



## Internalism and Agency

Stephen L. Darwall

*Philosophical Perspectives*, Vol. 6, Ethics (1992), 155-174.

Stable URL:

<http://links.jstor.org/sici?sici=1520-8583%281992%296%3C155%3AIAA%3E2.0.CO%3B2-T>

*Philosophical Perspectives* is currently published by Blackwell Publishing.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/black.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## INTERNALISM AND AGENCY

Stephen L. Darwall  
University of Michigan

As interest has refocused in recent years on fundamental questions of meta-ethics, a group of loosely-related ideas collectively referred to as *internalism* have come in for increasing attention and controversy. A good example would be recent debates about moral realism where question of the relation between ethics (or ethical judgment) and the will has come to loom large.<sup>1</sup> Unfortunately, however, the range of positions labelled internalist in ethical writing is bewilderingly large, and only infrequently are important distinctions kept clear.<sup>2</sup> Sometimes writers have in mind the view that sincere *assent* to a moral (or, more generally, an ethical) judgment concerning what one should do is necessarily connected to motivation (actual or dispositional).<sup>3</sup> This necessity may be conceptual, or perhaps metaphysical, the thought being that it is not merely a contingent matter that people have motives to do what they think or sincerely say they should. I call internalism of this variety *judgment internalism* and distinguish it from another set of theses that concern, not what it is to accept an ethical judgment, but what it is for such a judgment to be true.<sup>4</sup> According to *existence internalism*, someone morally (or ethically) ought to do something only if, necessarily, she (*the agent*) has (actually or dispositionally) motives to do so. Again, this necessity might be conceptual or metaphysical, the thought being that it is not merely a contingent matter that agents have motives to act as they ought.

Already, several comments are necessary. It may be wondered, for example, what the point is of adding 'ethical' to 'moral' in these two formulations. Falk is generally read as having originally introduced 'internalism' to refer to a view about the moral 'ought', viz., that, necessarily, an agent morally ought to do something, say *A*, only if she has a motive to *A*.<sup>5</sup> Falk did not distinguish between this claim and the thesis that, necessarily, an agent morally ought to *A* only if there exists *reason* or justification for her to do so. Indeed, the view he had in mind was that a moral 'ought' claim is true only if its grounds are themselves reasons to act, *both* justificational and (at least potentially) motivational.<sup>6</sup> These claims came to the same thing

for him because, evidently, he believed that the existence of (justifying) practical reasons also necessarily depends upon motivation.<sup>7</sup> This is controversial, however. Someone might think that there being reasons to act—indeed, an agent's having such reasons—has no necessary connection to motivation. (Justifying) reasons *for* someone *to* do something can be distinguished from an agent's actual (motivating) reasons for acting, and it can be held that whether there are good or justifying reasons for an agent to do something is completely independent of her motives, actual or hypothetical.<sup>8</sup>

Falk's suppressed premise was an internalism about practical reasons, but it was no less a *meta-ethical* thesis for that.<sup>9</sup> Whether there are reasons to act is a normative and thus, in the broad sense, an ethical matter.<sup>10</sup> It concerns what the agent rationally *ought* to do. Moreover, the rational 'ought' is almost never treated by philosophers as simply one 'ought' among others, on all fours with 'ought's *internal to*, e.g., etiquette, baseball, or bridge. It is regarded, if only implicitly, as *unqualifiedly normative*—not just an ought-according-to-the-norms-of-rationality, as there might be oughts-according-to-the-norms-of-etiquette, or -baseball, or -bridge. What a person rationally ought to do is whatever he ought to do *simpliciter*—*sans phrase*, as it were.<sup>11</sup> It is, with some redundancy, what he ethically ought to do.

Consider in this light the position of those who assert existence internalism of practical reasons, but deny it of morality, and conclude from this that morality has no warranted claim to unqualified normativity. Such a line of thought can impugn the unqualified normativity of the moral 'ought' only if the rational 'ought' has it.

A different picture would be that morality lacks unqualified normativity, but then, so does everything. There are 'ought's of morality, rationality, etiquette, bridge, baseball, and whatever other systems of norms one cares to imagine, but none of these bear on some further issue—what a person should do *simpliciter*; there is no such further issue. This, however, amounts to a form of ethical skepticism. Insofar as it is practical, ethics' fundamental question is what a person should do. Unqualified normativity would appear to be its subject.<sup>12</sup>

Judgment and existence internalisms are always asserted, I believe, of what their proponents *take* to be unqualifiedly normative. Judgment internalists about morality, for example, see their view as recommended by the fact that the (unqualified) normativity of moral judgment cannot be adequately understood except by way of connection to judge's-motivation. And existence internalists about morality reason that the (unqualified) normativity of the moral 'ought' requires agent's-motivation as a necessary condition of its truth.<sup>13</sup> Judgment or existence internalism about morality will seem plausible, I submit, only if one thinks that morality *is* unqualifiedly normative.<sup>14</sup> Thus, internalists about practical reasons who are externalists about morality

also deny that moral 'ought's have unqualified normativity. In the end, internalism is offered as a thesis about unqualified normativity and normative judgment.

Sometimes the view that morality is unqualifiedly normative is simply taken to be a form of internalism.<sup>15</sup> This makes reasons to act internal to morality, but the internalisms we shall be concerned with are views that hold the normative, or normative judgment, to be necessarily related to what is internal to us.

Or again, 'internalism' sometimes refers to a view about motivation—for example, to the view that belief by itself is motivationally inert and that all motivation derives from desire.<sup>16</sup> The issues to which ethical internalism is addressed, however, can be treated somewhat independently of these matters of philosophical psychology. If, for example, ethical judgment is necessarily connected to motivation, the latter might be carried by a desire, or by a distinctive conative state—such as Gibbard's "acceptance of a norm"—or, perhaps, by a peculiar variety of motivating belief.<sup>17</sup> *Ethical* internalism is addressed to the nature of ethics and ethical judgment, and even if issues of philosophical psychology cannot be postponed indefinitely, the initial focus is different.

