DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Practical Reason and the Possibility of Error

Citation
Lavin, Douglas. 2004. Practical reason and the possibility or error. Ethics 114(3): 424-457.

Published Version
http://dx.doi.org/10.1086/381695

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:3121076

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http:// nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Practical Reason and the Possibility of Error*

Douglas Lavin

I. THE ERROR CONSTRAINT

That there is a deep connection between reason and the possibility of error is these days a philosophical commonplace:

> For a creature to be correctly said to have a rule, it is necessary that it should be able to break the rule.
>
> The physical or causal possibility of making a mistake, or doing what one is obliged, by what one means, intends, believes, and desires, not to do, is essential to the conception of such states and shows the essentially normative nature of their significance.
>
> An agent may be mistaken about what he has reason to do. . . . This is essential to preserving the point that statements of what people have reason to do have normative force; no account that excludes this can be adequate.
>
> Reason-giving explanations require a conception of how things ideally would be, sufficiently independent of how any actual individual's psychological economy operates to serve as the basis for critical assessment of it. In particular, there must be a potential gap between the ideal and the specific directions in which a given agent's motivations push him.

* For many helpful suggestions and criticisms I am grateful to Chrisoula Andreou, Melissa Barry, Matthew Boyle, Andrei Bukareff, Stephen Engstrom, John McDowell, Elijah Millgram, Ram Neta, Sebastian Rödl, Tara Bray Smith, David Sobel, and Matt Weiner. I received insightful written comments from two anonymous readers and should like to thank them and the editors of *Ethics* for their help. Above all, I am grateful to my teacher Michael Thompson for his invaluable advice and encouragement.

Ethics 114 (April 2004): 424–457
© 2004 by The University of Chicago. All rights reserved. 0014-1704/2004/11403-0002$10.00

There is no normativity if you cannot be wrong.[1]

Insistence on the link between reason and the possibility of error is just insistence on the normative character of reasoning. On any account, reasoning is activity governed or guided by norms, rules, standards, or principles, and so the very idea of this activity must contain the distinction between correct and incorrect application of them. Indeed, to most anyone who has thought about what a principle is and what it is for an individual to be responsible or subject to such a thing, it will seem that: *a reasoner is subject to a principle only if the reasoner can go wrong in respect of it.* And on this basis, most philosophers have accepted as a condition of adequacy on any account of being subject or responsible to a principle that it be able to distinguish between a reasoner's correct and incorrect application of the principle, between her faithfulness and unfaithfulness to it.

Our abstract claim has application to reasoners in their theoretical and practical capacities—as both knowers and doers—and so constrains accounts of both the understanding and the will; it is only the latter, however, which is of interest to me here. In the sphere of action and will, then, we might put the point as follows: *an agent is subject to a principle only if the agent can go wrong in respect of it.* This is the *error constraint* and will be the focus of what follows. While the error constraint certainly has the ring of truth, I will argue that it is not entirely clear what proposition is expressed by this form of words and that very much hangs on the resolution of the ambiguity and the subsequent determination of which renderings of the error constraint are true. So, let me begin by sketching two ways of understanding how the possibility of error might figure as a condition of being subject or responsible to a principle.

1. Jonathan Bennett, *Rationality: An Essay toward Analysis* (Indianapolis: Hackett, 1989), p. 17; Robert Brandom, *Making It Explicit* (Cambridge, Mass.: Harvard University Press, 1994), p. 14; Bernard Williams, "Some Further Notes on Internal and External Reasons," in *Practical Reasoning*, ed. Elijah Millgram (Cambridge, Mass.: MIT Press, 2001), pp. 91–97, pp. 92–93; John McDowell, "Might There Be External Reasons?" in his *Mind, Value, and Reality* (Cambridge, Mass.: Harvard University Press, 1998), p. 105; Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), p. 161. Of course, many others have explicitly endorsed the connection: Stephen Darwall, "Internalism and Agency," *Philosophical Perspectives* 6 (1992): 155–74; James Dreier, "Humean Doubts about the Practical Justification of Morality," in *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut (New York: Oxford University Press, 1997), pp. 81–99; Donald Hubin, "The Groundless Normativity of Instrumental Rationality," *Journal of Philosophy* 98 (2001): 445–68; Peter Railton, "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action," in Cullity and Gaut, eds., pp. 53–79; John Searle, *Rationality in Action* (Cambridge, Mass.: MIT Press, 2001), chap. 3; R. Jay Wallace, "Normativity, Commitment, and Instrumental Reason," *Philosophers' Imprint* 1, no. 3 (2001), http://www.philosophersimprint.org/001003.

I will then raise some questions about the two renderings of the error constraint that will occupy us for the remainder of the essay.

The error constraint is a complex proposition and the crucial ambiguity resides in the *can go wrong* component. But before commenting on that, it is necessary to say something, however sketchy and incomplete, about the others—namely about the concept *principle* and the relation *being subject to*, or, as I will often put it, *being under*. As I will understand it here, being subject to or being under a principle is simply being positioned within the scope of its authority; loosely speaking, the typical case includes being in a position to act as authorized. And as I will understand it, a principle is simply what an agent can abide by or follow. I mean to include under this heading such refined objects of practical philosophy as those possessing universal and necessary application to agents; but I also mean to include more ordinary objects like positive law, etiquette, crafts, and games, which do not possess universality and necessity. A principle is what I will call a *formal principle* if necessarily, all agents are under it; a principle is what I will call a *substantive principle* if this does not hold.[2] While this gives us a highly antiseptic idea of a *principle* and of the relation *being under*, it does so with a view to allowing the initial scope of this essay to be as wide as possible and also to avoid settling by fiat many of the questions steering it.

Now, a putative principle formulable as "Do A or don't do A" or "You must either do A or else not do A" isn't something which an agent can violate. Its logical form prevents us from describing or thinking anything at all that would count as a deviation from it: for this reason we want, I think, to say that its violation is logically impossible. And for this reason in turn we want to deny that an agent can be under it. Quite generally, the *logical interpretation* of the error constraint says that *an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate the principle*. It quite reasonably insists, more or less, that an agent is under a principle only if there is something for the agent to go wrong in respect of; really, the logical

2. There is a tradition of thought about practical reason according to which it is simply a capacity for the mechanical application of general and statable rules in particular circumstances. And by adopting the expression "principle" to pick out norms of action, and the phrase "being under a principle" to pick out being assessable in the light of a norm of action, I may give the impression that I mean to endorse this model. But that would be a mistaken impression: in speaking of "principles" and "being under principles" I merely assume, with the authors quoted at the very beginning of this section, that agents and actions are assessable in the light of norms. Whether norms have the form of mechanically applicable rules is not settled by this assumption. There is also a tradition of thought about practical reason according to which principles of reason are formal principles in my sense. I have adopted the expression "formal principle" in order to remain neutral on that question as well.

interpretation is about the principle and demands that this make a division within the space of action types.

The logical interpretation is concerned with what is possible independent of facts about the agent, and this is why we can determine whether a "principle" can be violated by an agent, in this sense, simply by looking at its content, or more accurately, at its aspiration to have content. In contrast, the other interpretation is concerned with what is possible given facts about the agent. It has to do with the coupling of agent to principle, or again with the specific nature of an agent's relation to the principle. Here, "can go wrong" means that it is possible for the agent to act out of accord with the relevant principle, perhaps because of *akrasia*, ignorance, vice, or some other feature of the agent. On our second interpretation, *an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate the principle and it is possible for the agent to do it.* According to this interpretation of the error constraint (itself subject to further refinement in Sec. III) an agent is under a principle only if she is in one way or another imperfectly hooked up with it. For reasons to be explained later, I will call this the *imperatival interpretation.*

It can now be said that the target of this essay is the question of the truth of the error constraint, on its imperatival interpretation. Since I am assuming that being an agent, that is, having the capacity to employ concepts in the service of action, has an essential normative dimension which I have expressed as *being subject to principles*, we might also characterize the main thread of this essay as an inquiry into whether liability to error is a condition of agency or will.

I begin with a suspicion that the seeming plausibility of the imperatival interpretation derives in large part from conflation with the error constraint on its obviously true logical interpretation. Exposing this conflation is the aim of the next section: there I develop the logical interpretation and suggest that the main appearance of the error constraint in the tradition of practical philosophy has been in the guise of the logical interpretation. Because the error constraint is ambiguous, it is not easy to establish commitment to the imperatival interpretation of it, though the thought lurks in the background of much of the recent work on practical reason. And while our real target is imperativalism in all its forms, I will address Christine Korsgaard as its most dynamic and articulate advocate. She is especially clear about endorsing the imperatival interpretation and almost unique in advancing considerations on its behalf. Furthermore, Korsgaard's influential argument for the claim that instrumentalism is flatly incoherent employs the imperatival interpretation as a premise: to the extent that imperativalism is impugned,

we are in possession of a defense of the coherence of instrumentalism.[3] Section III investigates this connection between imperativalism and instrumentalism and, in the course of doing so, specifies weak and strong versions of imperativalism, directing all of our subsequent attention to the latter. Then, to put us in a position to evaluate the imperatival interpretation, Section IV reminds us of a few abstract features of Kant's discussion of imperatives and their distinctive role in the rational explanation of action. Section V tries to understand the allure of imperativalism, concluding that it is not forced on us by argument.

Even if the first five sections succeed in specifying several senses of the error constraint and undermining arguments purporting to compel adoption of the imperatival interpretation, the question of imperativalism's truth is still left open. The final three sections make some progress toward closing that, though they do not get it all the way shut. Still, even without proving that imperativalism is false, these sections reveal why it ultimately matters whether it's true or not.

