



**Does Hume Have an Instrumental Conception of Practical Reason?**

Jean Hampton

*Hume Studies* Volume XXI, Number 1 (April, 1995) 57-74.

Your use of the HUME STUDIES archive indicates your acceptance of HUME STUDIES' Terms and Conditions of Use, available at <http://www.humesociety.org/hs/about/terms.html>.

HUME STUDIES' 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 HUME STUDIES archive only for your personal, non-commercial use.

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

For more information on HUME STUDIES contact [humestudies-info@humesociety.org](mailto:humestudies-info@humesociety.org)

<http://www.humesociety.org/hs/>

# Does Hume Have an Instrumental Conception of Practical Reason?

JEAN HAMPTON

Many philosophers and social scientists regard the instrumental theory of practical reason as highly plausible, and standardly credit David Hume as the first philosopher to formulate this conception of reason clearly. Yet I will argue in this paper that Hume does not advocate the instrumental conception of practical reason as that conception is normally understood by contemporary theorists who endorse it. Although it is often thought that Hume's position on reason is the "common-sense" one, I will argue that in a very fundamental way, his view of reason defies common-sense.

## The Instrumental Conception of Reason

Consider the following definition of an instrumental theory of reason, understood normatively:

- 1) An action is rational to the extent that an agent believes (reasonably)<sup>1</sup> that it furthers the attainment of an end; *and*
- 2) Human reasoning involves the determination of means to achieve ends, in a way described by the theory (I will say, henceforth, that using reason to determine the extent to which an action is a means to an end is an *instrumental* use of reason); *and*
- 3) These ends are in no way fixed by reason operating non-instrumentally; i.e., what makes them our ends is something other than reason.

In this article I will consider as instrumental all and only theories that accept these three theses.<sup>2</sup> Thesis 3 is normally understood as the hallmark of the instrumental theory. Hume's famous remark "Reason is...the slave of the passions" (T 415) is a clear endorsement of this thesis. In contrast, theories of reason that do not count as instrumental maintain that reason not only pursues means to ends, but also defines ends of action, thereby denying thesis 3. On this sort of view, it is reason that is, at least sometimes, the master. For example, Kant portrays reason as concerned both with the efficiency and the morality of an action, and as that faculty which determines whether any proposed course of action is acceptable. So for Kant, reason "constructs" the ends of action insofar as it picks out from among the goals proposed by our desires those which are appropriate for us to pursue: it is the final determiner of what our ends of action ought to be. Moreover, after approving these ends, Kant says that reason provides a motivation to achieve them in virtue of the fact that it has approved them, a motivation which, if it directs a moral action, is always sufficient to determine that action, if the agent chooses to let it.<sup>3</sup> So on Kant's view, practical reason is a master that issues directives, provides the impetus for obeying them, and calculates how best to do so.

There are two reasons why philosophers have been troubled by the Kantian type of view. First, it is a conception of reason that seems unacceptable from the standpoint of science. What special "sight" or access to normative reality can we realistically ascribe to human reason, such that it can tell us our ends in life? And how does a scientific world view permit us to believe that there are unmotivated ends which we are rationally compelled to pursue? Science, after all, does not recognize such objects or properties with inherent prescriptive power. J. L. Mackie calls such objects and properties "queer"—indeed, too queer, given the strictures of science, for us to believe they obtain.<sup>4</sup> Moreover, no scientific description of human beings has identified a rational capacity within us that can determine these objects, respond to their inherent prescriptivity, and motivate action in compliance with their requirements.

The other problem facing any non-instrumental account of reason concerns motivation. Most theorists believe that whereas the question "Why be moral?" is deeply troubling, the question "Why be rational?" does not seem to be. Yet if reason is developed along Kantian lines, it becomes so expansive, and its directives so wide-ranging and divergent from interests the individual is readily able to recognize, that the issue of behaving rationally, when rationality is defined like *that*, is now a real one.

For those who are committed to a naturalistic account of human beings, the motivational puzzles of the non-instrumental view, and the non-natural role it accords reason, make this approach to reason unacceptable.<sup>5</sup> In particular, for Hume, who wanted to be the Newton of the science of Man, it was important to put forward a conception of reason that grants it no occult

powers, and that presupposes a foundation that is utterly acceptable from a scientific point of view. The instrumental approach to reason seems to meet both criteria: indeed, even the *normative* version of this approach seems to be able to explain the force and strictures of its rational directives in a way that is scientifically acceptable. This is because, on the instrumental approach, reason dictates only hypothetical imperatives, and these imperatives seem to be reducible to beliefs and desires,<sup>6</sup> both of which are (arguably) allowed by, or reducible to, entirely natural phenomena. And the way in which hypothetical imperatives appeal to desires seems to make them motivationally unproblematic, without the inexplicable "magic" that Philippa Foot and others say supporters of categorical imperatives have clothed them in.<sup>7</sup>

However, in the arguments to follow, I will maintain that Hume developed a conception of reason that is *not* a variant of the instrumental conception, because he appreciated that a true instrumental conception is problematic on naturalist grounds. Hume's sagacity regarding that danger of the instrumental view has not been shared by contemporary naturalists who generally adopt the instrumental conception uncritically. Alas, what Hume proposes is highly implausible as a theory of *reason*, precisely because of its attempt to be plausible on naturalist grounds.