### **Existence internalism, a preliminary distinction**

Existence internalism is a metaphysical claim. It is the nature of unqualified normativity, it holds, to be necessarily related to motivation. But this general characterization admits of two very different kinds of view. On one, the ethical is such that motivation is necessarily an effect of engaging it epistemically. On the other, motivation is a constituent of ethical facts themselves. The most familiar versions of the former are rational intuitionisms such as Plato's which hold that it is of the nature of the Good that it cannot be fully grasped with indifference. Richard Price held a similar view about moral obligation: "When we are conscious that an action...*ought* to be done, it is not conceivable that we can remain *uninfluenced*, or want a *motive* to action."<sup>18</sup> These may seem to be versions of judgment internalism, but they are not. They say, not that ethical belief or sincere assertion necessarily motivates, but that actual consciousness of or cognitive contact with the ethical does.<sup>19</sup> But Price also stresses that motivation has nothing to do with what ethical facts themselves are. Motive is "the *effect* of obligation perceived, rather than *obligation itself*."<sup>20</sup> This, however, is precisely what views of the second kind assert. Seeking an understanding of the unqualifiedly normative—of what it is for an act to be something an agent ought to do—they look to motivation. They hold that, ultimately, normative force must be understood as a kind of motivational force.

Thus, although both these varieties of existence internalism assert that, necessarily, an agent ought to do something only if, necessarily she would have under certain circumstances a motive to do so, only the latter is a *constitutive internalism*. Only on the latter sort of view does the agent's (dispositional) motivation have anything to do with what makes it the case that he ought to act. Only constitutive internalism, we might say, situates normativity itself *within* agency.

This difference reflects two different ways of looking at practical reason. For rational intuitionists such as Plato and Price, there is no fundamental difference between practical and theoretical reason. Reason is the faculty of cognition, and ethical facts are no less available to it than, say, mathematical ones. The difference consists in the objects of cognition. Unlike other objects of knowledge, the ethical is such that, necessarily, to know it is to be moved. For constitutive internalists, on the other hand, practical reason is not simply theoretical reason applied to the discovery of ethical truths. Rather, as I shall argue, constitutive internalists view practical reason as realized in the practical reasoning that *agents* do. And facts concerning what agents *ought* to do are constituted by motives they can acquire through practical reasoning. The formula "reason can be practical" is thus ambiguous as between two quite different ideas.

Some kind of non-constitutive existence internalism appears to be the crucial premise in J. L. Mackie's well-known argument that ethical properties are too "queer" to exist.<sup>21</sup> If anything really were "objectively valuable," as he puts it, it would have to have an unqualified normativity that would necessarily motivate when grasped. "An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it."<sup>22</sup> But it is mysterious how there could be such a property, precisely because it seems to require a bridge between knowledge, or its object, and *action*. Constitutive internalisms also hold that ethical knowledge necessarily motivates, at least when it is gained by an agent for herself within the deliberative process of practical reasoning and judgment.<sup>23</sup> But that is because, for such views, practical knowledge is *of motives* deriving from the agent's practical reasoning. For non-constitutive internalisms, however, there is no such connection between the *object* of ethical knowledge and agent's-motive. And practical reason, on such views, can refer only to the fact that when the general faculty through which we cognize truth grasps ethical truth, motivation necessarily results. How this connection might be effected nonetheless remains a mystery. There is no necessary connection in general between knowledge and motive, and non-constitutive existence internalists have traditionally eschewed any account of the object of ethical knowledge in particular that would relate it to agent's-motive.

Recently, however, another form of non-constitutive existence internalism has been proposed that promises to relate ethics to motivation in a new way.<sup>24</sup> Inspired partly by the concerns just registered, but also by a resistance to non-cognitivism *and* to constitutive internalism, John McDowell and David Wiggins have articulated a meta-ethical position, which because it aims to model our understanding of ethical properties on that of secondary qualities such as color, we might call *sensibility theory*.<sup>25</sup> Simplifying significantly, sensibility theorists claim (a) that we have distinctively ethical sensibilities, which have affective and motivational aspects; (b) that these sensibilities cannot be understood projectively, but only in terms of ethical properties they implicitly attribute; and (c) that, nonetheless, ethical properties can themselves adequately be understood only by relation to a relevant sensibility, including its motivational aspects.

As an analogy, consider the sense of humor. A person lacking in a sense of humor does not perceive the humor in a situation and simply fail to be amused; sensing humor apparently includes its distinctive affect. But neither can we well understand what it is to *be* humorous independently of amusement and a sense of humor. Sentiment and property seem to require each other in something like the same way color and color sense interdepend.

Sensibility meta-ethical theories make similar claims about the relation between ethical sensibility and ethical properties. Someone lacking moral sense, they claim, will not perceive that a certain situation, say the avoidable suffering of a child, calls for action, but simply fail to be moved. Sensibility theory is thus an existence internalism, but unlike rational intuitionism, it holds motivation not to be an effect of ethical cognition, so much as a component or condition. And contrary to constitutive internalism, sensibility theorists argue that what is cognized is a *sui generis* space of reasons that apply to an agent irrespectively of her motives—not something constituted in any way out of her practical reasoning. An agent suffering no defect of practical rationality might fail to perceive an obligation she was genuinely under simply through lacking the power to be appropriately moved.

Sensibility theories may thus be thought of as non-constitutive existence internalisms that seek to explain an otherwise puzzling connection between moral cognition and motivation by the mutual dependence of moral property and sensibility. It is not clear, however, that this can suffice for an adequate account of unqualified normativity. Even if it is part of moral sensibility that, as McDowell argues, some properties it enables one to discern have the effect, when recognized, of silencing or outweighing other sorts of reasons in practical deliberation, we might imagine this also to be true of other matched sets of sensibility and property. What about the sense of the sacred, for example, or even the diabolical? Just as, according to sensibility theorists, only properly motivated agents can perceive the right, and then do so as authoritative and silencing other considerations, it may be no less true that

only the devout can discern the sacred, and that they discern it as properly regulating conduct. And similarly, perhaps, for the diabolical. So long as the justificatory force of the moral can only be grounded within a moral sensibility that can in principle conflict with other “practical” senses, there seems nothing to give its reasons any weight except, as it were, *within* that sensibility, and certainly nothing to give them greater weight than those tied to any other. But if this is so then sensibility theory is apparently an inadequate account of unqualified normativity.<sup>26</sup>

### Judgment internalism

Judgment and constitutive internalists share the belief that the unqualified normativity which is the hallmark of ethics is best explained by the idea that ethics is an inescapably *practical* subject. Consider Moore’s famous open question argument against any attempt to reduce (what he regarded) the fundamental ethical property, intrinsic goodness, to any complex of naturalistic or metaphysical properties. We can always intelligibly ask of any such complex, Moore argued, whether that complex is itself good. The open question argument has had remarkable staying power, despite the fact that Moore’s formulation has been known to rest on widely rejected assumptions in the philosophies of language and mind.<sup>27</sup> Why is this? Recall that intrinsic goodness for Moore (in *Principia*) has unqualified normativity built into it.<sup>28</sup> Moore alternately formulated an option’s being intrinsically good as its “ought[ing] to exist for its own sake,” and this entails that anyone who can realize it *ought*, other things equal, to do so. We can therefore put the thought behind the open question argument in this way: for any complex of naturalistic or metaphysical properties, it seems we can intelligibly ask of an arbitrary agent in a position to realize this complex, whether she *ought* to, other things being equal. But what explains why this question should prove so difficult to close? Judgment and constitutive internalists believe the explanation has to do with the necessary connection between unqualified normativity (or normative judgment) and motive.