3. The argument is developed at length in Christine Korsgaard's "The Normativity of Instrumental Reason" (in Cullity and Gaut, eds., pp. 215–54). Of course, neutralizing Korsgaard's argument does not prove that instrumentalism is coherent—there may be other ways of administering the poison. Even so, it is worth asking why the more general project of vindicating the coherence of instrumentalism is of any interest. The most straightforward answer would be that it is a step toward establishing its truth. Seems a small step in that direction, though: consider how anemic a defense of the truth of, say, the phlogiston theory of combustion would be on the grounds of its coherence. Leaving aside the question of instrumentalism's truth, for the moment, we can still see that its mere coherence would create trouble for anyone hoping to anchor more-than-instrumental principles in an abstract analysis of rational agency, a strategy typically associated with the Kantian tradition in practical philosophy. Here's the trouble. If instrumentalism is coherent, then it is possible that there is some sort of creature whose practical reason operates in such a way as to be completely characterized with the resources of instrumentalism. Maybe this is how it is with the Martians? The Martian would then be a rational agent, of a sort, and yet would be entirely limited to employing reason in the service of achieving further aims. On its face, this mere possibility places a wedge between possession of practical reason and being subject to more-than-instrumental standards and disrupts the attempt to derive the latter from the former. What, then, of the truth of instrumentalism? This is a much longer story and one entirely independent of this essay. Still, what I would like to say is that the question itself becomes difficult to place if it is fixed that instrumentalism is coherent and also that we human beings exhibit reason giving and asking behavior that outstrips anything adequately described with the limited resources of instrumentalism. In that case, instrumentalism would capture how it is with Martian practical reason and yet fail to capture how it is with ours. But now what are we to make of the very question of instrumentalism's truth? Hopefully, this compressed line of thought can, when expanded, bring us closer to seeing the place for an analysis of the concept of practical reason on which it admits of species or forms; the various positions, e.g., instrumentalism, might then be seen as simply purported specifications of forms of agency. The tendency of the current literature is very different; it is to treat the various positions as articulations of competing conceptions of some single concept of practical reason. But this is all offstage here.

In Sections VI and VII, I extract a pair of suspect implications of the imperatival interpretation—the conception of freedom as the liberty of indifference and the impossibility of perfectly rational agency. That imperativalism entails such theses is not a reductio ad absurdum, though many may thereby be spurred to reject it, perhaps because of prior and deeper commitments, say, to determinism or theism. However, my focus is elsewhere, chiefly on the way these implications, and thus imperativalism, confine our understanding of what it is to be under a principle in the first place. Our conception of the metaphysics of practical normativity is confined, I argue in Section VIII, because imperativalism is incompatible with *constitutivism*, a strategy for elucidating the authority of a principle by revealing it to be internal to what is under it and the main alternative to *platonism* among nonreductive and realist accounts of practical norms. That constitutivism is also a strategy Korsgaard perspicuously articulates and adopts as her own is of local interest but does not, of course, decide the question of how far to write the possibility of error into agency. About that we must settle for more modest, conditional results: if we go as far as imperativalism would have us go, we are prohibited from a constitutivist understanding of principles and then forced to choose between improbably writing them into "a world beyond, which exists God knows where," or drearily giving up on the idea of there really being standards of correctness for action.[4]

II. THE LOGICAL INTERPRETATION

To see the force of the logical interpretation of the error constraint it will help to consider two occasions of its use. In book 1, chapter 3 of *On the Social Contract*, "On the Right of the Strongest," Rousseau tells us that the principle "Obey those in power," where it has the sense of "yield to superior physical force," is merely apparently meaningful. Its absurdity is revealed by the fact that "it will never be violated," that is, its violation is inconceivable. But why can't we think its violation? In a characteristically adept turn of phrase Rousseau reveals the reason: "As soon as force makes right, the effect changes along with the cause. Any force that overcomes the first one succeeds to its right."[5] His idea here

4. G. W. F. Hegel, *Elements of the Philosophy of Right*, ed. Allen Wood, trans. H. B. Nisbett (Cambridge: Cambridge University Press, 1991), p. 20. Limited space requires developing the incompatibility claim in the restricted context of formal principles; though I believe that a parallel point can be made in connection with substantive principles, our official results are nevertheless suitably restricted. That so much of the current debate about the content of the principles of practical reason gets played out in terms of a debate about what the formal principles are ensures that this restriction does not sever our discussion from the core of the current discussion.

5. Jean-Jacques Rousseau, *On the Social Contract*, ed. Roger D. Masters, trans. Judith R. Masters (New York: St. Martin's, 1978), p. 48.

is that the truth of any judgment of relative physical power is determined or fixed by the outcome of actual struggle, so that "X is more powerful than Y" just means "X overcomes Y in a struggle." Given this, it must be that the superior physical power is always obeyed. And if having a right is understood in terms of being strongest or having superior physical power, it must be that the strongest has what is his by right.

In his second sermon at the Rolls Chapel, "Upon Human Nature," Bishop Butler employs a structurally identical argument against the principle "Act as you please," which he interprets as recommending "following that principle or particular instinct which is for the present strongest."[6] Butler develops the argument as follows: "If by following nature were meant only acting as we please, it would indeed be ridiculous to speak of nature as any guide in morals, nay, the very mention of deviating from nature would be absurd; and the mention of following it, when spoken of by way of distinction, would absolutely have no meaning. For did ever any one act otherwise than as he pleased?"[7] Butler's thought is that if every action arises from desire, and if the truth of any judgment of the strength of desire is determined by the outcome of actual struggle, then it must be that the strongest desire is acted on. Here, it is inconceivable that an agent act, yet not act on the strongest desire; thus it is impossible that an agent be subject to the principle "Act as you please."

The preceding interpretation of Rousseau's objection to the doctrine of the right of the strongest isn't without its difficulties. I would like to raise one with a view to saying a bit more about the structure of the positions that run afoul of the logical interpretation. I have in mind someone who points out that "power" has more than one sense, only one of which is amenable to my account. On the one hand, and perhaps most commonly, we think of a power as a capacity—as something with a kind of generality. Power, in this sense, is something one might have at a time even if not exercising it just then, like Harry Houdini's powers of sight and locomotion when he's blindfolded and stuffed inside a coffin. There is also a sense of "power" as a particularized and occurrent force—the sort of force an individual tsunami rushing toward Honolulu is thought to have. If Rousseau is operating with the former understanding of "power," then the truth of a judgment of relative power is not determined by the outcome of conflict and in a conflict between powers the greater power might lose. We could then correctly say "Monsieur R. had the power to defeat Monsieur D." even when he didn't. We might

6. Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, reprinted in *Five Sermons Preached at the Rolls Chapel and A Dissertation upon the Nature of Virtue*, ed. Stephen Darwall (Indianapolis: Hackett, 1983), p. 35.

7. Ibid., p. 36.

explain, "If he hadn't slipped on that banana, he certainly would've won the duel." Consequently, the strongest might not have what is his by right. So, if Rousseau's objection to the doctrine of right of the strongest is as I have taken it to be, then we ought to find him appealing to the particularized sense of power. And this is precisely what he is doing when he says that "yielding to force is an act of necessity, not of will."[8] Were "yielding to force" a species of yielding to something general, to a capacity, then yielding would have to be an act of will, what we might regard as an expression of prudence or, perhaps, terror. This is because one must be able to yield to such a thing even when the power or capacity isn't being exercised, and this sort of yielding can only be the expression of will. So, we must take Rousseau to be operating with the particularized sense of power and force in his argument against the very sense of the attempt to explain what it is for X to have a right against Y in terms of facts about the relative strength of X and Y. Now, I take it that behind Butler's gloss of "Act as you please" as recommending "following that principle or particular instinct which is for the present strongest" is an acute awareness of this complication about the general and particular senses attaching to "power" and "strength," as well as a commitment to a highly particularized sense.

Our original "Do A or don't do A" wears its emptiness on its sleeve and so only appears in real life as the expression of helplessness or frustration: "Keep drinking or don't. I don't care." Because they are much better dressed, "Yield to the superior physical force" might appear as the basis of a doctrine of political right, and "Act as you please" might appear as the basis of a doctrine of human flourishing—nevertheless, each is as empty.

The emptiness of these purported principles can be brought into somewhat better focus by attending to the following. The execution conditions of a principle are just the truth conditions of a certain proposition. And in all of our cases the relevant propositions are necessary truths. In the "Do A or don't do A" case, the necessity of the corresponding proposition derives from its logical form, while in the interesting cases the necessity of the corresponding proposition derives from a relation between the contents of the component expressions. The shift from the transparently absurd to the philosophically tempting cases is something like the shift from "Kill a bachelor or don't kill a bachelor" to "Kill a bachelor or don't kill an unmarried male." The particularized understanding of power, force, and strength that Rousseau and Butler employ is really in the service of establishing a meaning connection between "X is more powerful than Y" and "X defeats Y in a struggle" and also one between "Z is the strongest desire" and "Z is the desire

8. Rousseau, p. 48.

acted on." With such connections in place "Yield to the superior physical force" and "Act as you please" become very much like the second bachelor case.

Rousseau's thought is not simply that it is false that the right of the strongest provides the foundation for political right, but that the idea of the right of the strongest is an "inexplicable confusion" and "meaningless." According to Butler, "Act as you please" is "absurd" and "ridiculous" as a principle of life. The incoherence of these derives from their purporting to be standards of correctness while also being necessarily in accord with every describable state of affairs. This is so because their supposed content is specified in terms of what, in fact, happens. One would like to say: whatever happens is what ought to happen and that only means that here we can't talk about what ought to happen.[9]

III. IMPERATIVALISM AND INSTRUMENTALISM

In this section, I develop the imperatival interpretation of the error constraint, including a specification of weak and strong versions, by considering an occasion of its use. In particular, I consider Korsgaard's employment of imperativalism in her attempt to show that instrumentalism is simply incoherent.