## Hypothetical Imperatives

According to the instrumental conception of reason, the directives of reason are—and can only be—hypothetical imperatives. Conventional wisdom has it that there is nothing mysterious nor especially troubling about the force of hypothetical imperatives. J.L. Mackie probably speaks for most philosophers when he explains the force of a hypothetical imperative as follows:

'If you want to do X, do Y' (or 'You ought to do Y') will be a hypothetical imperative if it is based on the supposed fact that Y is, in the circumstances, the only (or the best) available means to X, that is, on a causal relation between Y and X. The reason for doing Y lies in its causal connection with the desired end, X; *the oughtness is contingent upon the desire*.<sup>8</sup>

But what *exactly* does it mean to say that the "oughtness" in a hypothetical imperative is in some way a function of the desire predicated by the imperative?

Most philosophers have not realized that this question about hypothetical imperatives needs answering. The contingency of the directive in a hypothetical imperative on a certain desire, does not, by itself, explain why we *ought* to follow the directive. While it is true that the "ought" statement in a hypothetical imperative should be withdrawn if the action that has been

directed is not effective in satisfying an agent's desire, or if the agent does not have the desire,<sup>9</sup> nonetheless, if the agent has the desire and the action is effective, that ought statement "holds." But what does it mean to say this? How does this hypothetical imperative give the agent a reason for action, that is different from the desire assumed by the imperative? And how does the hypothetical imperative motivate us, by virtue of the fact that it gives us this reason?

These questions about the force of hypothetical imperatives can be answered in quite different ways, by theories that offer different accounts of the normativity of these imperatives. All of them sharply distinguish between the authority and the motivational efficacy of these imperatives and the reasons they give us. I will say that a reason generated by a hypothetical imperative is authoritative if it dictates an instrumentally effective way to act in the circumstances. But dictating the best action is not necessarily motivating that action. I will say that a reason given to us by a hypothetical imperative is motivationally efficacious if it can, *by itself*, move us to act as it directs (albeit, perhaps with insufficient force to effect the action it directs).

Just because we know we have an authoritative instrumental reason to do some action *x*, it doesn't follow that this reason is motivationally efficacious. Alas, philosophers have persistently assumed that it is. To see why that assumption is false, consider a popular account of what makes something a reason for action, put forward by Bernard Williams. In order for an agent to have a reason for acting, according to Williams, there must be a deliberative connection between the reason's directives for action, and an agent's motives.<sup>10</sup> Williams calls such reasons "internal," and he denies that there can be any "external" reasons for action, i.e., reasons that do not have this deliberative connection with our motives. Note, however, that this distinction between "internal" and "external" reasons concerns the issue of what it is that makes something a reason for action, and not the issue of whether or not a reason is also a motive.

Hence Williams puts forward a variant of what I will call *Justificational Internalism*, which is the view that an agent has a reason to *x* if and only if *x*-ing is connected, via deliberation (correctly performed) with an internal feature of the agent. It is opposed by *Justificational Externalism*, which is the view that an agent has a reason to *x* iff *x*-ing can be connected, via deliberation, with some aspect of the world which need not be, and sometimes is not, an internal feature of the agent. But justificational internalism, by itself, takes no stand on the issue of whether reasons are also motives. Hence one can be a justificational internalist and a motivational externalist, denying that the reasons we have (by virtue of their deliberative connection with certain of our internal features) must also be motives themselves.<sup>11</sup>

But justificational internalism does not imply motivational internalism. Suppose that after a deliberation process, an agent arrives at the conclusion

that she has a reason to do *p* in order to achieve some object *q* which would satisfy her desire *s*. This reason is related to a motive (her desire *s*), and discovered by her in a deliberation process, so it would seem to be a reason that she *has*, and believes she has. But having said that, why does it follow, by virtue of the fact that it is a reason that she has and believes she has, that she is also motivated to perform *p*? To say that the motivation exists requires some further argument—justificational internalism does not, by itself, provide any argument at all with respect to the motivational efficacy of reasons.

Once we distinguish between the authority and the motivation of an instrumental reason, we can isolate a number of accounts of the “normative force” of a hypothetical imperative, which differ in how they understand the authority of these imperatives, and in how—if at all—they link the motivational efficacy of these imperatives with that authority.