This is clearest with judgment internalism. Moore explained the open question by the hypothesis that goodness is a *nonnatural* property, but exactly what he meant by this was not entirely clear.<sup>29</sup> Later Moore remarked that the contrast he had in mind could be put by saying that

an intrinsic property is ‘natural’ if, in ascribing it to a natural object you are *to some extent* “describing” that object...and that hence an intrinsic property, e.g. the sense of ‘good’ with which we are concerned, is not ‘natural’ if, in ascribing it to a natural object you are not... describing that object *to any extent at all*.<sup>30</sup>

This may provide some confirmation of Stevenson’s diagnosis that the seeds

of non-cognitivism were already implicit in Moore's thinking. A nonnatural, intrinsic property, Moore is here saying, is one the attribution of which involves no description at all. Stevenson concluded that Moore's idea of a "nonnatural quality...must be taken as an invisible shadow cast by confusion and emotive meaning."<sup>31</sup> Similarly, Hare argued that the open question can be explained by the hypothesis that the judgment that something is good, unlike any descriptive judgment, necessarily expresses the speaker's *commendation*. Were 'good' to mean some complex of descriptive properties, *C*, it would "become impossible to commend [something] *for being C*."<sup>32</sup> Judging that it was good would add nothing to the judgment that it was *C*.

The cleanest way to be a judgment internalist is to be a *motive non-cognitivist*—to hold, that is, that the reason why ethical judgments are necessarily connected to judge's-motivation is that they *express* such motives, and that this exhausts their distinctively ethical content. Not all forms of non-cognitivism are judgment internalisms. For example, if Hume is read (anachronistically) as a non-cognitivist, rather than, say, a kind of naturalist or, perhaps, an error theorist, it is at least doubtful that he is best read as a motive non-cognitivist or, on these grounds, as a judgment internalist.<sup>33</sup> Humean moral judgments express *sentiments* directed towards contemplated motives or character traits, and these sentiments are only doubtfully motives themselves. Any form of non-cognitivism can "explain" the open question, of course, by the familiar fact that what is non-cognitive—whether sentiment, motive, or itch—cannot follow from anything, but if what is to be explained is the irreducibility of practical 'ought' judgments, then the appropriate non-cognitive candidate would seem to be a motive. Hare's view is that 'ought' judgments express a universal prescription, a preference for anyone doing a kind of thing in a kind of circumstance.<sup>34</sup> And Gibbard holds that judgments about what is reasonable or sensible express the acceptance of a norm permitting it, and that the latter state of mind includes a tendency to be governed by the norm oneself in like situations.<sup>35</sup> I say this is the cleanest way to be a judgment internalist since, for a motive non-cognitivist, the state of mind expressed by a judgment of unqualified normativity is exhausted by its motivational powers. Since it is not also a belief, the question cannot even arise, how can a belief of this sort motivate?

We have, then, three fundamentally distinct kinds of view, all of which have been called "internalism": judgment internalism and two sorts of existence internalism, constitutive and non-constitutive. What lies behind judgment internalism—at least in its purest form—is a view about, to paraphrase Stevenson, the "non-cognitive content of ethical terms." The reason why sincere assent to ethical judgments is necessarily connected to judge's-motivation is that expression of such motivation is all that ethical judgment *is*. What such judgments express is non-cognitive, a motivating attitude: a preference, perhaps, or the acceptance of a norm.<sup>36</sup>



Non-cognitivists do not offer this, of course, as an account of normativity (as distinct from normative judgment), since for them there really is no such thing. But unlike error theorists, who agree with this latter thesis, motive non-cognitivists believe that ethical judgments express no belief which can be confounded by this philosophical discovery. And if there need be no such thing as unqualified normativity to make sense of normative judgment, no question arises concerning what it might consist in, and no philosophical incentive to think that 'ought's have anything essential to do with agent's-motive. To be sure, a motive non-cognitivist might hold the *ethical* position that agents ought only to do what they can be motivated to do under certain conditions, but his meta-ethics provide no particular ground for this view.

### **Constitutive internalism and philosophical naturalism**

Constitutive internalists maintain, on the other hand, that there is something deserving to be called unqualified normativity, and seek to understand it as having a realization in the *practical realm*, from an *agent's* point of view, in motives generated through practical reasoning. Consider, for example, how the laws of nature function on Hobbes's view. On the one hand, Hobbes refers to these as "*dictates of reason*" and defines them as "precept[s] or general rule[s], found out by reason, by which a man is *forbidden* to do that which is destructive of his life, or taketh away the means of preserving the same; and to omit that, by which he thinketh it may be best preserved."<sup>37</sup> On the other, he says that the laws of nature are "but conclusions, or theorems concerning what conduceth to the conservation and defence of themselves."<sup>38</sup> And while he writes that "we ought to judge those actions only wrong, which are repugnant to right Reason," what he means by right reason is not some faculty of rational intuition, but simply a correct "reckoning" of consequences.<sup>39</sup>

But how is it possible for reason to *dictate* anything if the only thing it can tell us is what is likely to happen if we act in one way rather than another? What turns the latter, *theoretical* conclusion into an 'ought to do'? Hobbes never faces this question directly, but we can see how he must be thinking. Any of us, Hobbes believes, "by natural necessity...intend[s] the procurement of those things which are necessary to his own preservation."<sup>40</sup> Self-preservation is an inescapable human end. Thus reckoning concerning what is necessary to achieve self-preservation has an inescapable *practical* relevance for us. As *agents*, unavoidably viewing the world *sub specie* the end of self-preservation, our conclusions regarding how our lives are "best preserved," give rise to dictates, to 'ought to do's. Of course, could we give up this end, the most we could conclude would be that we ought either to do what is necessary for self-preservation or renounce it as end, but the latter, Hobbes believes, is not an option that is open to us.

According to this line of thinking, theoretical reasoning becomes practical when placed in the service of ends—from the agent’s point of view. Viewed from an observer’s perspective, the thought that keeping covenants is, for anyone, the strategy necessary to achieve self-preservation is simply an observation about how things lie in nature. But if self-preservation is the end at which we unavoidably aim, it becomes, from our point of view as agents, a sufficient ground for the thought that keeping covenants is the strategy one *ought* to pursue. For Hobbes, this strategy’s being unqualifiedly normative for human beings just *is* its being the case that human agents have a rationally conclusive  *motive* for pursuing it.