9. I allude to Wittgenstein's famous remark at sec. 258 of the *Philosophical Investigations*, trans. G. E. M. Anscombe (New York: Macmillan, 1958) to suggest that he employs the logical interpretation in the course of attacking attempts to understand the content of "sensation language" on the model of private inner ostensive definition. His point there is that on such a picture "sensation language" could not provide a "criterion of correctness" and so could not be said to mean anything at all. It might be worth noting that what sits at the foundation of the projects Rousseau and Butler devastate is very much like the organizing impulse of the dispositional account of meaning or content Saul Kripke considers in *Wittgenstein on Rules and Private Language* (Cambridge, Mass.: Harvard University Press, 1982), pp. 22–37. While not interested in what it is to have a right, or what it is to follow nature, Kripke is nevertheless investigating something normative—in particular, the fact in virtue of which an expression has meaning. The dispositionalist answer is that facts about how a language user is disposed to use an expression determine the meaning of the expression. Kripke argues that the dispositional view is ultimately an equation of performance and correctness and, for this reason, can't make sense of the possibility of mistake. This, in turn, reveals that the dispositionalist can't meet a basic condition of adequacy on accounts of meaning—namely, that any candidate for the fact in virtue of which an expression has meaning must be such as to ground the normativity of meaning, i.e., must contain a specification of the correct use of that word. In the end, each of the views attacked aims at grounding and explaining the relevant norms in baldly naturalistic and reductive terms, and each consequently washes up on the logical interpretation of the error constraint. Although I have not touched on revealed preference theory or its connection with the theory of rational choice, it seems clear that the criticisms this coupling has undergone in, say, the work of Simon Blackburn and David Gauthier are of the Butler-Rousseau type. See Blackburn's "Practical Tortoise Raising," *Mind* 104 (1995): 695–711, as well as his *Ruling Passions* (New York: Oxford University Press, 1998), pp. 161–68. And see Gauthier's *Morals by Agreement* (New York: Oxford University Press, 1986), pp. 26–29.

Most philosophers are familiar with G. E. M. Anscombe's poisoner—"I'm pumping in order to replenish the house water supply," "I'm replenishing the house water supply in order to poison the inhabitants"—or if not with this fellow, then with Donald Davidson's scribe: "I'm writing the letter 'a' in order to write 'act'," "I'm writing 'act' in order to write 'action'." But one needn't have read Anscombe or Davidson to be familiar with the subject matter of their philosophical reflections—the sort of explanation of action exhibited here, broadly speaking explanation of the form "X is doing A in order to do B." That is, we are all familiar with teleological explanation of action or, again, explanation of action in terms of an agent's further purpose. And we are, I suppose, equally familiar with the form of advice and form of practical inference internally related to this. Let's call an agent whose intentional operations admit only such explanation, who takes only such advice, and who conducts only such practical inferences, a *purely instrumental agent*. Moreover, we might call the view that this is all there is to practical reason *instrumentalism*.

Recently it has become fashionable to speak of the *instrumental principle* and then to characterize cases of doing A in order to do B as following the instrumental principle, to characterize being suited to receive advice like "you ought to do A seeing as you propose to do B" as being subject to the instrumental principle, and along the same lines, to characterize the instrumentalist as holding that the instrumental principle is the only formal principle of practical reason—that is—as holding that the principle enjoining one to take the (necessary) means to one's end is the only principle of which it is true that necessarily, if X is an agent, then X is under it.[10] With such ideas and terminology in place, we can make the following claim: something is an agent only if it is subject to the instrumental principle.

Now, a starting point of this essay is that being an agent, or having the capacity to employ concepts in the service of action, has an essential normative dimension, which I have formulated as *being subject to a principle*. Thus, when we are asking about the conditions of something's being subject to a principle we are also asking about *the conditions of agency*. Imperativalism is a component of a view about this; it is a specification of an essential or necessary feature of being an agent. And like imperativalism, the claim that something is an agent only if it is subject to the instrumental principle is also a condition of agency claim, though of a slightly different sort; it is a substantive claim about which principles

10. Elsewhere I would complain about this way of describing instrumentalism and its conception of end-oriented agency, but such complaints would be a distraction here. See my "There Is No Such Thing as the Instrumental Principle" (unpublished manuscript, University of Pittsburgh, 2003).

an individual must be under if that individual is to be an agent at all. It is a substantive view about formal principles and not, like imperativalism, an abstract specification of the conditions of being subject to principles simply.

Some, like James Dreier, have argued that the real appeal of instrumentalism derives from the special status of the instrumental principle, a status he argues no other principle has: "The special status of instrumental reason is due to its being the *sine qua non* of having reasons at all."[11] Others, like Korsgaard, have argued that while being under the instrumental principle is indeed a condition of agency, there is only an appearance of uniqueness here and that when we tease out the conditions of instrumental agency itself, we will appreciate that "if there are any instrumental requirements, then there must be unconditional requirements as well."[12] The reason I am raising such matters in this essay is that Korsgaard's analysis makes an ineliminable appeal to the imperatival interpretation. I would like to look into this.

In "The Normativity of Instrumental Reason," Korsgaard argues against the neo-Humean doctrine that "the instrumental principle is the *only* requirement of practical reason." That is, she argues against the doctrine that purely instrumental agents are the only agents there are. More specifically, Korsgaard defends the thesis that "the view that all practical reason is instrumental is incoherent." But why are we to accept that there cannot be purely instrumental agents? According to Korsgaard, a purely instrumental agent "cannot violate the instrumental principle" and for this reason cannot be under it.[13] But how are we to hear her claim that a purely instrumental agent can't go wrong? Along logical or imperatival lines?

Since the logical interpretation is weaker than the imperatival interpretation, it is obviously permissible for an imperativalist to rule out a view on the grounds that it does not meet the weaker logical condition. So, someone might grant that Korsgaard is an imperativalist, which she undoubtedly is, but still recommend reading her attack on the coherence of instrumentalism as structurally similar to the Butler-Rousseau arguments I described earlier.[14] Were this interpretation correct, putting pressure on the imperatival interpretation would not help to vindicate the coherence of instrumentalism, as I am hoping it can.

But this reading of Korsgaard cannot be correct. Recall Butler's

11. Dreier, p. 99.

12. Korsgaard, "The Normativity of Instrumental Reason," p. 252.

13. Ibid., pp. 220, 251, 231.

14. Melissa Barry raised the possibility of this reading of Korsgaard's argument in helpful comments on an ancestor of this essay. For Korsgaard's explicit endorsement of imperativalism, see "The Normativity of Instrumental Reason," pp. 236 and 240, as well as *The Sources of Normativity*, pp. 29–30, 137–38, 146, and 161–64.

argument: from the fact that X is not acting on a particular desire, it follows that that desire is not strongest. If so, it is impossible that X fail to act on the strongest desire. And thus, an agent can't be under the principle "Act on the strongest desire." The principle under consideration in Korsgaard's essay is the instrumental principle: take the (necessary) means to your end. It is clear that this principle cannot fall to a Butler-like argument once in the instrumentalist's hands. This is because the instrumentalist does not, and need not, make the relevant meaning connection between "X has an end" and "X is doing what is in fact necessary for achieving an end." According to a reasonable instrumentalist, from the fact that X is not doing what is necessary for doing A, it does not follow that X does not aim at doing A. This is enough to enable the instrumentalist to describe a state of affairs which would be out of accord with the instrumental principle. For example, according to the instrumentalist, from the fact that D. D. is not turning left, when turning left is necessary for getting to Katmandu, it does not follow that D. D. doesn't have the objective of getting to Katmandu—he might be reading the map upside down. What does this tell us about the place of the error constraint in Korsgaard's argument against instrumentalism?

Since the instrumentalist can satisfy the error constraint on the logical interpretation, charity recommends reading Korsgaard's core argument against instrumentalism as employing another interpretation. But, as the plausibility of our example shows, it is also true that the instrumentalist can satisfy the error constraint on the imperatival interpretation—at least, on the assumption that instrumental agents are susceptible to error in action that derives from some defect of theoretical reason. Moreover, Korsgaard grants that the instrumentalist is entitled to admit cases of error in action of this character, miscalculations and mistakes. But she also says "the possibility of mistake is not in general very interesting" and does not permit the instrumentalist to locate the requisite place for the possibility of error by appealing to it.[15] Since the instrumentalist can satisfy the error constraint even on the imperatival interpretation, and because Korsgaard seems to see this, charity requires reading her core argument as employing another interpretation of the error constraint. But what is it?

To the extent that exercises of practical rational capacities depend on exercises of theoretical rational capacities, error in the former can be derivative of error in the latter: a faithful man who has nevertheless forgotten a promise made some time ago might, for that reason, fail to keep it; a physically adept woman who is simply ignorant of a relevant means-end connection might thereby fail to achieve her end. However,

15. Korsgaard, "The Normativity of Instrumental Reason," p. 227.

in such cases the error in action does not lie specifically in the misuse of a practical capacity. Instead, it derives from error that is already present in the conditions of the exercise of the relevant practical capacity. And such error does not express genuinely practical imperfection.[16]

Given the distinction between derivative and genuine practical imperfections, we are in a position to distinguish weak and strong versions of the imperatival interpretation. The *weak imperatival interpretation* is what we introduced early in this essay: an agent is subject to a principle only if there is some kind of action such that if the agent did it she would violate the principle and it is possible for the agent to do it, and it doesn't matter why the agent does it. The *strong imperatival interpretation* adds a further condition, one having to do with the source or origin of the error. It says that *an agent is subject to a principle only if there is some kind of action such that if the agent did it she would violate the principle and it is possible for the agent to do it and her doing it would be a genuinely practical error.* That is, the error in action must be error that does not depend on error in theoretical reason quite generally and is the expression of a genuinely practical defect.

Once the distinction between weak and strong versions of the imperatival interpretation is in play, we must consider where weak imperativalism fits in the scheme of this essay. In Section I, I said that I would not take up the question of the way in which the possibility of error figures as a condition of a theoretical reasoner's being subject to norms. But the question of the truth of weak as opposed to strong imperativalism will raise exactly that. So, I put it to the side and focus on the truth of strong imperativalism in what follows.