First, there is what I will call the *Kantian* position—i.e., the Kantian position on the normativity of *instrumental* reason; I am not referring to Kant’s larger, non-instrumentalist conception of rationality. Kant is famous for the view that when we act morally we are acting on reason, and he contrasts such moral action with action that is (merely) caused by desires. This rhetoric, however, obscures the fact that he also recognizes a non-moral component of reason, from which we act when we are motivated by hypothetical imperatives. And a careful look at his account of the force of hypothetical imperatives shows he believes we are motivated to follow them *not* by the desires assumed by such imperatives, but by *reason*—that is, by the instrumental (and not the moral) component of reason. Although predicated on a desire, Kant maintains that hypothetical imperatives “present the *practical necessity* of a possible action as a means to achieving something else which one desires.”<sup>12</sup> This “necessity” is the substance of the authority of the imperatives, and it is *in virtue of that authority* that it has motivational effect on the agent:

Whoever wills the end, *so far as reason has decisive influence on his action*, wills also the indispensably necessary means to it that lie in his power.<sup>13</sup>

I have emphasized the phrase that shows that Kant believes, when we act from a hypothetical imperative, that it is not desire but (instrumental) reason that moves us, just as, when we act morally, it is not desire but reason (in this case, moral reason, not instrumental reason) that moves us, by virtue of its authority. (That is, these imperatives command us with “practical necessity.”) And while Kant is aware that many human beings resist the means they ought to take to achieve their ends, his position is that in virtue of the fact that we *ought* to will means appropriate to our ends, we will do so for as long as “reason has decisive influence” on our actions. So a Kantian position on the force of hypothetical imperatives is that when a rational agent acts on a hypothetical

imperative, she is not motivated to do so by the desire assumed in the imperative, but by the rational authority of the imperative, so that the imperative is motivationally efficacious *by virtue of its authority*.<sup>14</sup>

The main problem with the Kantian view is that postulating the idea that a reason's motivational efficacy is necessary by virtue of its authority is postulating a non-natural cause of action. Whether or not the authority of the hypothetical imperative is taken to motivate action directly, or via some psychological intermediary such as a desire, the Kantian view still insists that it is *this authority* that effects us. But that effect seems unlike any natural cause-and-effect event recognized or studied by science. Indeed, Kant himself accepts this, and strives, in his *Critique of Pure Reason* and in subsequent works on moral theory, to develop a defence of the non-natural motivational efficacy of this authority. As I have noted, the instrumental view has been popular with naturalists and social scientists generally, because it seems not only a highly plausible answer to the question "Why be rational?" but also one that is acceptable to naturalists insofar as it does not appear to rely on some kind of "magical" force which human beings are supposed to have the capacity to sense, and be motivated by. But the Kantian explanation of the motivational efficacy of a hypothetical imperative by virtue of its authority makes it just as magical as any categorical imperative. However contingent the hypothetical 'ought' is on a desire, it is still *not* the same as a desire; to say, therefore, that it can move us to action is to believe in the motivational efficacy of the authority of such a statement, and the idea that "ought" statements can motivate us has been the main reason many theorists have dismissed the idea that categorical imperatives have motivational efficacy.

The second position on instrumental reason understands a hypothetical imperative as involving only one component, namely, normative authority. On this view, an imperative is authoritative, but *never*, by virtue of that authority, also motivationally efficacious. That is, it gives us reasons for action, but those reasons cannot directly motivate us. They can, at best, be motivationally efficacious indirectly, for example, by causing the activation in us of certain desires (or any other motivationally efficacious psychological material) that motivate us to act as it directs; they cannot motivate us to act by themselves by virtue of their authority. Whereas Kantian instrumental reasons have what I will call "authoritative motivational force," this view attributes to instrumental reasons (at most) what I will call "causal motivational force." In virtue of the way in which Mill distinguished the "correctness" of a moral imperative from the motivation we have to follow that imperative, I am calling this position *Millian*, even though I am describing a position on the nature of instrumental reason, not moral reason.<sup>15</sup>

The Millian position would license criticism of people who fail to act rationally in virtue of the way they fail to act on the reasons supplied by instrumental rationality, even though their motivational structure is such that

they have no capacity for doing so. Whereas the Kantian position is a thoroughgoing form of motivational internalism, i.e., the view that to know that *x* ought to be done is to have a motive to do *x*, the Millian position embraces a form of motivational externalism, denying that it follows from knowing that I ought to do *x* that I have a motive to do *x*. So although both the Kantian and the Millian positions insist that to know one ought to do *x* is to have a reason to do *x*, they differ in the positions they take on whether these reasons are motivationally efficacious. Because the Millian position avoids postulating a non-natural motivational connection between an instrumental reason and our action, those who are interested in a form of instrumentalism that is motivationally plausible may believe that the Millian position is superior to the Kantian position on this matter. I do not believe it is, but I will not have time to pursue that issue here.<sup>16</sup>

The third view of the force of hypothetical imperatives was explicitly advocated by Hume; hence I am calling it *Humean*. This view agrees with the Millian position that instrumental reasons have no motivational efficacy by virtue of their authority, but at most have a (merely) causal effect on our motivational structure. On this view, hypothetical imperatives simply set out causal connections between ends and means, which have been worked out using logic and causal reasoning (both of which are understood to be components of theoretical reason). And this cause-and-effect information can at most be only indirectly involved in getting us to act, by causing in us the development of desires to perform the instrumentally effective action; it cannot effect action by itself:

...reason, in a strict and philosophical sense, can have an influence on our conduct only after two ways: Either when it excites a passion by informing us of the existence of something which is a proper object of it; or when it discovers the connexion of causes and effects, so as to afford us means of exerting any passion. (T 459)

But in this passage Hume is saying not only that reason cannot move us to action alone, without any help from desires at all, but more importantly that it has a (merely) causal effect on action, and no motivational effect by virtue of any (supposed) authority over action.