Here, then, is a way of maintaining a view about the distinctive normativity of ethics within a comprehensively naturalist metaphysics and a broadly empiricist epistemology. There is no *sui generis* normative metaphysical order; nor have we a cognitive faculty that might access it if there were. Viewed from the observer’s perspective, ethical truths do not differ fundamentally from those regarding any other natural connections in the world. They have a distinctive subject—the relations between human action and consequence—but they nonetheless concern the same natural order as does natural philosophy. Viewed, however, from the agent’s point of view, ethical truths have an inescapable motivational force, since they concern the necessary means to an unavoidable end. And this is what their distinctive normativity consists in. Its being the case that a person *ought* to keep covenants, as this is the necessary means to an inescapable end, consists in his therefore having, as agent, an unavoidably conclusive motive for so doing. Ethics’ distinctive normativity consists in practical relevance from the agent’s point of view.

Although Hobbes pioneered this way of viewing normativity in general, and that of morality in particular, his near contemporary Richard Cumberland was more methodologically explicit.<sup>41</sup> Cumberland quite straightforwardly announces his belief that “the whole of moral philosophy, and of the laws of nature, is ultimately resolved into natural observations known by the experience of all men, or into conclusions of true natural philosophy.”<sup>42</sup> How then is morality’s normativity to be understood? For Cumberland, as for his contemporaries more generally, this was the same as asking how morality can *obligate* agents, how it can *bind*.<sup>43</sup> This latter idea, however, is “somewhat obscure from metaphors; for the mind of man is not properly tied with bonds.” “There is nothing,” he continues, “which can superinduce a necessity of doing or forbearing any thing, upon a human mind deliberating upon a thing future, except thoughts or propositions promising good or evil to ourselves or others, consequent upon what we are about to do.”<sup>44</sup> Since any rational agent “necessarily seeks his own greatest happiness,” propositions concerning what is necessary for that have an inescapable practical relevance, and this is their normative force.

There is, Cumberland believes, a single law of nature: “the greatest bene-

volence of every rational agent towards all, forms the happiest state of every, and of all the benevolent, as far as is in their power."<sup>45</sup> So stated, this is a finding of natural philosophy. To appreciate its normative force, we must view it from the perspective of an agent "deliberating upon a thing future." From this standpoint, it has an unavoidable practical relevance, as it concerns the necessary means to a necessary end, and this is its normative force. Cumberland holds that the law of nature is a "practical proposition" which can be expressed variously as a proposition of natural philosophy concerning the connection between acts and consequences, as an imperative ("Let that action, which is in thy power, and which will most effectually, of all those which you can't exert, promote the common good in the present circumstances be exerted"), or "in the form of a gerund; 'Such an action ought to be done.'" "In my opinion," he concludes, "these several forms of speech...mean the same thing."<sup>46</sup>

It is important to realize that even in its purest, instrumentalist form, the sort of view we are considering requires a distinction between actual motivation of actual agents, on an occasion, and motives they would have if their practical reasoning were in order. Both Hobbes and Cumberland take it that practical reasoning transfers motivation from end to necessary means, but it is surely conceptually possible for a being to have an end, believe that the only means to it is an action at hand, and not be motivated appropriately. Cumberland, indeed, recognizes this is a real possibility. "Upon discovering the law of nature," he writes, "all men are obliged, whether it be of so great weight with them, as perfectly to incline their minds to what it persuades" or not. They are obliged because belief in this proposition "would certainly prevail; unless the ignorance, turbulent affections, or rashness of men, like the fault in the balance, opposed their efficacy."<sup>47</sup> The law of nature is normative, therefore, because it would motivate an agent whose practical reasoning was not faulty. And correct *practical* reasoning involves, not just having correct beliefs about practice, but also deliberating about these, bringing them to bear on choice, in the right way. For an instrumentalist, this requires that motivation transfer from end to means. A failure to do so is a failure of practical reasoning.<sup>48</sup>

At this point, the sort of view we are now considering faces a dilemma. A purely instrumentalist version confronts the problem that our actual, uncriticized ends seem to provide an inadequate basis for normativity since we can surely also ask what ends we should have. The connection the instrumentalist proposes between means and ends seems uncontroversially defensible, at least, on grounds of something like practical coherence.<sup>49</sup> But that it is incoherent to maintain an end while refusing the only available means can provide no more reason to take the means than to renounce the end. On the other hand, any attempt to strengthen instrumental rationality to provide a more adequate basis for normativity, through specifying rational

ends, seems to bring an accompanying increase in controversy, if not implausibility. Why, for example, must a rational agent aim at his own happiness?

An increasingly popular strategy has been to enhance instrumental reasoning with an account of the rational criticism of ends that looks primarily to an uncontroversial ideal of theoretical rationality. An agent's end may be based on ignorance in the straightforward sense that, were she to know more about it, she would renounce it. When the pursuit of an end is rooted in error, then, as Brandt (among others) has argued, it is plausible to suppose that this undermines the justification for taking necessary means.<sup>50</sup> Normativity cannot derive simply from the relation between means and end; it is also necessary that the agent would be motivated to adopt (or retain) the end were she to be perfectly informed about it and its place in the natural order. The first step in this strategy, then, is to hold that a necessary condition of its being the case that an agent ought to do something is that she would be motivated to do so were she *epistemically* perfect and instrumentally rational.<sup>51</sup> This will not be sufficient, however, since mere knowledge may fail to enter into deliberation or practical reasoning at all. As in Cumberland's example, "turbulent affections, or rashness" may impede judicious consideration of what one knows.<sup>52</sup> Correct theory must be brought properly to bear on practice in the agent's deliberative reasoning.

Of course, the constitutive internalist's standard for correct practical reasoning cannot be set externally, by whether the agent is motivated to do what she (independently) ought. The independent variable must be some procedure of practical reasoning whose correct exercise can be specified internally and not by its approximation to some external, independently specifiable result.<sup>53</sup> Thus, for Brandt, whether an action is rational depends on whether there is a stable truth concerning what the agent would be dominantly motivated to do were she knowledgeable about her situation *and* vividly to represent and think about what she knows.<sup>54</sup>

We may think of this as the appropriate outcome of a certain kind of philosophical naturalism. Concerned to place normativity in the natural order, the philosophical naturalist can hold that normativity consists in the *practical* consequences of accurately representing *that very order*. Thus, not only does normativity become a natural fact; it becomes the natural fact concerning whatever stable effect an agent's accurately representing the natural order itself has on her motives. So conceived, normative force is motivational force, not from just any motives, but from those arising from the contemplation of experience and correct natural theory. This situates normativity within an agency maximally improved by the experience and knowledge of nature.