There is direct evidence that Korsgaard endorses the strong imperatival interpretation. The sort of violation, error, or going wrong that Korsgaard has in view is exemplified in cases in which "people's terror, idleness, shyness, or depression is making them irrational and weak-willed." Unlike cases of mistake, the actions that result here are "strictly speaking, irrational," they exhibit "true irrationality, by which I mean a

16. What are the distinctively practical imperfections? An adequate answer to this question would require the specification of the forms of practical thought, the specification of the modes of reason in action. This is because if there is a kind of practical defect then there must be a kind of practical thought or form of practical reason correlated with it. If this is right, then we can't even begin to touch on a substantive characterization of the forms of practical defect. My thinking about the idea of distinctively practical error and defect has benefited from Stephen Engstrom's "Contradictions in the Will" (unpublished manuscript, University of Pittsburgh, 2001).

failure to respond appropriately to an available reason."[17] All signs point to an interpretation of Korsgaard's main argument against instrumentalism on which it employs the strong imperatival interpretation and is really this: a purely instrumental agent can't be distinctively practically defective (truly irrational or irrational strictly speaking) with respect to the instrumental principle, and for that reason can't be under it.[18] But why is the possibility of distinctively practical error necessary? Why is strong imperativalism true?

IV. KANT ON IMPERATIVES

To put us in a better position to assess the imperatival interpretation, I want to remind us of two very general features of Kant's own account

17. Korsgaard, "The Normativity of Instrumental Reason," pp. 229, 228; Christine Korsgaard, "Skepticism about Practical Reason," in her *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), pp. 311–34, p. 318.

18. Among other things, the second section of "The Normativity of Instrumental Reason" develops a complex argument for the claim that a purely instrumental agent can't be distinctively practically defective and, as a whole, has prompted much discussion. Two contributions to the discussion are especially relevant to the line of thought I am pursuing here. In "The Groundless Normativity of Instrumental Rationality," Donald Hubin argues that the crux of Korsgaard's argument against instrumentalism cannot be that it fails to meet the error constraint, or in Hubin's terms, that it fails to be "normatively demanding." This is because he sees how easily instrumentalism can meet both the logical interpretation and the weak imperatival interpretation. Nevertheless, Hubin does not see that Korsgaard's interest is in the strong imperatival interpretation, and thus, I think, he fails to address her real concern. David Sobel's "Subjective Accounts of Reasons for Action," *Ethics* 111 (2001): 461–92, also records the ease with which the instrumentalist can make sense of derivative error in action. Although Sobel expresses some concern about how the instrumentalist might meet what I am calling the strong imperatival interpretation, he does not develop an account of that. As I hope is obvious, I am instead recommending that the putative condition of agency be resisted. Elsewhere I suggest that behind the strong imperatival interpretation is the thought that action is not the upshot of theoretical reason operating with a special set of contents—"mince pie syllogizing" as Anscombe puts it—but, rather, the upshot of a rational capacity of an entirely different kind. When conjoined with the intuition that kinds of rational capacity are in some sense tightly connected with kinds of error, we come close to the strong imperatival interpretation. However, we can capture the heart of the intuition without making imperativalist commitments by appeal to the much weaker *practical defect constraint*: for every kind of practical thought there is a unique kind of practical defect connected with it. It is a further and nontrivial step to claim that each individual bearer of a capacity for thought of the relevant kind must potentially exhibit such a defect or really be able to mess up in the relevant way. This raises a further difficulty. In light of the practical defect constraint, someone might claim that Korsgaard's argument against instrumentalism appeals only to this and not at all to the imperatival interpretation. That is, someone might insist that Korsgaard's complaint with instrumentalism is that it can't make sense of the very idea of genuinely practical error. While there is little textual support for this interpretation, and much against it, there is good reason to reconstruct Korsgaard's argument along these lines, if only to see where it might ultimately go wrong. I pursue this project in "There Is No Such Thing as the Instrumental Principle."

of imperatives and their distinctive role in the rational explanation of action, features which have largely gone missing in subsequent debates nevertheless conducted in its terms. We might put the first this way: while the imperative is prior in the order of what is known to us, it is not, according to Kant, prior in the order of nature. The second concerns the restricted place of explicitly prescriptive or deontic thoughts, for example, I ought to do A, in the explanation of action, as Kant understands it.

Before moving on, a word of caution: we are considering the content and credentials of imperativalism, a thesis attaching to the abstract genus of *principle* which potentially contains, say, rules of a game, craft, and positive law as species. Yet most of what Kant has to say about imperatives takes place against the background of a conception of principles as general, of rational principles as formal principles (in my sense), as well as a highly determinate conception of what the formal principles are. In this section, I bracket any divergence over such matters and restrict my attention in a similar fashion. This is not because I am here giving up on the generality of the claim I am investigating; rather, ease of exposition of Kant's conception of an imperative recommends such a route.

It is no mere coincidence that in each of Kant's most elaborate discussions of the nature of an imperative in general he first introduces the concepts of a practical principle and a practical law: "Practical *principles* are propositions that contain a general determination of the will. . . . They are objective, or practical *laws*, when the condition is cognized as objective, that is, as holding for the will of every rational being."[19] Such laws contain specifications of what is practically necessary. By "necessary" Kant means "in accord with a rule" and by "practically necessary" he means "in accord with the rule given by objective laws of the good" or again, the rule given by the concept of the good.[20] A practically rational being is, for Kant, simply something with the capacity to act according to a conception or representation of a law. Or again, a practically rational being is something with the capacity to employ the concept *good*, or to act in the light of the rule given by that concept. As with other abilities, capacities, or powers, this can be possessed either perfectly or imperfectly. In the first case, where reason infallibly determines the will and what is done is necessarily in accord with the relevant

19. Immanuel Kant, *Critique of Practical Reason*, 5:19. Also see *Groundwork of the Metaphysics of Morals*, 4:412–14, and *The Metaphysics of Morals*, 6:221–22. All translations are from *Practical Philosophy*, ed. and trans. Mary Gregor (Cambridge: Cambridge University Press, 1996). All references to Kant's writings are given in the notes by the volume and page number of Kant's *Gesammelte Schriften*, ed. Royal Prussian Academy of Sciences (Berlin: de Gruyter, 1902–).

20. Kant, *Groundwork*, 4:412–14.

law, "no imperatives hold . . . the 'ought' is out of place."[21] It is only with the introduction of the idea of an imperfectly rational being, one which has the capacity to act on the representation of laws, but which might also misuse that capacity, that the idea of an *imperative* has any room to appear: "An imperative differs from a practical law in that a law indeed represents an action as necessary but takes no account of whether this action already inheres by an *inner* necessity in the acting subject (as in a holy being) or whether it is contingent (as in the human being); for where the former is the case there is no imperative."[22] So, it is no accident that Kant's discussion of the nature of an imperative is preceded by discussion of practical principles and the subset of these which are laws, because an imperative is, in some sense, derivative of these. An imperative is, in essence, a law considered in relation to a certain sort of being, namely, an imperfectly rational one: "Imperatives are only formulae expressing the *relation* of objective laws of volition in general to the subjective imperfection of the will of this or that rational being, for example, of the human will."[23]

But when, according to Kant, is something imperfect in such a way as to receive principles as imperatives? Not just any imperfection will do. Were, for example, a creature to lose its capacity for growth prematurely, it would have an imperfection to be sure, though this need not have any bearing on the facility and grace of its reason. But does Kant think that any rational imperfection suffices for turning laws into imperatives, as it were?

Given the distinction between derivative and genuine practical imperfections sketched in the previous section, we are in a position to claim that an agent is related to a principle as an imperative just when the agent has the relevant distinctively practical imperfection. Sometimes this seems to be Kant's view: "[Imperatives] say that to do or omit something would be good, but they say it to a will that does not always do something just because it is represented to it that it would be good to do that thing." However, there are other passages which lend support to a different interpretation—in particular, one on which the possibility of, say, mere ignorance of fact is sufficient for turning principles into imperatives. "The imperative thus says which action possible by me would be good and represents a practical rule in relation to a will that does not straightaway do an action just because it is good, *partly because*

21. Ibid., p. 414.

22. Kant, *Metaphysics of Morals*, 6:222.

23. Kant, *Groundwork*, 4:414, italics mine. Of course, Kant distinguishes between hypothetical and categorical imperatives, and some might read his remarks about the good as having application only to categorical imperatives and the sublime rational capacities which underlie them. I am not reading him this way and think that the thought experiment described in the *Critique of Practical Reason* at 5:58–59 gives me some license to do so.

the subject does not always know that it is good, partly because, even if he knows this, his maxims could still be opposed to the objective principles of a practical reason."[24] Any attempt at a resolution of this interpretative tension is another essay; nevertheless, it is interesting to describe these conceptions of the conditions under which an agent is related to a principle as an imperative, if only to emphasize the parallel with weak and strong versions of the imperatival interpretation, each of which Kant would reject.

Now that we have said something about the conditions under which an agent is related to a principle as an imperative, we can ask, "What, if anything, is distinctive about the *relation*?" Since the principles we have in view here are practical, that is, they figure as the determining grounds of action, we should expect to find something distinctive about the way in which they figure in the determination of an imperfectly rational will. The concept which Kant introduces to locate what is special here is *necessitation* (*Nötigung*) or *constraint* (*Zwang*) and is to be distinguished from practical necessity, which he glosses as *goodness*. What he says is that an imperfectly rational being is necessitated or constrained to act in a certain way by the representation or conception of objective practical principles, or again an imperfectly rational being is necessitated to do A by the appreciation that doing A is practically necessary or good.[25]

According to Kant, practical and deontic thoughts, for example, I must do A, I ought to do A, and I should do A, are expressions of rational constraint or necessitation. Thus they figure in the practical thinking or among the grounds of action only of imperfectly rational beings.[26] This restricted understanding of the place of deontic thoughts in deliberation and action explanation, if it is in fact Kant's, would not be idiosyncratic. Hume's view on the matter seems to be much the same. Although Hume famously refuses to give "the motive of duty" a foundational, or even significant, place in his account of moral motivation, he nevertheless wants to provide some account of action on its basis: "But may not the sense of morality or duty produce an action, without any other motive? I answer, it may: But this is no objection to the present doctrine. When any virtuous motive or principle is common in human nature, a person, who feels his heart devoid of that principle, may hate himself upon that account, and may perform the action without the motive, from a certain sense of duty, in order to acquire by practice, that virtuous principle, or at least, to disguise to himself, as much as

24. Kant, *Groundwork*, 4:413, 414, italics mine.
25. See, e.g., ibid., 4:412–14, and *Metaphysics of Morals*, 6:222.
26. Kant, *Metaphysics of Morals*, 6:379.

possible, his want of it."[27] Whatever else we have to say about the passage, this much is certainly contained in it: only what is imperfect, indeed only what appreciates its own imperfection, acts on the motive of duty, and, I would say more generally, on explicitly prescriptive or deontic thoughts. It is a striking fact that these are here a secondary case of acting for a reason. And this should make us pause. In much contemporary work, curiosity about the motivational powers of reason is focused in the first instance on the question of the truth of judgment internalism—the thesis that it is a necessary condition of a genuine instance of a judgment that I ought to do A, that I am disposed to act in a way appropriate to it. But the tradition pulls us in a very different direction in holding that acting on explicitly deontic thoughts is not the grassroots level of acting for a reason, and moreover that it has a place only where the agent is imperfect.[28] This is not, I suggest in Section VIII, a mere historical artifact of Kant's conception of practical reason but crucial to the explanatory strategy he is pursuing.