However, what is striking about the Humean conception is that it also says that instrumental reason has no authority over our actions!<sup>17</sup> Consider Hume's (famous) remarks in Book II:

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me. 'Tis as little contrary to reason to



prefer even my own acknowleg'd lesser good to my greater, and have a more ardent affection for the former than the latter. A trivial good may, from certain circumstances, produce a desire superior to what arises from the greatest and most valuable enjoyment; nor is there any thing more extraordinary in this, than in mechanics to see one pound weight raise up a hundred by the advantage of its situation. (T 416)

Note Hume's contention that one is not being irrational in preferring an "acknowledg'd lesser good" to a "greater" good. The meaning of these terms is somewhat obscure: in order for the phrase in which they occur to make sense, there must be some way that we can judge one good "greater" than another, apart from the strength of our preferences for either one. An appeal to instrumental reason gives us a way to make such a judgement. Suppose going to the dentist is a means to avoiding pain in the long term. Suppose further that an agent prefers avoiding long-term pain to spending an afternoon reading. Because going to the dentist is a means to the more preferred end, it is a greater good than spending the afternoon reading. And yet it makes perfect sense to imagine this agent saying that she prefers spending the afternoon reading to going to the dentist. Hume's position would be that *because* spending the afternoon reading is what one prefers to do, it is "not against reason" to do so. Our desires are not only the sole motivational force within us but also the only force within us that can "tell us what to do." And a desire can act upon us in a way that causes us to do something other than the course of action consistent with reason's information—for example, when it fails to cause within us a desire to perform the means that is sufficiently strong to counteract an opposing desire. (Remember Hume's mechanics metaphor: sometimes a pound weight can raise up a hundred by the advantage of its situation.)

Of course there can be some psychological process within us that is usually initiated by reason's instrumental deliberation, eventuating in a desire to do that which one has instrumental reason to do. Once again:

reason, in a strict and philosophical sense, can have an influence on our conduct only after two ways: Either when it excites a passion by informing us of the existence of something which is a proper object of it; or when it discovers the connexion of causes and effects, so as to afford us means of exerting any passion. (T 459)

In the latter case, reason has an "influence on our conduct" (perhaps it operates along the lines of what Hume calls a "general rule"), but note that this influence does *not* come about by virtue of its authority over our action. And

its causal effect on the development of motives in us is a contingent phenomenon, derailed in certain circumstances:

Men often act knowingly against their interest: For which reason the view of the greatest possible good does not always influence them. (T 418)

So while we can act "against reason" in the sense that we can fail to act according to its information, we cannot act against its (motivationally efficacious) authority over our action, because such authority does not exist. Again, to quote Hume:

actions do not derive their merit from a conformity to reason, nor their blame from a contrariety to it; and it proves the same truth more *indirectly*, by shewing us, that as reason can never immediately prevent or produce any action by contradicting or approving of it, it cannot be the source of the distinction betwixt moral good and evil.... Actions may be laudable or blameable; but they cannot be reasonable or unreasonable. (T 458)

Compare Hume's remarks on character and action in T 458, in which he once again insists that actions cannot be judged as rational or irrational:

Reason is the discovery of truth or falshood. Truth or falshood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact. Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. Now 'tis evident our passions, volitions, and actions, are not susceptible of any such agreement or disagreement; being original facts and realities, compleat in themselves, and implying no reference to other passions, volitions and actions. 'Tis impossible, therefore, they can be pronounced either true or false, and be either contrary or conformable to reason. (T 458)

So for Hume, it is, strictly speaking, incorrect to call an action irrational, even if that action fails to achieve the stated aim of the agent.<sup>18</sup>

At the deepest level, Hume's position should be understood as a view of what reason *is*: whereas philosophers such as Wollaston and Clark considered it to be a normative faculty with both authority and motivational efficacy over our actions, for Hume, reason is a purely informational faculty, working out relations of ideas and causal connections. Although he accepts that this information might play a causal role in the creation of a motive to perform a means to a desired end given the psychological processes of the human mind

(and Book II is full of accounts of psychological processes generating various passions), he rejects completely the idea that it has any normative authority over action, or any capacity to move us by virtue of that supposed authority.

Contrast the Humean view with the Kantian and Millian conceptions of rational imperatives, both of which assume that reason can have normative authority. On the Kantian or Millian views, if you say about someone, "He ought to do y to achieve x in the circumstances," you are invoking what you take to be an objective norm to criticize him. I call this the "instrumental norm" which directs us to pursue those objects and perform those actions that will be the most effective means to a desired end. For the Kantian or Millian instrumentalist, to say that "reason has authority over us" is really to say that *this instrumental norm* has authority over us.<sup>19</sup> But Hume rejects the idea that there is such an objective authoritative norm. For Hume, acting on a hypothetical imperative does not involve accepting (or being motivated by) the authority of such a norm, but instead involves being caused to act in this way by a process that is affected both by one's desires and by the information about how to satisfy them supplied by reason. That information, in conjunction with one's desires, might have the "feel" of an authoritative, natural norm, but in reality, one is being directed by a conjunction of entirely natural (and scientifically recognized) forces. And when those forces fail to direct us, Hume insists that we do not violate any authoritative code of reason applying to action—because no such code exists.