### Constitutive internalism and autonomy

Historically, philosophical naturalism was a major force leading to the original constitutive internalisms of the early modern period, as it continues to be today. But it is not the only line of thought leading to constitutive internalism whose influence dates from that period. A second begins by thinking about the capacity to be subject to norms, with the thought that only agents of a certain kind *can* be normatively bound. It then seeks to understand what normativity is by relating it to this capacity. Locke, for example, identifies himself in the *Essay* with “those who cannot conceive, how any thing can be capable of a Law, that is not a free Agent.”<sup>55</sup> And, beginning in the second edition, he provides an account of the distinctive capacity for self-determination “intellectual beings” have owing to their ability to step back from present desires and critically examine alternatives. Moreover, while he never renounces his usual theological voluntarism, Locke includes along with his new theory of autonomy an argument that agents can only be obligated by motives created by self-determining practical thinking, and at points suggests that obligation just consists in this.<sup>56</sup>

We are most familiar, of course, with the shape this line of thought takes in Kant—viz., that practical laws are realized in the self-legislation of a free rational agent—but in some form or other it runs deep in the thinking of many philosophers. Among the early modern British moralists, for example, it can be found in Cudworth, Shaftesbury, and Butler (as well as Locke).<sup>57</sup> And in addition to Kant and post-Kantian idealists, it often appears in recent philosophers influenced by Kant, such as Thomas Nagel.<sup>58</sup> One way into this way of thinking is to begin with Kant’s discussion of his claim that we unavoidably regard ourselves as free to do what we ought.<sup>59</sup> Sometimes the doctrine that ‘ought’ implies ‘can’ is understood to mean that a person must be able to do what he ought in the sense that, were circumstances to be such that he wanted so to act, he would—nothing is stopping choice from eventuating in action. What is most interesting about Kant’s discussion, however, is that he explicitly distinguishes between the sense in which we might say that someone could have refrained from doing something, meaning that he would have if, say, a policeman had been at his elbow, and the sense with which he is concerned. Kant’s claim is that when, for example, one is asked by a tyrant to betray an honorable man, and thinks one ought not, one is committed thereby to the thought that one *can* not do so *for the very reasons one thinks one ought not*. By Kant’s lights, this is not a theoretical conjecture that might falsified, say, by a proof of determinism. It is an irreducibly *practical* thought, one we cannot avoid *as agents* deliberating about what to do, or in judgments of others (or ourselves) that involve an imaginative projection into their deliberative standpoints.<sup>60</sup>

The *deliberative* thought that, for certain reasons, one ought to choose

an alternative requires the thought that one can choose that alternative for those very reasons. Nor is this latter thought confined to actual deliberation. Imputations of responsibility, either to ourselves or others, also require it, albeit by way of imaginative projection. To impute responsibility is to regard someone as a person—to relate to her as someone who can determine herself by reasons.<sup>61</sup> When we reproach ourselves for not having acted as we ought, and thereby impute responsibility, we must suppose we could have so acted for the reasons we (now think) we ought—that these could have been *our* reasons for acting. We project ourselves into our earlier deliberative standpoint from which we take such action to be (have been) a practical alternative. Likewise, when we blame, and impute responsibility to, another, we commit ourselves to a deliberative judgment we think he should have made and could have acted on from his (agent's) viewpoint.<sup>62</sup> We judge there were reasons *for him to act*, and, in so thinking, think he could have acted for those reasons.<sup>63</sup> In this way, reproach and blame differ from mere sanction or retaliation. Reproach does not just aim to *give* the agent a reason to do what we think he ought; it expresses the judgment that there *existed* sufficient reason, reason on which the agent could have acted. We can hardly blame others for doing what we think they had good and sufficient reason to do.

Suppose, then, that thinking a person is genuinely bound by a norm, and appropriately held responsible for following it, commits us to thinking she has the capacity to determine herself to conform to it. We will then require a philosophical understanding of normativity that can explain this connection. We will need an explanation of what normativity is such that a person is normatively bound to do *A* only if she can autonomously choose to do *A*. Constitutive internalists who take these concerns seriously give various versions of the following answer: being obligated consists in having a dominant motive resulting from the proper exercise of the capacity for autonomy. They proffer an account of autonomous practical thinking, and hold obligation to consist in motives resulting from the correct exercise of that. Thus Locke, holding that the only rational end is the agent's greatest happiness, and that autonomy involves the exercise of a free agent's power to step back from the press of present desire and, by imaginatively considering the objects of desire "on all sides," make a genuinely practical judgment of available future happiness with appropriate motivational force, maintains that this motive is an obligation to happiness.<sup>64</sup> And what makes morality, the law of nature, obligatory for Locke, is that God has divinely sanctioned it with eternal punishment and reward, and given human agents sufficient evidence of this that they can determine themselves to act morally.

For Kant, of course, this is a description of heteronomy. No material end, whether the agent's happiness or the happiness of all, can serve as an autonomous agent's governing motive. A Kantian autonomous agent governs her

conduct by a conception of unqualified normativity, by a conception of law. Autonomy is *normative self-regulation* or *self-government*. This involves two, somewhat distinguishable elements in Kant's thought. First, an autonomous agent can act on a *rule* or "maxim" he gives to himself. And second, an autonomous agent can act on a conception of *law* (of unqualified norms that apply to any agent capable of normative self-regulation) by virtue of a distinctive second-order capacity for self-regulation (a capacity to regulate her self-regulation by maxims). "Nothing is left," ultimately to ground autonomous action than the idea of "the bare conformity of actions to universal law as such, this alone must serve the will as its principle." And the only way an autonomous agent can discipline rule-guided conduct by this idea is, Kant argues, to do so by the Categorical Imperative. An agent fully governs herself by a conception of law only if she subjects her will to the form of law—only if, he argues, she constrains her maxim-guided conduct by those she could will *as law*. "That is to say," he concludes, "I ought never to act except in such a way *that I can also will that my maxim should become a universal law...*"<sup>65</sup>

Locke's agent determines himself by judgments about his greatest happiness or good, but is unable to question whether this is an (or the only) end at which he *ought* to aim. Nor, indeed, does he aim at his own happiness because he thinks he ought to do so.<sup>66</sup> The idea of the normative plays no deliberative role. For Kant, however, autonomy is realized only by an agent for whom the idea of the normative (a conception of law) fundamentally governs practical deliberation. And it can do so only if an agent can act on rules and discipline her doing so by the Categorical Imperative.<sup>67</sup> If, for a given situation, she finds only one of a set of contenders she can will universally, she acts on it. She wills her acting on the maxim as an instance of willing that all do. And her being obligated to do so consists in the fact that she has a conclusive motive for doing so as a result of the practical thinking that realizes autonomy.<sup>68</sup> Thus are practical laws "laws of freedom." Unqualified normativity is the practical force for the agent of guiding herself by a conception of the unqualifiedly normative.