With so much said about the imperative and its Kantian origins, let's return to the main thread of the essay. It is, we now know, plainly and trivially true that an agent is subject to an imperative only if it is possible for the agent to violate it. This much is simply contained in the definition of an imperative. But why think that this is a condition that applies to being under a principle quite generally? Why think that liability to error is a condition of reasoning or principled agency? Why think that principles are imperatives? Or in our terms, why think that the imperatival interpretation is true?

V. THE ALLURE OF IMPERATIVALISM

Thus far the focus has been the description of three, progressively stronger interpretations of the error constraint: the logical interpretation, and the weak and strong imperatival interpretations. In the course of presenting these, the soundness of Korsgaard's influential argument against instrumentalism was linked to the truth of strong imperativalism.

27. David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and rev. P. H. Nidditch, 2d ed. (Oxford: Oxford University Press, 1978), p. 479.

28. For a helpful classification of substantive theses in the theory of practical reason going by the names of "internalism" and "externalism," see Stephen Darwall, "Internalism and Agency," and *The British Moralists and the Internal 'Ought': 1640–1740* (Cambridge: Cambridge University Press, 1995), pp. 9–12. Darwall is right that judgment internalism has figured prominently in contemporary arguments for practical noncognitivism. Still, as Darwall recognizes, the general orientation is not limited to noncognitivists. See, e.g., John Broome's "Reason and Motivation," *Proceedings of the Aristotelian Society* 71, suppl. (1997): 131–46. I have not argued for this restricted conception of action on deontic thoughts. Still, if it is correct, and if imperativalism is false, then we do not address reason's practicality at the fundamental level by addressing the question of judgment internalism.

Now the focus will be an investigation of its truth. Fair enough. But is our interest in the error constraint limited to resolving ambiguity and defending the intelligibility of instrumentalism? It is not. When we reach for the error constraint in the first place, even before the various renderings are set out, we do so as a way of bringing some substance to the distinction between occupying a position in the space of reasons and occupying a position in the realm of law, to borrow a suggestive remark of John McDowell's. The several renderings present us with alternatives, and, I will argue, the subsequent decision about which to endorse ultimately restricts how we can understand the nature and authority of principles in the first place. Our reflection on the possibility of error will, thus, turn out to be a partial meditation on the metaphysics of practical normativity. As a first step, in this section we try to understand the allure of imperativalism and ultimately argue that it is not forced on us by argument.

Let's begin by describing a couple of obstacles to supplying a proper defense of imperativalism. We have already encountered one, the ambiguity of the error constraint itself. Inasmuch as it is easy to conflate the logical and imperatival interpretations, it is easy to transfer illicitly the obviousness of the former to the latter.[29] There is another, more elusive danger—conflating a discussion of what an imperative is with an argument on behalf of the imperatival interpretation. Consider Korsgaard's most robust formulation of imperativalism:

> [Kant] does not explicitly give up the view that the will's imperfection is what makes us subject to an *ought*, but it seems to me that he should have, for imperfection is a red herring here. Even a perfectly rational will cannot be conceived as *guided* by reason unless it is conceived as capable of resisting reason. It may be true, as Kant insists, that a divine will is not subject to temptation and so just would do what reason requires, but it is not true, as he seems to infer, that no *ought* applies to the divine will. . . . Obviously, one of the central ideas of this essay is that we can be subject to normative principles only if we can resist them, because without that possibility they cannot function as guides. But I do not agree with Kant that the absence of any specific temptation to resist them removes the possibility of resistance in the sense needed for normativity. It is not imperfection which places us under rational norms, but rather freedom, which brings with it the needed possibility of resistance to as well as of compliance with those norms.[30]

29. John Searle employs strong imperativalism as a premise in an argument for a certain conception of freedom, what he calls "the gap," not worrying that it might itself be subject to challenge. See Searle, pp. 16–17, 66–67.

30. Korsgaard, "The Normativity of Instrumental Reason," p. 240.

Earlier I mentioned that it is plainly and trivially true that an agent is under an imperative only if it is possible for that agent to violate it. If an agent under an imperative is the topic of this passage—if Korsgaard has in mind *imperatives* and the kind of direction they provide when using expressions like "normative principles," "rational norms," "norms," "guides," and "normativity"—then all of what is said would be plainly and trivially true. Clearly, it is a condition of adequacy on any defense of the imperatival interpretation that it not hang on a simple and straightforward appeal to concepts such as these, that is, concepts which might equally be used to supply the definition of an imperative. This remark on terminology is not intended to be a criticism of imperativalism but only to show how little work can be done on its behalf by a straightforward appeal to norms, guidance, et cetera.[31]

The allure of imperativalism does not vanish altogether when the relevant ambiguities of the error constraint are resolved or when the threat of mere stipulation is revealed. And we should not expect it to—the attraction of the thought runs deep. But where does it run to? What is the substantive consideration underwriting the imperatival interpretation? It is, I think, something like the following. No mere mechanism can be under a principle or follow a rule; such a thing might only be a locus of mere regularity. But a would-be perfectly rational will—something whose will is in a state of perfection and thus which can't really go wrong—must amount to no more than a strange sort of mechanism or automaton.

Korsgaard expresses this thought when she says that a "perfectly rational will" is something "whose own conduct is not guided by normative principles at all but, instead, describable in a set of logical truths."[32] The claim is that principles can be no more than "merely descriptive" of a perfectly rational will, just as physical laws are no more than "merely descriptive" of the trajectories of asteroids and planets. This chain of ideas also finds expression in John Searle's recent monograph, *Rationality in Action*: "In order to behave rationally I can do so only if I am free to make any of a number of possible choices and have

31. Korsgaard's own writings do not indicate that she is entirely clear about the risk. There are many passages that suggest that she does not distinguish principles (rules, norms, standards, etc.) from imperatives ("The Normativity of Instrumental Reason," pp. 217, 236). She takes normativity to be a species of necessitation and suggests that necessitation is a component of any adequate account of how "reasons direct, guide, or obligate us to act or judge in certain ways" (*The Sources of Normativity*, p. 226). Recall that for Kant, necessitation is a feature of an imperfectly rational being's following a rational principle, and yet, she says that Kant exhibits confusion in holding that a perfectly rational will is both under laws of reason and not necessitated to follow them ("The Normativity of Instrumental Reason," p. 239).

32. Korsgaard, "The Normativity of Instrumental Reason," p. 240.

open the possibility of behaving irrationally. Paradoxically, the alleged ideal of a perfectly rational machine, the computer, is not an example of rationality at all, because a computer is outside the scope of rationality altogether."[33] One might deny that a mechanism can be under a principle for any number of reasons, though what matters here is that Searle seems to deny that there might be a "perfectly rational machine" on grounds that equally rule out the possibility of a "perfectly rational animal" or a "perfectly rational spirit" or whatever, namely, the absence of "the possibility of behaving irrationally." The threat seems to be that at the limit of perfection the agent goes out of view, and the purported solution seems to be that only with the imperatival interpretation can we distinguish the agent-principle connection and the mere object–physical law connection. Let's investigate this.

All sides agree that the way in which agents are directed by principles is radically different from the way in which mere bits of stuff are directed by physical laws. The distinction between these two sorts of "being directed" is often marked as that between active direction and passive direction, activity and passivity. Korsgaard puts the point this way: "It is in the nature of activities, as opposed to mechanical processes, that one who engages in them is self-guided"; "The rationality of action depends on the way in which the person's own mental activity is involved in its production, not just on its accidental conformity to some external standard."[34] Let's call the thought expressed in these passages the *participation requirement* and formulate it as follows: the efficaciousness of a principle must be mediated by the thought of what follows it. According to the participation requirement, the relevance of a principle to what happens in the world is mediated by the conceptual activity of that which is under it. The contrast is, of course, with the rather unmediated grip that a physical law has on what is, in some other sense, subject to it. With the advocates of imperativalism, let's grant that the participation requirement must be met by any adequate substantive account of the way in which agents are related to principles.