### **Does Hume Have An Instrumental Conception of Reason?**

Given Hume's position on the nature of the force of a hypothetical imperative, can he really be said to have an instrumental conception of reason, as I have defined it above? Consider that if the imperatives of instrumental reason have no authority, and motivate us to act only in virtue of certain psychological effects they tend to have on most (normally constituted) human natures, then if they fail to motivate some people, we need only explain this by pointing to the (unusual) natures of such people, which are unresponsive to these imperatives. But there would be nothing *wrong* or *irrational* about those natures. We could only say that, in these circumstances, these natures are different from usual. But this means the Humean view does not count as an instance of the instrumental theory of reason as I have defined it above because it violates thesis 1, i.e., that "an action is rational to the extent that it furthers the attainment of an end." The inclusion of that thesis in the definition of the instrumental conception is inescapable; for how could one endorse an instrumental theory of reason but not criticize, for example, the man who wanted to cure his tuberculosis but refused to take the medicine that would do it, or the woman who wanted a college B.A. but refused to work to pass any of her classes? To say, as Hume does in the passage quoted above, that

the actions of such people are merely "natural existences" that, as it happened, were not caused by correct cause and effect reasoning, but are not in a literal sense irrational, is to refuse to allow reason to have *any* critical impact on human behavior at all. Yet isn't this exactly the position one is taking when one says, as Hume does, that reason only has a theoretical role? Ironically, that position is inconsistent with the instrumental theory of reason, understood as involving a certain kind of rational criticism of behavior, that is standardly attributed to Hume!

Might it be possible to argue that Hume allows criticism of irrational people indirectly, using adjectives that characterize various kinds of vices, especially the vice of imprudence? Hume sanctions all sorts of criticisms in his discussion of natural vices, and the criticism of imprudence (i.e., behaving in a way that is advantageous in the short term but not in the long term) might be thought to apply to at least some of the "irrational" people I am describing. However, not only is imprudent action not the same as instrumentally irrational action (for example, some imprudent actions can nonetheless be instrumentally rational, in virtue of the fact that they satisfy short-term goals), more importantly, calling someone imprudent is *not* calling her mistaken—and the concept of mistake is central to the charge of irrationality. When, according to Hume, we criticize someone as vicious in some way, we are projecting on to her a property that is created from our displeasure at her actions, occasioned in us after sympathetic identification with her and/or others, where this projection is governed in certain ways by conventional rules generated by our society.<sup>20</sup> But upon reflection, we must admit that this criticism is a function of how her actions strike us. It does not mean that, by her lights, she made any mistake. And even if our criticism of her as (in some way) vicious is supposed to mean that she violated some (external) ideal, it is nonetheless a "projected" criticism that is based on certain emotions experienced by (many but not all) onlookers as they contemplate her behavior, and not a criticism of her behavior as violative of (external) standards of reason (for none exist), nor a criticism of her actions as (internally) mistaken given her motives.

It may seem that Hume's position is the "sensible" one, insofar as it eschews the idea that reason has strange powers, relies on desires as motives, and reduces hypothetical imperatives to foundations that seem metaphysically unproblematic from a scientific point of view. Yet on reflection it is highly counter-intuitive and fails as a theory of reason. Consider that according to Hume, one who hears a (correct) hypothetical imperative "Do y in order to achieve x" and doesn't perform y *makes no mistake* when he doesn't perform y. He cannot be criticized internally (as a Kantian would wish) because he has no desire to perform that action which is a means to his end and thus can do nothing else but act on the desire precluding the instrumentally directed action. Even more strikingly, he cannot be criticized

externally (as a Millian would wish) because on the Humean view reason does not provide us with a normative standard by which to judge action. So someone who fails to act so as to achieve his ends, in a situation where he has no desire to perform the actions required to achieve those ends, does nothing wrong. He violates no rational standards of action; and indeed, that's the point of this Humean view—*there are no rational standards of action*. The Humean position must be that human beings can only act *non-rationally*, since reason is neither the standard nor the motivation (by virtue of being the standard) of human action. Instrumental assertions amount merely to psychologically efficacious ways of talking, and not normative prescriptions setting out what we *ought* to do if we wish not to be among those whose behavior violates the standards of reason.

The following thought experiment illustrates the extent to which Hume has lost the notion of irrationality with his eschewal of the idea that there is such a thing as practical reason.<sup>21</sup> Suppose there is a person who (as a matter of fact) only acts on his occurrent motives (where these motives are defined either as desires or as whatever a Humean psychology takes to be a human being's motives), and knows this fact about himself. Such a person is unlike the notorious semi-moral agents criticized by Hobbes and Hume: Whereas Hobbes's "foole"<sup>22</sup> or Hume's "knave" believes "I should do that action which I can establish, using instrumental reason, it is in my interest to perform," this individual, whom I will call a "curmudgeon," believes "I will [not I should] only do that action which is prompted in me by the strongest occurrent motive." Note that whereas the fool's and knave's remark is a prescriptive principle setting out what counts as legitimate reasons for action, the curmudgeon's remark has no prescriptive overtones (he is not saying that he ought only to act on occurrent motives) but is merely setting out what he takes to be an accurate descriptive statement about how and why he behaves as he does.