There are, then, two quite different lines of thought leading to constitutive internalism. For the philosophical naturalist, concerned to place normativity within the natural order, there is nothing plausible for normative force to be other than motivational force, perhaps when the agent's deliberative thinking is maximally improved by natural knowledge. This gives the naturalist a way of holding that while there is a sense in which, as Cumberland put it, "the whole of moral philosophy,...[can] ultimately [be] resolved into...conclusions of true natural philosophy," moral philosophy is not the very same subject as natural philosophy. What is distinctive about ethics, unqualified normativity, is the practical force some natural propositions have for us as agents.

For the line of thought we have just been considering, however, the attraction of constitutive internalism is primarily that it provides a philosophical understanding of the connection between being appropriately held subject to a norm and the capacity for normative self-determination. Only persons, it seems, can be subject to norms. But if to regard someone as a person subject to a norm is to think her able to determine herself by that very norm, then we need some understanding of what it is to be subject to a norm that assures this connection. But if normativity just is the force of motives resulting from self-determining practical reasoning, this connection is assured.

We have, then, two traditionally distinct philosophical rationales for constitutive internalism.<sup>69</sup> The thought that normativity can only adequately be grasped from within the agent's standpoint is, however, common to both—both situate normativity within agency. In this they contrast with the purest form of judgment internalism, motive non-cognitivism, which takes normativity to be a projection of the motives of the ethical judge, and with non-constitutive existence internalism, which locates it independently either of agency or judgment, in a distinct metaphysical order.

#### Notes

1. See, e.g., David Brink, *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press, 1989), pp. 37-80.
2. Brink is a notable exception.
3. For a recent example see James Dreier, "Internalism and Speaker Relativism," *Ethics* 101 (1990): 6-26. Robert Audi calls this position "motivational internalism" in "Internalism and Externalism in Moral Epistemology," *Logos* 10 (1989): 13-37; see esp. p. 22. Mostly, however, he is concerned with internalist and externalist theses regarding moral knowledge, by analogy with a familiar contrast in epistemology.
4. I make this distinction in *Impartial Reason* (Ithaca: Cornell University Press, 1983), p. 54. See Brink, p. 40 for a similar distinction between what he calls "appraiser internalism" and "agent internalism."
5. Putting it this way is a little misleading, since Falk's point was that one conception we have of morality is a wide one in which we take the moral 'ought' to dictate whatever, on balance, there is reason to do. See "'Ought' and Motivation," *Proceedings of the Aristotelian Society* 48 (1947-8): 111-138; reprinted in Falk, *Ought, Reasons, and Morality* (Ithaca: Cornell University Press, 1986), pp. 21-41. Further references will be to the latter. For convenience, I will sometimes omit the qualifiers "necessarily" and "actually or dispositionally."
6. The other classic article on internalism and externalism is William K. Frankena's "Obligation and Motivation in Recent Moral Philosophy," in A. I. Melden, ed., *Essays in Moral Philosophy* (Seattle: University of Washington Press, 1958), pp. 40-81.
6. Falk writes that Prichard is an externalist because "he holds that the man who while granting a duty doubts he has also a motive has a real axe to grind: there is no convincing him that he has a motive except by considerations additional



to those which already convince him that he has the duty." (27)

7. As he argues in "Action-Guiding Reasons," in *Ought, Reasons, and Morality*, pp. 82-98.
8. See Kurt Baier, *The Moral Point of View* (Ithaca: Cornell University Press, 1958), pp. 148-156, for the distinction between *justifying* and *motivating* reasons. Baier is an example of a philosopher who holds that justifying reasons are independent of motivation. For Baier's critique of an internalism about reasons, see "The Social Source of Reason," *Proceedings of the American Philosophical Association* 51 (1978): 707-733.
 

The remark has sometimes been made (including by the present author), that the distinction between justifying and motivating reasons derives from Hutcheson's distinction between "justifying" and "exciting" reasons. This is misleading since Hutcheson meant by the former a ground for moral approval (of character, motive, or motivated action), and he also held that such grounds cannot be reasons to act. Thus they are not justifying reasons in the present sense. See Francis Hutcheson, *Illustrations on the Moral Sense*, ed. Bernard Peach (Cambridge: Belknap Press of Harvard University Press, 1971), p. 121.
9. Note the emphasis on "ethical" rather than "meta". We need presuppose no sharp line between meta- and normative ethics.
10. On this point, see Sidgwick's chapter "Ethical Judgments," in *The Methods of Ethics*, 7th ed. (London: MacMillan, 1967), pp. 23-38.
11. Some of the resonances of this latter phrase are misleading, since the intended contrast is orthogonal to one between a *prima facie* (or *pro tanto*) 'ought' and one "all things considered."
12. It is notable, in this connection, that G. E. Moore, whose ethics are fundamentally axiological, nonetheless understands intrinsic goodness to be essentially unqualifiedly normative. He regards 'what ought to exist for their own sakes' as a synonym for 'what is good intrinsically' or 'absolutely'. (*Principia Ethica* (Cambridge: Cambridge University Press, 1966), p. viii) It is essential to Moorean intrinsic goodness that, other things equal, everyone ought to do what they can to realize it. On this point see William K. Frankena, "Obligation and Value in the Ethics of G. E. Moore," in P. A. Schilpp, ed., *The Philosophy of G. E. Moore*, 3rd ed. (La Salle, Ill.: Open Court, 1968), pp. 99-110. See also Moore's reply, pp. 560-581, 592-606.
13. This is evident, for example, in Harman's argument that moral relativism follows from a certain form of existence internalism about morality. Unless this latter premise were true, he in effect argues, morality would not be unqualifiedly normative. See Gilbert Harman, "Moral Relativism Defended," *The Philosophical Review* 85 (1975): 3-22; and "Metaphysical Realism and Moral Relativism: Reflections on Hilary Putnam's *Reason, Truth, and History*," *The Journal of Philosophy* 79 (1982): 568-575.
14. Hare may seem to be a counterexample since he is a judgment internalist about morality, but holds, nonetheless, that whether it is rational to be moral may best be approached by Brandt's method of determining what course of conduct the agent would prefer under cognitive psychotherapy. On such a view there will be no necessary connection between judging what a person morally ought to do and judging what it would be rational for him to do. But by Hare's lights the latter is a descriptive rather than a normative question. In fact, not only does Hare hold moral judgment to be of the unqualifiedly normative, he holds it to be of the *overridingly* unqualifiedly normative. R. M. Hare, *Moral Thinking* (Oxford: Clarendon Press, 1981), pp. 20-24, 214-216.
15. E.g., by Brink, *Moral Realism*, p. 39.