The imperativalist seems to picture matters this way: there is not enough space or enough of a gap between a would-be perfectly rational agent and a principle for her to be said to act on a conception of it. But if the connection between a principle and what an individual actually does is not mediated by a conception of the principle, then what the individual does cannot be put down to an agent's activity or participation. What it does must instead be merely the result of the play of mechanical forces. The imperativalist might continue: even if there is

33. Searle, p. 66.

34. Korsgaard, *The Sources of Normativity*, p. 236, and "The Normativity of Instrumental Reason," p. 236.

some space for self-consciousness or reflection in such a thing, it would be practically impotent. The relevant powers of reflection could place such beings merely in the position of "some leaves blown about by the wind and saying 'Now I'll go this way . . . now I'll go that way' as the wind blew them."[35]

This chain of ideas gives us a somewhat clearer picture of what it is about would-be perfectly rational wills that purportedly renders them unfit for genuine practically rational activity, though a link is missing. We still want to understand why there isn't enough of a gap between a would-be perfectly rational agent and a principle for her to act on a conception of it. The thought linking unswerving accord with a principle and mere mechanism seems to be something like the following: if X always does A in C, then the explanation of why X did A in C can't be an explanation by reasons. But this goes undefended.[36] And I am now simply asking why we should expect the *that* of unswerving accord to travel together with the *because* of mere mechanism—that connection is not baldly contained in the participation requirement.[37]

35. The quotation is from Ludwig Wittgenstein, "Lectures on Freedom of the Will," as cited in G. E. M. Anscombe, *Intention* (Cambridge, Mass.: Harvard University Press, 2000), p. 6. Kant describes a similar case in the teleological argument of the opening pages of the *Groundwork*: "And if reason should have been given, over and above, to this favored creature, it must have served it only to contemplate the fortunate constitution of its nature, to admire this, to delight in it, and to be grateful for it to the beneficent cause, but not to submit its faculty of desire to that weak and deceptive guidance and meddle with nature's purpose. In a word, nature would have taken care that reason should not break forth into *practical use* and have the presumption, with its weak insight, to think out for itself a plan for happiness and for the means of attaining it. Nature would have taken upon itself the choice not only of ends but also of means and, with wise foresight, would have entrusted them both simply to instinct" (*Groundwork*, 4:395).

36. The slide is easy to make. In the course of elaborating his earlier remark about the impossibility of a "perfectly rational machine," Searle says, "A computer is neither rational nor irrational because its behavior is entirely determined by its program and the structure of its hardware. The only sense in which a computer can be said to be rational is observer-relative" (pp. 66–67). Here Searle shifts from complaining about the absence of "the possibility of behaving irrationally" to complaining that the behavior is entirely determined by nonrational elements—very different grounds for holding that something is "outside the scope of rationality altogether."

37. It might help to forestall misunderstanding by begging the question for a moment. As I have it, a perfectly rational agent does what it does through the exercise of practical judgment. One way to start to spell this out is to say that a perfectly rational agent can ask and answer "Why?" questions as, say, these are characterized by Anscombe in *Intention*. In querying whether the possibility of error is essential to agency, I do not mean to query whether the possibility of reflection on one's reasons is essential to agency as this is exhibited in the asking and answering of "Why?" The ability to ask and answer such questions is sometimes linked to the ability to construct philosophical theories of practical reason. According to Korsgaard, ordinary reasoning and philosophical reasoning are absolutely continuous: "A person who starts out reasoning in some perfectly ordinary way . . . finds himself on a route that has no natural stopping place short of the unconditioned

At this point, the imperativalist might challenge my formulation of the participation requirement. What is needed to capture the special sort of contribution that agents make to what they do, the imperativalist might insist, is a strengthening, perhaps along the following lines: only if an agent can determine for herself whether to comply with such principles as she grasps will her behavior be guided by her and be a result of her activity as an agent. Other strengthenings are no doubt possible.[38] For our purposes, the question is whether any such strengthening can remedy the imperativalist's argument. I doubt that such a move will help.

Either the liability to error is packed into a proposed strengthening or it is not. If it is not, then prima facie it will be possible to place a wedge between "participation" and potential error, as I did above. The subsequent burden would then fall on the imperativalist to defend the connection between unswerving accord and mere mechanism, as I suggested above. Of course this would have to be done on a case by case basis. If the liability to error is packed into a proposed strengthening, then obviously the participation requirement, so construed, entails imperativalism. However, it would then no longer be possible to take it as common ground from which to argue for imperativalism. We have not then, I think, been forced by argument to adopt the imperatival interpretation, though no compelling reason has been given to reject it either.

VI. IMPERATIVALISM AND THE LIBERTY OF INDIFFERENCE

How, then, can we make any progress toward resolving the question of imperativalism? As I have indicated, my plan is to investigate the way in which imperativalism constrains our understanding of the nature and authority of principles. To do that it will help to first spell out two

---

Ideas of Reason and the metaphysical perplexities to which they sometimes lead" ("Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind," *Ethics* 109 [1998]: 49–66, p. 61). Kant seems to see matters along similar lines: the principle of morality "is really an obscurely thought *metaphysics* that is inherent in every human being because of his rational predisposition" (*Metaphysics of Morals*, 6:376). Of course, there are other, very different ideas about the degree of continuity of ordinary and philosophical reflection. I raise this immensely difficult and important topic merely to say that in urging us away from imperativalism, I do not mean to urge us away from or toward any particular view about the degree of continuity of ordinary deliberation and philosophical reflection. If such reflection is simply a continuation of ordinary deliberation, then the capacity for such reflection will be a noncontingent feature of being under a principle. If it is not, then it is much harder to see how a case for its noncontingency might be made out. In any case, these questions are simply left open by my investigation of practical reason and the possibility of error. I am grateful to an anonymous referee for helping me to see the risk of conflating these issues.

38. I owe the instructive challenge as well as the specific suggestion to an anonymous referee.

implications of imperativalism: in this section I argue that imperativalism entails the liberty of indifference conception of freedom, and in the next that it entails the impossibility of a perfectly rational agent.

At the end of the long passage quoted at the beginning of the previous section, Korsgaard says that it is by appeal to the concept of freedom that we understand what it is to be under a rational principle: "I do not agree with Kant that the absence of any specific temptation to resist them [normative principles] removes the possibility of resistance in the sense needed for normativity. It is not imperfection which places us under rational norms, but rather freedom, which brings with it the needed possibility of resistance to as well as of compliance with those norms."[39] Korsgaard is careful to deny that being subject to temptation or felt desire can be the source of the possibility of resistance in the sense she requires; this is because being a subject of felt desire is *not*, while being able to resist *is*, part of what it is to be an agent under a principle, the former can't account for the latter. This complaint with Kant's view of the matter reveals the way in which Korsgaard writes the possibility of resistance into the very idea of an agent under a principle. This is also revealed by her treatment of the source of resistance as itself a capacity or ability: "Even a perfectly rational will cannot be conceived as *guided* by reason unless it is conceived as capable of resisting reason."[40] What, then, are we to make of this freedom which "places us under rational norms"?

Korsgaard is sympathetic to Kant's notion of negative freedom: "Nothing in human life is more real than the fact that we must make our decisions and choices 'under the idea of freedom.' When desire bids, we can indeed take it or leave it."[41] However, the sort of freedom involved in accepting the imperatival interpretation must be quite different; it must be far more radical than freedom from determination by felt desire. Indeed, to be free in this more radical sense seems to involve both the capacity to make a choice for any principle and the capacity to make a choice against any principle. That is, accepting imperativalism on the ground that the freedom of a rational agent brings with it the needed possibility of resistance to and compliance with rational standards commits one to such a radical conception of freedom,

39. Korsgaard, "The Normativity of Instrumental Reason," p. 240.

40. Ibid. At certain points Korsgaard takes herself to be involved in a dispute with Kant about how to understand the capacity to resist: "Kant apparently identified our capacity to resist the dictates of reason with the imperfection of the human will" ("The Normativity of Instrumental Reason," p. 239). I think it is unlikely, however, that Kant would accept that there is such a capacity, ability, or power. See n. 48 below.

41. Korsgaard, *The Sources of Normativity*, p. 97.

which, following the tradition, we might call *libertas indifferentiae* or the liberty of indifference.[42]

It might be useful to locate the imperativalist in an exchange between Kant and Carl Leonhard Reinhold over closely related, though somewhat narrower, matters.[43] Reinhold criticizes the core of Kant's project—the inseparability of freedom and practical reason—on the grounds that it makes it inconceivable how one could freely violate the moral law. Then with a view to accommodating the fact of free and immoral action, Reinhold adopts the position that freedom is the capacity to determine oneself either in accordance with, or contrary to, reason. Kant responds:

> But freedom of choice cannot be defined—as some have tried to define it—as the ability to make a choice for or against the law (*libertas indifferentiae*), even though choice as a *phenomenon* provides frequent examples of this in experience. For we know freedom (as it first becomes manifest to us through the moral law) only as a *negative* property in us, namely that of not being *necessitated* to act through any sensible determining grounds. But we cannot present *theoretically* freedom as a *noumenon*, that is, freedom regarded as the ability of the human being merely as an intelligence, and

42. In the context of a distinct but closely related inquiry—one about the nature of "moral responsibility" rather than practical rule following—Susan Wolf describes a kind of agent that is free in this very sense: "The autonomous agent must be one who is able to act in accordance with Reason *or not*. That is, she must be able to regard the rational course of action, insofar as there is one, as just one alternative among others. . . . This ability to choose among the rational, irrational, and nonrational alternatives alike is not an ability to choose on some higher-than-rational basis. Rather, it is an ability to choose on no basis whatsoever, an ability, if you will, to choose whether to use any basis for (subsequent) choice at all" (*Freedom within Reason* [New York: Oxford University Press, 1990], p. 54). While I object to Wolf's attribution of this position to Kant—*Metaphysics of Morals*, 6:226–27, is decisive evidence that he does not hold it—I accept her characterization of just how radical it is. In the course of considering various arguments for the claim that something must be radically "autonomous" in order for it to be a possible locus of "moral responsibility," there is one which Wolf raises only to leave to the side: "A second possibility is that those who continue to insist that radical autonomy is necessary for responsibility do so not because they disagree with my view that the ability to act in accordance with Reason is sufficient for responsibility but because they think that this ability itself requires at least a kind of radical autonomy" (*Freedom within Reason*, pp. 61–62). The imperativalist, I think, might then be seen as arguing for "autonomism" on similar grounds, and this essay might be seen as a contribution to Wolf's defense of what she calls the Reason View.

43. The matters are somewhat narrower because Reinhold and Kant reserve the concept of freedom for characterizing a being with distinctively moral capacities, while in our discussion a being with any practically rational capacity is free or active in some sense. The matters are closely related because in every case we are concerned with the conditions of being under rational norms. It is just that they reserve the concept of freedom for characterizing the species or form of activity characteristic of beings with moral capacities.