Given their statement, the fool and the knave believe that they can have reason on occasion to perform actions for which they need have no immediate occurrent motive if and when these actions will causally effect states of affairs leading to the satisfaction of desires or preferences that they have. However, the curmudgeon will believe no such thing: if he, say, has no occurrent motive to go to the dentist, he will not go, even if he knows that such a visit would enable him to satisfy highly important self-regarding desires. A Kantian might contend that, were the curmudgeon to know such a thing, this knowledge would produce in him an occurrent motive to go to the dentist; but in order to illuminate aspects of the Humean view on instrumental reason, I will assume, as this view insists, that knowing that one has an instrumental reason is not sufficient to give one a motive to follow its directives. So the curmudgeon is someone who can know that he has a desire for *x*, and that *y* is a means to *x*, but will not, as a consequence of this knowledge, always have

an occurrent motive to do *y*. And when he does not, he will not do *y*.

Now suppose you saw this curmudgeon refusing to do *y* even though he admitted to wanting *x*, and you issued a hypothetical imperative to him to the effect that he ought to do *y* to achieve *x*. If, as a Humean instrumental theorist would insist, you took your hypothetical "imperative" to be merely a statement of a causal connection between an action and an outcome you believe to be desired by the agent, then if he ignores you after you have uttered it, you would merely shrug and turn away. After all, you would believe you could give him neither a reason nor a motive for acting otherwise. But if you would mean your hypothetical imperative to be a real *imperative* and not merely a statement of the causal facts, you would charge him with making a *mistake* if he didn't follow it, calling him "wrong" or "irrational." ("He *ought* to do *y*; he's irrational not to do it," you would insist.) But to respond to him this way is to attribute to him an objectively authoritative reason for action despite the fact that he refuses to recognize it, and to take yourself to be warranted to do so by virtue of the fact that you believe this reason applies to him whether he likes it or not. Indeed, Philippa Foot takes it that our criticisms of such a person should be understood in this way.<sup>23</sup>

So in essence you are maintaining that, no matter the peculiar psychological or social facts that explain his curmudgeonly behavior, it is normatively *necessary* (objectively) that he do *y* in order to achieve *x* in these circumstances—a reason which, at the moment, he does not recognize. Or to put it another way, when you say that he ought to do *y* to achieve his desired end *x*, then you believe, given that he wants *x*, and *y* is the means to *x*, that he has an objectively authoritative reason to do *y*. But this means you are judging him—and criticizing him—using a norm that you are assuming to be objectively authoritative (albeit perhaps not motivational by virtue of that authority), in just the way moral theorists take their moral norms to be valid. And as I said above, you are really appealing to the objective authority of what I have called the Instrumental Norm: "Act so as to perform the most effective means to a desired end." This norm is implicit in the fool's and knave's account of how they determine what action to perform, and it is the norm that you are using to attribute a certain reason for action to the curmudgeon, a reason which you believe he has necessarily, no matter his views about what actions he should, or will, take. If you believe—despite his disclaimers—that he is capable of acting from this norm, you are adopting the Kantian position on instrumental reason, attributing motivational efficacy to this authoritative instrumental norm. On the other hand, if you doubt its motivational efficacy, then you are embracing the Millian position, accepting the norm's authority, but admitting its motivational inadequacy in the circumstances. In either case, you will believe that, no matter his occurrent motives, he has a reason to do *y*.

But the Humean instrumental theorist has no room for this instrumental norm in his theory, because such a norm permits us to do what the Humean says is impermissible, namely, criticize actions as irrational. Accordingly, a Humean's hypothetical "imperatives" are merely assertions of causal connections, which in *no* sense give "reasons" for an agent to act. There is nothing either authoritative or motivationally efficacious about the Humean's hypothetical imperatives. And a Humean (curmudgeon-like) agent only follows them if, as it happens, the agent has a pre-existing motive to perform *y*, or if the agent is affected by the assertion such that she develops a motive to do *y*. In this case, the force of the 'ought' is explainable in terms of contingent psychological human responses.

### Is and Ought

Hume abandons the idea that there is practical reason, and thus the idea that actions can be condemned as irrational, because he understands, better than many contemporary proponents of instrumental reason, that even this (seemingly minimal) understanding of practical reason is still positing a kind of normativity that will be problematic for any naturalist. To say that the curmudgeon should have acted to secure the means to his end, *no matter what his occurrent desires were*, is to say that he is governed by an authoritative reason. And to say this, is to believe that there are "governing directives" in the world, to which we have access, and which bind us regardless of our occurrent desires. The idea that such directives exist—that there are things with "inherent prescriptivity" to use a term of J. L. Mackie<sup>24</sup>—has been derided by those who consider themselves moral skeptics. The idea that there is a "to be doneness" in the world, to which human beings are able to respond, is a hallmark of moral objectivism that its critics dismiss as too fantastic (from a scientific point of view) to take seriously. But what I believe Hume understood is that this same "to be doneness" is implicit in the idea of practical reason construed (merely) as an instrumental faculty. The hypothetical imperative has an 'ought' that is just as commanding as any categorical 'ought', and as the thought experiment with our curmudgeon makes clear, the 'ought' in that imperative is not the same as an 'is', and cannot be reduced or in some way "assimilated to" the desire assumed in the imperative.