16. See, e.g., Alfred Mele, "Motivational Internalism: The Forms and Limits of Practical Reasoning," *Philosophia* 19 (1989): 417-436.
17. Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge: Harvard University Press, 1990), pp. 55-82. For the last alternative, see, e.g., David McNaughton, *Moral Vision* (Oxford: Blackwell, 1988).
18. Richard Price, *A Review of the Principal Questions in Morals*, ed. D. D. Raphael (Oxford: Clarendon Press, 1974), p. 186. This passage is reprinted in D. D. Raphael, ed., *British Moralists* (Oxford: Clarendon Press, 1969), v. ii, p. 194.
19. Note that this is also different from thinking that being in the right motivational state is necessary for epistemic access.
20. Price, *A Review*, p. 114; Raphael, *British Moralists*, v. ii, p. 167.
21. J. L. Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1978), pp. 38-42.
22. *Ibid.*, p. 40.
23. Such deliberation may be only imaginative, of course.
24. By, independently, John McDowell and David Wiggins. For McDowell, see, e.g., "Are Moral Requirements Hypothetical Imperatives," *Proceedings of the Aristotelian Society*, supp. vol. 52 (1978): 13-29; "Virtue and Reason," *Monist* 62 (1979): 331-350; "Values and Secondary Qualities," in *Morality and Objectivity*, ed. Ted Honderich (London: Routledge and Kegan Paul, 1985); and "Truth and Projection in Ethics," Lindley Lecture (University of Kansas, 1987). For Wiggins, see, e.g., "Truth, Invention, and the Meaning of Life," "Truth as Predicated of Moral Judgments," and "A Sensible Subjectivism?," reprinted in Wiggins, *Needs, Values, and Truth: Essays in the Philosophy of Value* (Oxford: Basil Blackwell, 1987).
25. For a discussion of sensibility theories, see Stephen Darwall, Allan Gibbard, and Peter Railton, "Towards Fin de Siècle Ethics," *The Philosophical Review*, forthcoming, on which the discussion in the text draws.
26. For further discussion of these and related points, see *Ibid.*
27. William K. Frankena, "The Naturalistic Fallacy," *Mind* 48 (1939): 464-477; Gilbert Harman, *The Nature of Morality* (New York: Oxford University Press, 1977), pp. 19-20; David Brink, *Moral Realism*, pp. 151-154, 162-163.
28. See note 12 above.
29. Including perhaps, it is fair to say, to Moore himself. See Moore's "Reply to My Critics," in *The Philosophy of G. E. Moore*, pp. 581-592.  
 For an extremely illuminating discussion of 'nonnaturalness' and the open question argument within the context of G. E. Moore's early thought, see Darryl Wright, *Refuting Idealism: G. E. Moore's Metaethics in Historical Context* (Ph.D. diss., University of Michigan, 1991).
30. *The Philosophy of G. E. Moore*, p. 591. See also note 24 above.
31. Charles Stevenson, *Ethics and Language* (New Haven: Yale University Press, 1944), p. 109.
32. R. M. Hare, *The Language of Morals* (London: Oxford University Press, 1964), p. 85.
33. For Hume as naturalist, see, e.g., Barry Stroud, *Hume*, (London: Routledge & Kegan Paul, 1977), pp. 171-218; for Hume as error theorist, see J. L. Mackie, *Hume's Moral Theory*, (London: Routledge & Kegan Paul, 1989); for doubts about Hume's judgment internalism, see Charlotte Brown, "Is Hume an Internalist?," *The Journal of the History of Philosophy* 66 (1988): 69-87.
34. On the latter point, see Hare, *Moral Thinking*, p. 91.
35. *Wise Choices, Apt Feelings*, pp. 68-82. Note that, unlike Hare's view, this does not imply any preference for others being governed by the norm.

For Gibbard, moral judgments (in the narrow sense) do *not* directly express judge's motive; rather they express acceptance of norms for having certain *reactions* (viz., feelings of guilt and resentment). Note also that the open question argument takes a distinctive form in Gibbard's hands. For him, the core phenomenon is that genuine disagreement about what it is rational to do apparently survives complete agreement about factual matters. For example, those who think it rational to take one box in Newcomb's Problem, and those who disagree and think both boxes should be taken, can agree completely about the features of the situation and the respective properties of Bayesian and causal decision theory. The best way to understand the issue between them is, Gibbard argues, non-cognitively, viz., that these different judgments express conflicting non-cognitive states of mind: the acceptance of conflicting norms. On this point see *Wise Choices, Apt Feelings*, pp. 9-22, 36-54.

36. Obviously, on the latter view, we must be able to understand what it is to accept a norm non-cognitively.
  37. Thomas Hobbes, *Leviathan*, chs. xv and xiv, respectively.
  38. *Ibid.*, ch. xv.
  39. *De Cive*, ch. ii; *Leviathan*, ch. v.
  40. *De Cive*, ch. iii.
  41. Richard Cumberland (1631-1718) was known as one of the three great modern writers on natural law, the other two being Grotius and Pufendorf. His *De Legibus Naturae disquisitio philosophica* (1672) was published in an English translation by J. Maxwell as *A Treatise of the Law of Nature* in 1727. Selections from this work can be found in J. B. Schneewind, ed., *Moral Philosophy from Montaigne to Kant* (Cambridge: Cambridge University Press, 1990), v. i, pp. 138-155; and in D. D. Raphael, ed., *British Moralists* (Oxford: Clarendon Press, 1969), v. i, pp. 79-102.
- Although Cumberland may be considered a follower of Hobbes in some significant methodological respects, a major purpose of his *Treatise* was to criticize Hobbes.
42. *A Treatise of the Law of Nature*, trans. J. Maxwell (London: 1727), facsimile edition, (New York: Garland Publishing, Inc., 1978), p. 41.
  43. I sketch this phenomenon in "Motive and Obligation in the British Moralists," in Ellen F. Paul, et al, eds., *Foundations of Moral and Political Philosophy* (Oxford: Basil Blackwell, 1989), pp. 133-150. (Also published as v. 7, no. 1 of *Social Philosophy & Policy*.)
  44. *Treatise*, p. 233. Note that Cumberland says that propositions concerning good consequences for others can also "superinduce a necessity of doing of forbearing." Cumberland's position is actually a good deal more complex than I am presenting it here.
  45. *Treatise*, p. 41.
  46. *Treatise*, p. 180. The "fitness, which is expressed by a gerund, wants explanation, which is to be fetched, either from the necessary connexion of the means with the end, or from the obligation of a law." And ultimately, Cumberland resolves the latter itself into the "necessary connexion of the means with the end."
  47. *Ibid.*, p. 234.
  48. On this point, see Christine Korsgaard, "Skepticism About Practical Reason," *The Journal of Philosophy* 83 (1986): 5-25.
  49. For a discussion of this point, see *Impartial Reason*, pp. 43-50, 62-77.
  50. Richard Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979).
  51. I am indebted to Connie Rosati here.