> show how it can exercise constraint upon his sensible choice; we cannot therefore present freedom as a positive property. But we can indeed see that, although experience shows that the human being as a sensible being is able to choose in opposition to as well as in conformity with the law, his freedom as an intelligible being cannot be defined by this, since appearances cannot make any supersensible object (such as free choice) understandable.[44]

Since freedom is not knowable through experience, and since error is knowable only through experience, the latter cannot be a component of the definition of the former. The problem of imputable errors of reason must be solved in another way than by prying apart freedom and practical reason.[45] However, accommodating the mere fact of error, explaining what we encounter in experience, cannot be what moves the imperativalist to hold that "the possibility of self-government essentially involves the possibility of its failure."[46] Indeed, we can treat imperativalism as a later stage in the development of the Kant-Reinhold dispute. According to imperativalism, even if nobody has ever, in fact, acted out of accord with what reason prescribes, we would still be in possession of the same materials for insisting on the imperatival interpretation. This is because, as the imperativalist thinks, the liability to error is contained in the very idea of a free and rational agent.

VII. IMPERATIVALISM AND PERFECTLY RATIONAL AGENCY

According to the imperatival interpretation, the idea of a Kantian divine or holy will under, say, the principle of promise keeping is as absurd as the idea of an *arbitrium brutum* or mere animal will under the same. Neither sort of being can really go wrong with respect to this principle, and so, neither can be conceived as genuinely being under it as a principle. More generally, imperativalism implies that it is impossible for an agent to be essentially infallible or perfectly rational by nature. (Thus, imperativalism offers a novel proof of atheism, on the assumption that God is an agent and essentially infallible, certainly the position of classical theology.)

44. Kant, *Metaphysics of Morals*, 6:226.

45. For discussion of the central importance of this passage in Kant's practical philosophy, see Henry Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), pp. 129–36, and also see Allen Wood's essay "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy*, ed. Allen W. Wood (Ithaca, N.Y.: Cornell University Press, 1984), pp. 73–101, pp. 79–83. Notably the passage also figures in Korsgaard's own sympathetic interpretation of Kant in "Morality as Freedom," in *Creating the Kingdom of Ends*, pp. 159–87. This suggests that she will resist the attribution of the liberty of indifference conception of freedom which, of course, is fine, though I do not see how she can distance herself while continuing to maintain imperativalism. I return to this tension in Sec. VIII.

46. Korsgaard, "The Normativity of Instrumental Reason," p. 248.

However, even though the imperatival interpretation implies the incoherence of the very idea of an essentially infallible agent, Korsgaard sometimes seems averse to losing access to this concept, no doubt because she is averse to losing contact with the tradition of practical philosophy she has done much to help us understand. She insists, against Kant, that a "divine will" and a "holy will" are also capable of resisting reason and therefore subject to "oughts."[47] But this is very confusing. On the face of it, we can't make any sense of the idea of an essentially infallible agent that is even capable of resistance.[48] If X is capable of doing A, then it is possible for X to do A. So, if something is capable of resisting reason, then it is possible for it to resist reason. But it is simply not possible for an essentially infallible agent to resist reason. Thus, it is not true that an essentially infallible agent is capable of resisting reason. Korsgaard admits that "a divine will is not subject to temptation and so just would do what reason requires," but this isn't an adequate gloss of the excellence characteristic of a divine or holy will.[49] Instead, the sort of thing Korsgaard is describing here might be found in an Oliver Sacks anecdote: a human being who has lost any capacity to experience pleasure and felt desire. There is a significant difference between the idea of an agent that, as it happens, is not going to go wrong and the idea of an agent that cannot, under any circumstances, go wrong. Only the second is essentially infallible and it cannot also be capable of resistance.

Now, in Kant's system the holy will functions as "a practical *idea*, which must necessarily serve as a *model* to which all finite rational beings can only approximate without end and which the pure moral law, itself called holy because of this, constantly and rightly holds before their eyes."[50] Although the imperativalist cannot consistently let a perfectly rational being serve as an ideal, perhaps she has the resources to articulate another, more plausible, more terrestrial, conception of a perfectly rational agent; and perhaps this might serve as an ideal actually attainable in the life of that which operates with it. In the tradition, this is

47. Ibid., p. 240. See the passage quoted at length at the beginning of Sec. V above.

48. Although I am using the expression "capacity to resist," I think we are in some sense already on the wrong track when doing so. A capacity, ability, or power is, I would like to say, always something, in some sense, good. Indeed, one way of putting the underlying sentiment of this essay is that the idea of a capacity or power to resist reason is a confusion on the order of the idea of a capacity not to see, or the treatment of blindness as itself a capacity. See Wood's "Kant's Compatibilism," pp. 81–82, for an interesting discussion of this sense of "capacity." For the parallel between irrationality and blindness, see Saint Anselm's "The Fall of Satan," in *Truth, Freedom, and Evil: Three Philosophical Dialogues*, ed. and trans. Jasper Hopkins and Herbert Richardson (New York: Harper & Row, 1967), sec. 11.

49. Korsgaard, "The Normativity of Instrumental Reason," p. 240.

50. Kant, *Critique of Practical Reason*, 5:32.

done by shifting attention away from characterizing a kind of being—what is perfectly rational by nature, and what I have been calling essentially infallible—and toward characterizing a state of will, a disposition, *habitus*, or *hexis*. I will now argue that even the more modest conception of perfect rationality—as a state of will nonaccidentally issuing only in correct action, though only contingently possessed by its bearer—is unavailable to the imperativalist.

Any characterization of such a practical disposition must permit the distinction between a perfectly rationally ordered state of will and other states of will, some of which, nevertheless, often give rise to similar actions in similar circumstances. In other words, any characterization must permit the distinction between perfect rationality and its imitation. But, I will now argue, this can't be done without the use of materials inaccessible to imperativalism. If X is in a certain state and has not, in fact, gone wrong it does not follow that X is in an ideal state—X might have the unblemished record by accident. If X is in a certain state and will not, in fact, go wrong it does not follow that X is in an ideal state—X might acquire the unblemished record by accident. How, then, are we to distinguish the agent in an ideal state from the merely lucky agent, if not according to the totality of past, present, and future performance? Actual success doesn't guarantee counterfactual success, and it is the latter which is needed to distinguish the ideal from the lucky. In order to distinguish perfect rationality from its imitation, it seems that one has to say something like this: if X is in an ideal state of will, then it must be no accident that when X acts, X acts correctly. But can the imperativalist consistently say this? The imperativalist must accept that an agent is liable to error even when in possession of the ideal state of will. That is, within the imperativalist framework, given some agent X which is in the "ideal" state, there must be some circumstance in which X's power to err would be exercised. Were this denied it would be quite dubious whether the capacity for resistance had any explanatory role whatsoever. If there is some circumstance in which X, such as he is, would go wrong, and if it is only an accident that X does not find himself there, then for any state of will that X is in, it must be an accident that X does not go wrong. And so, the imperativalist is not entitled to a conception of an ideal agent as what has a perfectly rational state of will.[51]

Suppose I am right that the imperativalist is not entitled to either conception of a perfectly rational agent—as essentially infallible or only as possessing a flawless state of will. Should she care? Were we of a

51. I have not shown that these are the only terms in which to develop a conception of an ideal agent. So, strictly speaking, I have not proved that no conception of the ideal agent is available to the imperativalist. But what are the alternatives?

different era we might press her in something like the manner of Saint Anselm: "I don't think that freedom of choice is the ability either to sin or not to sin. Indeed, if this were its definition, then neither God nor the angels who are not able to sin would have free choice. But to say that they have no free choice is blasphemous."[52] Since we cannot, these days, get any purchase on our adversaries through accusations of blasphemy, we must resort to more subtle and less decisive maneuvers.

VIII. THE INCOMPATIBILITY OF IMPERATIVALISM AND CONSTITUTIVISM

Sections VI and VII articulated implications of imperativalism, respectively, the liberty of indifference conception of freedom and the impossibility of perfectly rational agency. Drawing on these results, this final section establishes the incompatibility of imperativalism and *constitutivism*—a strategy for understanding that in virtue of which an agent is under a principle—thereby bringing the question of imperativalism within the scope of problems about the metaphysics of practical normativity quite generally.

The problem of the normative or binding force of a principle is familiar enough, I suppose; still it might be worth trying to say a little bit more in order to provide some context for a description of the constitutivist position. On any account, if an agent is under a principle, then its performance of some action is recommended or sided with. This is just to make the benign point that principles are standards of correctness, something we can read right off the dictionary definition of "principle," whatever that's worth. Now, when an action is sided with we can say so explicitly by employing deontic vocabulary as in "X ought to do A," "X may do B," or "Everyone must do B in C." So, if an agent is under a principle, then some deontic claim with it as subject is true, though I will concentrate on "oughts." Constitutivism is, then, a reply to "In what does the binding force of a principle consist?" or again to "What is it for an action to be something an agent ought to do?"

Let's narrow this question a bit: in what does the binding force of a formal principle consist? As I will understand it here, the constitutivist finds a way in through reflection on the nature of agency considered as such. Here is Korsgaard's helpful formulation of the position:

> There are in our tradition two things which philosophers have meant by 'reason'. Reason refers to the active as opposed to the passive capacities of the human mind, and 'reason' also refers to certain sets of principles—logical principles, moral and other practical principles, and the principles that Kant associates with the pure concepts of the understanding. What Kant did . . . was to

52. Anselm, p. 122.

> try to bring these two conceptions of reason together: to explain the normative force of the principles by showing that they are constitutive of reflective mental activity itself. To choose *is* to follow the hypothetical and categorical imperatives; to understand *is* to employ the concepts and principles of the understanding, and so on. And in the same way, my own aim is to portray moral principles [and principles of practical reason generally] as constitutive of, and so as essential to, making human choices, and leading a human life.[53]

Let me try to rephrase this. The constitutivist provides an answer to the metaphysical question "What is agency?" with a view to elucidating the binding force of formal principles. Very generally, the constitutivist's objective is to demonstrate that any particular exercise of rational agency, or action, involves a commitment to comply with certain principles. The objective is to be achieved by showing that the relevant principle is simply a partial description of agency itself; that is, the aim is to be achieved by showing that the capacity to act just is, in part, the capacity to follow the relevant principle. If so, then when one does exercise that capacity one is committed to abiding by the relevant principle—it is a standard internal to the capacity. Obviously the greatest difficulty for any constitutivist account is in showing that there are, in fact, standards essentially contained in such spare materials as the bare capacity for action. But skepticism about that aside, we can grant that were such an analysis correct, there would be no way for an agent to get far enough outside the relevant standards to doubt that it is under them: indeed, no such way for any agent, necessarily.[54]

So, constitutivism is an attempt to extract binding principles from the nature of agency; imperativalism is a thesis about the nature of

53. Korsgaard, *The Sources of Normativity*, p. 236. Also see "The Normativity of Instrumental Reason," p. 243, "Motivation, Metaphysics, and the Value of the Self," p. 65, "Self-Constitution in the Ethics of Plato and Kant," *Journal of Ethics* 3 (1999): 1–29, esp. pp. 12–15, and her unpublished "Self-Constitution: Action, Identity and Integrity," delivered as the John Locke Lectures (Oxford University, 2002), which develops this theme in great detail.