Nor can it be "derived" from a cause-and-effect statement. Hume's famous dictum that you cannot derive an "ought" from an "is" has been forgotten by moral skeptics who believe, nonetheless, in the existence of an instrumental practical reason. While it may be true that *x* is the means to some end *y*, one cannot derive from it the claim that an agent if he desires *y* ought to do *x*. As Hume would put it, even if this change from "is" to "ought" is "imperceptible" (T 469) and ubiquitously made, it is nonetheless undefended and violative of naturalistic scruples every bit as much as moral oughts. The fashion for seeing

moral imperatives as hypothetical rather than categorical has assumed that naturalists are able to accommodate the hypothetical 'ought' in a way that they cannot accommodate the categorical 'ought'. But Hume implicitly understood that this is not so; even if hypothetical imperatives strike us as more congenial or more understandable by virtue of their connection with desires, nonetheless, insofar as they generate authoritative reasons for action, which "apply" to us no matter what our occurrent desires, then their prescriptivity is just as "queer" and problematic from a naturalistic point of view as the prescriptivity of categorical imperatives. Even if we believe that, by virtue of their connection to desires, hypothetical imperatives do bind us (in a way that we cannot believe of categorical imperatives, because there is no connection between them and our desires), nonetheless the fact *that* they are taken to *bind* us, is what Hume and any naturalist who does not believe there is "inherent prescriptivity" in the world, must reject.

Yet to give up our ordinary understanding of hypothetical imperatives as binding upon human agents is to defy common-sense, in a way that I believe even Hume found it difficult to do. There are signs in the *Treatise* that Hume could not sustain his commitment to this reason as (merely) a theoretical faculty, particularly in his discussion of the artificial virtues. I will give one brief example. In his discussion of justice as an artificial virtue Hume calls for men to restrain their self-interest so as to better satisfy it:

There is no passion...capable of controlling the interested affection, but the very affection itself, by an alteration of its direction. Now this alteration *must necessarily take place* upon the least reflection; since 'tis evident, that the passion is much better satisfy'd by its restraint, than by its liberty....(T 492; my emphasis)

This passage suggests that, as self-interested creatures, we are at the mercy of our reason to explain to us how best to satisfy our interests.<sup>25</sup> But what does Hume mean by "this alteration must *necessarily take place*" after reflection? Does he mean that it is psychologically inescapable? But how can he mean this, when he has already acknowledged in Book II that "Men often act knowingly against their interest" (T 418)? One gets the feeling he means that such an alteration "ought" to take place, and yet that would mean recognizing the authority of something like the instrumental norm (understood to be partially constitutive of reason) discussed in the last section. I suspect that Hume "slips" here because the way in which we normally understand reason includes the idea that it necessarily has authority over action when it supplies accurate cause-and-effect information regarding action—which is what thesis 1 maintains. Yet such necessary authority does not fit with the sort of naturalism to which any Humean is committed.



So if we want to be naturalists, is it impossible to have not only an objectivist moral theory but also any (plausible) conception of practical reason—including a conception of reason as (merely) instrumental? I have not said enough here to show that the answer to that question is “yes”—but I have argued elsewhere that this is indeed the answer that naturalist moral skeptics must accept.<sup>26</sup> Suffice it to say here that it has not been sufficiently appreciated by contemporary philosophers that naturalism is actually a very costly metaphysics, and one of its costs is that it may preclude us from accepting a sensible—and genuinely instrumental—view of reason.

## NOTES

I am grateful to the Social Philosophy and Policy Center at Bowling Green State University for supporting me during some of the time this paper was written. I would also like to thank Tom Christiano, Don Garrett, Ken O'Day, David Owen, and Terry Price for discussions about some of the issues raised in this paper. Finally, I would like to thank the Pew Foundation, whose grant to me through the Pew Evangelical Scholars Program, made much of the research for this paper possible.

1 That is, the agent's belief must satisfy epistemic norms regarding belief formation.

2 But I believe there are positions that are reasonably called instrumentalist that have a somewhat different characterization. See Jean Hampton, “On Instrumental Rationality,” in *Essays in Honor of Kurt Baier*, edited by J. Schneewind (La Salle: Open Court, 1996).

3 See Immanuel Kant, *Religion Within the Limits of Reason Alone* for his mature account of the sufficiency of moral motives, and the role of the human will in choosing these motives to prevail against any conflicting motives.

4 J.L. Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1977).

5 For one well-known discussion of the ontological commitments science seems to require of us, see W.V. Quine, “Facts of the Matter,” in *American Philosophy from Edwards to Quine*, edited by R.W. Shahan and K.R. Merrill (Norman: University of Oklahoma Press, 1977).

