contracts. The first approach is in the spirit of Rawls, the second in the spirit of Nozick.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

duties to ourselves, but they account for a good part of what we naturally think of as the moral realm.

How does this substantive theory of morality—this theory of a central subset of moral obligations—relate to the theories already canvassed? The most striking point of contrast is that, in Hume's terms, it is a theory of artificial obligation and virtue. It presupposes an artificial, cultural practice, one of conversation and co-reasoning, and presents the relevant obligations as preconditions for the survival and prosperity of that practice. The obligations are not presented as preinstitutional, inescapable laws of nature (or self-interest or benevolence or reason or justifiability), but as requirements of an institutional practice in which co-reasoning and respect are at the center. They are laws, in a word, of civility.

Their institutional character means that the laws of respect only have hypothetical or conditional status. They identify what is required of people if they are to have access to the practice of co-reasoning and to all the associated benefits. But this conditionalization is relatively innocuous, since it involves only a very small "if." The practice of co-reasoning is at the very center of human life and has an inescapable hold on us, at least in the normal case where it serves to promote the value associated with respect-enjoyment. If the laws of respect are requirements for access to that practice, as argued here, then they are conditional on a goal that is close to unrejectable. Unlike most hypothetical imperatives, they are tied to an aspiration that human beings can scarcely disavow, and thus they have the assured status that we would expect moral requirements—as distinct, say, from requirements of etiquette—to display.[45]

Substantive moral theories, as I have emphasized, seek out unifying patterns among the requirements that pass as moral demands, by the tests associated with our schematic theory of rightness. While a number of theories might each have something to be said in their favor, especially if they focus on different subsets of demands, they are clearly in competition. So how, finally, does the theory sketched here compare with the five theories we looked at earlier?

The laws of respect identify conditions for human flourishing, and connect on that front with the laws of nature. But they identify conditions that it could scarcely be in anyone's interest to violate, at least not in any general way, and thus they also make contact with the tradition that links morality with egoistic interests. How do the laws of respect relate to the remaining three theories we surveyed? These laws point us to requirements of benevolence that emphasize the evil of disrespect and the neutral ideal of prompting respect-enjoyment overall. They constitute ideals of reason in the sense of prescribing a mode of treatment that is saliently universalizable. And they connect with the idea of justifying oneself to

[45] See Pettit and Smith, "The Truth in Deontology," and Mark Schroeder, "The Hypothetical Imperative?" *Australasian Journal of Philosophy* 83 (2005): 357–72.

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press

others insofar as the regime they put in place generally allows each individual to exercise only the sort of influence that others should welcome; this would facilitate individuals' rationality without restricting their power of choice.

These observations suggest that the theory sketched here is on the right track, picking up a pattern that would explain why the other approaches have each seemed plausible to so many. If such a substantive theory is sound, then it means that there are two different standpoints from which we can survey the terrain of moral obligation. From the standpoint of the schematic theory described earlier (in Section II), moral obligations present themselves as requirements to which we are pointed by the battery of connections and paradigms that establish the very reference of terms like "right" and "wrong." From the standpoint of the substantive theory developed in this section, a core set of obligations stand out as requirements that derive from the code of co-reasoning, the ethic of respect, on which the very possibility of civil life is premised. Morality, under this picture, has a dual aspect and can be investigated in philosophy on either of its two fronts.

Philosophy, Princeton University

https://doi.org/10.1017/S0265052508080011 Published online by Cambridge University Press