54. What I am calling constitutivism has been explored recently, though in very different ways and under different headings, by others: Stephen Darwall, *Impartial Reason* (Ithaca, N.Y.: Cornell University Press, 1983); David Gauthier, "The Unity of Reason: A Subversive Reinterpretation of Kant," *Ethics* 96 (1985): 74–88; Thomas Nagel, *The Possibility of Altruism* (Princeton, N.J.: Princeton University Press, 1970); Railton; David Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000). For the constitutivist, the class of formal principles contains those which are fundamental and those which are derivative, i.e., arrived at through the application of fundamental principles. The constitutivist is fixed on the fundamental principles, but with an interest in understanding the basis of those derived from it. It is important to note here that if the fundamental principles are not internal to agency, then neither are any of the principles derived from it.

agency itself. The question is whether an explanatory strategy like constitutivism can be employed while conceiving of agency in such terms. Intuitively the answer is no: in trying to understand what agents ought to do, constitutivism draws us toward what agency is, whereas imperativalism forces us away by insisting that agents cannot be as they ought.

Constitutivism begins from a conception of the capacity to act and a conception of action. Imperativalism insists that to be an agent is, in part, to possess a capacity to act out of accord with any principle whatsoever; and so to be an agent is, in part, to possess the capacity to act out of accord with any formal principle. Let's hold that fixed for the moment and imagine someone trying to apply the constitutivist's argument schema: the capacity to act just is, in part, the capacity to follow some principle, say, the instrumental principle. And so, when one does exercise that capacity one is thereby committed to abiding by the relevant principle, say, the instrumental principle. The difficulty is, I am sure, obvious. Given imperativalism, it is also true that the capacity to act just is, in part, the capacity to violate some principle, say, the instrumental principle. But if it is constitutive of agency both to follow and to violate a principle, then we can no longer derive an intelligible commitment simply to follow from the nature of agency itself.

The problem is that a constitutivist account of the authority of formal principles must make use of a conception of a perfectly rational agent. It must begin from there, using the ideal to fix the standard constitutive of the activity in question—treating deviations from the relevant standard as we might treat, say, blindness in a human being, an accident, and not as an expression of what rational agency is. Yet the imperativalist cannot help herself to that very starting point. Trying to combine imperativalism and constitutivism is like trying to ground one's moral philosophy in a conception of human nature while maintaining the doctrine of original sin.

Here's another perspective on the tension. Recall the way reason, freedom, and the possibility of error are linked in the defense of imperativalism canvassed in Section V. There we were told that it is by appeal to the concept of freedom that we understand what it is to be under a principle, reminding us that freedom here includes the capacity to resist the guidance of these. But, as I argued above, given the requirement of the possibility of resistance, this freedom amounts to the liberty of indifference. Given this conception of freedom, one exercises one's will in accord with its nature as free or active no matter whether one complies with or violates any principle. It seems that we lose all hope of explaining an agent's subjection to a principle by appeal to the active nature of the capacity whose exercise is action.

So, for this reason too, an imperativalist is prohibited from elucidating the authority of a formal principle, or explaining the commit-

ment to comply with it, by revealing it to be constitutive of reflective mental activity, or again, internal to agency or acting considered as such. Trying to combine imperativalism and constitutivism is like trying to maintain a voluntarism according to which a fiat of God's unconstrained will creates morality, while also holding that God's will drafts him into the moral order.

What is the relevance of the incompatibility to the truth of imperativalism? How far does the incompatibility take us toward deciding the question of whether part of being under a formal principle is being able to err with respect to it? Obviously this is not settled by the incompatibility argument itself. It is important to the assessment of the incompatibility that the chief nonreductive and realist alternative to constitutivism is platonism.[55] So it might be of use to say something very brief about that first.

According to platonism, we reach the explanatory bedrock in our understanding of being under a principle with explicitly normative facts themselves. The idea is that these make up a practical normative order that is prior to and independent of the subjectivity of what is under it. On this sort of view, a principle sides with action because a principle is simply a fact like "Everyone ought to do A" or "X should do B in C." For platonists, ultimately there is no explanation to be given of facts of this kind, even if there is much to say about their content. Now, to account for the practical efficacy of the grasp of a principle, platonists have sometimes maintained that unlike other such objects of knowledge, the relevant facts have intrinsic to-be-pursuedness and are such as to elicit movements of the will when grasped; but more recently, advocates of platonism have tended to postulate a background disposition to act in accord with judgments about what one ought to do. This depiction is very spare but still portrays something familiar, and, I think, something connected with great difficulties.

Now on the assumption that platonism and constitutivism are exhaustive of the nonreductive and realist alternatives for understanding the authority of such principles, the question of imperativalism poses a stark choice. If one opts for imperativalism, one is forced either to adopt

55. A nonreductive realist, in my sense, holds that normative claims are truth-apt (against noncognitivism) and sometimes true (against an error theory) and denies that such truths are reducible to nonnormative truths. That is, a nonreductive realist in my sense holds that there are facts about what an agent ought to do and that these are not reducible to nonnormative facts. I have not argued that constitutivism is the only nonreductive and realist alternative to platonism but am following a plausible and well-developed tradition of carving up the possibilities, at least so far as formal principles go. Both Korsgaard and Derek Parfit seem to work with this picture of the options, though Parfit leans in the direction of platonism. See his "Reasons and Motivation," *Proceedings of the Aristotelian Society* 71, suppl. (1997): 99–130, esp. pp. 107–9.

a mysterious platonism about the nature and authority of formal principles, or else to give up on the idea that there really is such a thing as being under one. Whether imperativalism particularized to formal principles is false will largely depend on whether platonism is as hopeless, and constitutivism as promising, as Korsgaard and others have argued, but that must be pursued elsewhere.[56]

I want to conclude by noting another limitation of our results and advertising a direction for further work. Even if imperativalism is incompatible with constitutivism particularized to formal principles, we are still far from having forged a lever with which to unsettle imperativalism in general. Recall that when raising the question of imperativalism, I asked not only whether the liability to error is a condition of being under a formal principle, but also, whether the liability to error is a condition of being under more ordinary, substantive principles like those of baseball, etiquette, and the laws of Pennsylvania, which do not have the same jurisdiction. But understanding what, if anything, it is to

56. But is constitutivism so promising? I should mention that many think that it inevitably runs into a terrible difficulty, one recently raised by R. Jay Wallace for Korsgaard ("Normativity, Commitment, and Instrumental Reason," sec. 1) and also raised by Philip Clark for David Velleman ("Velleman's Autonomism," *Ethics* 111 [2001]: 580–93). To illustrate the problem, let's look at Clark's essay. There he criticizes Velleman's attempt to argue from a conception of the nature of action as an activity whose constitutive aim is autonomy to a conception of the standards governing action, on the following grounds: "Autonomy, as Velleman conceives it, is a goal that is achieved in every fully intentional action and so cannot serve as the standard of rational assessment for action" (p. 593). Velleman has subsequently expressed concern that there is a real problem here (Velleman, p. 30). And one might think that the difficulty is so forbidding that one should prefer imperativalism to having to resolve it. But what exactly is the source of the difficulty? Even with so much interesting work being done in the constitutivist tradition, we still lack, I think, a correct conception of the schema for such a position. In particular, we lack a correct conception of the logical form of the claims describing the essence or nature of agency, the claims in virtue of which we are supposed to understand the force of "oughts" applying to particular agents. The default tendency is to treat the description of essences as universal generalizations in which all members of the relevant kind are said to possess a property. Indeed, the difficulty arises for Velleman just when his description of the nature of action, i.e., his characterization of action's constitutive aim, is taken to have the underlying form of a universal generalization in which anything that is an action has the property of being autonomous. My thought is that the difficulty is not for the general constitutivist attempt to, as it were, understand norms in terms of natures, but only for a certain conception of the form of descriptions of natures or essences. A promising direction for the constitutivist to go, I think, is to resist the urge to assimilate such descriptions to universal generalizations and instead look toward generics to describe "the what it is" which is to serve to underwrite standards of assessment. For recent work in this direction, see the following: Julius Moravcsik, "Essences, Powers, and Generic Propositions," in *Unity, Identity and Explanation in Aristotle's Metaphysics*, ed. T. Scaltsas, D. Charles, and M. L. Gill (Oxford: Clarendon, 2000), pp. 229–44; Michael Thompson, "The Representation of Life," in *Virtues and Reasons: Philippa Foot and Moral Theory*, ed. Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn (Oxford: Clarendon, 1995), pp. 247–96.

be under one of the latter is not within the reach of constitutivism, as we have considered it. And so the incompatibility argument just given is not directly relevant to the question of imperativalism in these more ordinary cases. I come to this topic with a conviction that our dialectic can be developed even when particularized to substantive principles. But whether an attempt to extract substantive principles from the nature of what is subject to them—an attempt of the kind John Rawls makes for the practice of baseball in "Two Concepts of Rules" and Philippa Foot makes in her recent *Natural Goodness* for the human life-form—is also obstructed by imperativalism must be pursued elsewhere.

Copyright of Ethics is the property of University of Chicago Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.