6 See Gilbert Harman, “Human Flourishing, Ethics and Liberty,” *Philosophy and Public Affairs* 12 (Fall 1983): 319.

7 Philippa Foot, “Morality as a System of Hypothetical Imperatives,” reprinted in her *Virtues and Vices* (Berkeley and Los Angeles: University of California, 1978), 167.

8 Mackie, 28, my emphasis.

9 See John McDowell, “Are Moral Imperatives Hypothetical?” *Proceedings of the Aristotelian Society* 52 (1978): 13

10 See Bernard Williams, “Internal and External Reasons,” in his *Moral Luck* (Cambridge: Cambridge University Press, 1980).

11 Jonathan Dancy, *Moral Reasons* (Oxford: 1993), Appendix I, 253–257. Dancy also notes the way in which Williams's internal/external distinction comes apart from a more traditional motivational way of making that distinction.

12 Immanuel Kant, *Foundation of the Metaphysics of Morals*, translated by Lewis White Beck (Indianapolis: Bobbs-Merrill, 1959), 33.

13 Kant, *Foundation*, 34.

14 So defined, the Kantian position is really a *family* of positions, and not a single view, for there are a number of ways of elaborating the idea that, to be rational, a person who acts from a hypothetical imperative is motivated by the authority of that imperative. For example, one can hold that the authority of hypothetical imperatives is *directly* motivational, in a way analogous to the direct motivational efficacy of the authority of a (moral) categorical imperative. Alternatively, one can hold that the authority of these imperatives motivates us indirectly, for example by having an effect on our psychological structure such that a desire to do what the imperative directs is created. Kant suggests this second idea himself in the section "The Incentives of Pure Practical Reason" in *Critique of Practical Reason*. Indeed, there are times in this section when he seems to assume that the only way to explain how reason could be motivational is to credit it with the power to effect a desire in us to do the moral action. But note that this view still credits reason with the causal power to create a motive, even if it does not credit it with the power to directly motivate. And both positions preserve what is fundamental to the Kantian view, namely, the idea that our motivation to follow the imperative is derived from the authority of the imperative, so that we are appropriately said to be acting *for the sake of* the reason given us in the imperative. If one holds the former view, one might argue that while it is possible for the authority of imperatives to work indirectly, it is better if it does not; i.e., it is better if we act directly from the reasons we have for acting. In contrast, one could maintain that reason does not demand that in all circumstances our motivation to follow its directives must come from its authority rather than, say, from our desires. For a defense of this latter view, see the work of Joseph Raz, especially his *Practical Norms and Reasons* (Princeton: Princeton University Press, 1975 and 1990), and the concluding postscript to the second edition.

15 Mill distinguishes between authority and motivational efficacy in the context of presenting his moral theory in *Utilitarianism*. In chapter 1, after setting out his principle of utility, Mill insists that we should not mistake the (compelling) meaning of a moral principle with the *motive* one might have to obey it. In chapter 2, he goes on to explore the motives we might have for following this principle, concluding that they are largely the feelings of conscience of those who have been raised to have them, and in chapter 3, entitled "How Proved" he embarks upon the task of constructing a proof to establish that only *this* principle constitutes all of morality. But that proof (which generations of students know is problematic) is *not*, for Mill, the source of our motive to follow the principle, nor do our motives justify the principle as the correct standard of moral action. So Mill sharply distinguishes what *gets* us to be moral, from the content and compelling force of morality. A Millian position on instrumental reason distinguishes, in the same way, between what

gets us to follow the directives of instrumental reason, and the authority of instrumental reason.

16 But see my "On Instrumental Rationality" where I argue that the Millian position is not acceptable from the naturalist's point of view.

17 For discussions on Hume's position on reason, I am greatly indebted to Tom Christiano and Don Garrett.

18 While he notes that we sometimes call a passion irrational, strictly speaking it is the false judgement attending the passion that is irrational, and not the passion itself; see T 416.

19 Note that the authority of this instrumental norm has to be understood non-instrumentally; because it is the foundation of the idea that we ought to act on means appropriate to the achievement of our ends, it cannot itself be defended consequentially. Thus, understood as an imperative, it is categorical and not hypothetical (a delightful bit of irony in this story).

20 The importance of conventional rules determining the content of natural virtues and vices is an implication of Hume's discussion in T 581–582.

21 See my "Rethinking Reason," *American Philosophical Quarterly* 29 (3): 219–236.

22 See Thomas Hobbes, *Leviathan*, chap. 15.

23 In her criticism of the amoral man, Foot says, "If he himself is an amoral man he may deny that he has any reason to trouble his head over this or any other moral demand. Of course he may be mistaken, and his life as well as others' lives may be most sadly spoiled by his selfishness" ("Morality as a System of Hypothetical Imperatives," 167). Note that she stresses that he would be *mistaken* insofar as the behavior would "spoil" his life, indicating that she believes regardless of what his desires with respect to his behavior happen to be, such a person will only be acting rightly if he acts instrumentally.

24 See Mackie, chap. 1.

25 I am grateful to Ken O'Day for pointing out how this passage violates the position that reason is only theoretical.

26 The arguments are in chap. 4 of my *For the Sake of Reason* (forthcoming, Cambridge University Press); and see "On Instrumental Rationality."

*Received May 1994*