52. This is the theme of W. D. Falk's "On Learning about Reasons," in *Ought, Reasons, and Morality*, pp. 67-81.
53. Compare Rawls's distinction between pure procedural and perfect (or imperfect) procedural justice. In the former case, as in a fair gaming wheel, there is no independent specification of a just outcome to which a just procedure approximates. The independent variable is the idea of a just procedure, and a just outcome is whatever eventuates. In the latter case, as when one person divides and the other chooses, the independent variable is the notion of a just outcome and a just procedure is what best approximates that. See *A Theory of Justice* (Cambridge: Belknap Press of Harvard University Press, 1971), pp. 85f.
54. Richard Brandt, *A Theory of the Good and the Right*, pp. 110-129. Peter Railton provides a similar, but interestingly different account of a person's good in "Moral Realism," *The Philosophical Review* 95 (1986): 163-207. The major wrinkle in Railton's account involves the idea of an epistemically-improved second-order desire. Roughly, *X* is part of *A*'s good [and hence something at which *A* ought to aim, other things equal], if a fully knowledgeable, experienced, and imaginative *A* (call her *A\**) would want *A*, given *A*'s actual situation, to want *X*.
55. John Locke, *An Essay Concerning Human Nature*, Peter H. Nidditch, ed. (Oxford: Clarendon Press, 1985), p. 76.
56. *Essay*, pp. 263-264, 266-267, 271; for the latter point see, especially, pp. 266-267.
57. For a short discussion of Shaftesbury on this topic see Darwall, "Motive and Obligation in the British Moralists." For Butler, see Darwall, "Conscience as Self-Authorizing in Butler's Ethics," in Christopher Cunliffe, ed., *Joseph Butler's Moral and Religious Thought* (Oxford: Clarendon Press, forthcoming). Cudworth's version of this view can be found in his manuscripts on free will, held in the British Library. One of these has been published: John Allen, ed., *A Treatise of Freewill* (London: John W. Parker, 1838). An excerpt appears in D. D. Raphael, ed., *British Moralists* (Oxford: Clarendon Press, 1969), v. i, pp. 120-134.
58. Especially in Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986), pp. 110-137. I discuss this aspect of Nagel's views in "How Nowhere Can You Get (and Do Ethics)?" *Ethics* 98 (1987): 137-157. The view I sketch in *Impartial Reason* can also be understood as a version of this line of thought, as I try to make clearer in "Impartial Reason," in *Contemporary Ethics*, James Sterba, ed. (Englewood Cliffs, N. J.: Prentice-Hall, 1989).
59. Immanuel Kant, *Critique of Practical Reason*, Lewis White Beck, trans. (Indianapolis: Bobbs-Merrill Co., Inc., 1956, p. 30.
60. Thus, Kant writes of the agent deliberating about refusing the betrayal: "Whether he would or not he perhaps will not venture to say; but that it would be possible for him he would certainly admit without hesitation." For an excellent discussion of Kant's ideas here, stressing their practical character, see Christine Korsgaard's "Morality as Freedom," in Y. Yovel, *Kant's Practical Philosophy Reconsidered* (Dordrecht: Kluwer Academic Publishers, 1989), pp. 23-48.
61. Thus Locke's famous passage: "*Person*...is a forensic term appropriating actions and their merit; and so belongs only to intelligent agents capable of a law..." *Essay*, p. 346.
62. As Nagel puts it, the "defendant" in a judgment of responsibility is an agent, and in imputing responsibility we implicitly "enter into the defendant's point of view as an agent" and "evaluat[e] his action from within it." *The View from Nowhere*, p. 120, 121.
63. It is, of course, consistent with this (indeed, it is required) that we take diminished capacity to act for reasons to diminish responsibility.

64. *Essay*, pp. 266-267.
65. Immanuel Kant, *The Groundwork of the Metaphysics of Morals*, H. J. Paton, trans. (New York: Harper Torchbooks, 1964), Ak. p. 402.
66. For an argument that such naturalisms as Brandt's and Railton's concerning an agent's good are incapable of adequately capturing normativity, because they do not connect up with autonomy in the right way, and that the normativity of ethics highlighted by Moore's open question argument is best understood in relation to autonomy, see Connie Rosati, "Agency and the Open Question Argument," unpublished manuscript, and *Self-Invention and the Good* (Ph.D. dissertation, University of Michigan, 1989).
67. For further discussion of the connection between autonomy and the search for unqualified normativity, see Darwall, "Autonomist Internalism and the Justification of Morals," *Nous* 24 (1990): 257-267.
68. Various qualifications are necessary here. Kant generally reserves 'autonomy' for "the property the will has of being a law to itself (independently of every property belonging to the objects of volition)." (Ak. p. 440) It is clear, however, both that Kant takes our capacity to determine ourselves by a conception of law as essential to what makes us subject to law, and that he thinks this is a capacity we can succeed or fail in exercising properly. In the text I am using 'autonomy' to refer to the result of the successful exercise of this capacity.  
Second, he generally reserves 'obligation' and 'ought' for a constraint the objective law imposes on a "will which is not necessarily determined by this law in virtue of its subjective constitution." (Ak. p. 413)
69. It may, of course, be possible to combine them. I argue that David Gauthier can be understood as having done this to some extent. See my "Rational Agent, Rational Act," *Philosophical Topics* 14 (1986): 33-57; and "Kantian Practical Reason Defended," *Ethics* 96 (1985): 89-99. Both rationales are also at work in Locke's thought.