

# Against All Reason? Skepticism about the Instrumental Norm<sup>1</sup>

Stephen Finlay

A naturalistic project descended from Hume seeks to explain ‘ought’ and normativity as a product of motivational states such as desires and aversions.<sup>2</sup> Following Kant, rationalists reject this thesis, holding that ‘ought’ rather expresses a command of *reason* or intellect independent of desires. On Hume’s view the only genuine form of practical reason is theoretical reason operating in the service of desire, as in calculation of means to ends. Reason at most *discovers* normative requirements, which exist through the interrelation of subjective desires and objective world. The Humean desire-dependence view of the source of normativity is commonly associated with *instrumentalism*, an influential theory of normative content according to which agents ought always and only to act so as to optimize satisfaction of their own desires. But rationalists (including Thomas Nagel, Jean Hampton, and Christine Korsgaard) have recently argued that proponents of desire-dependence are not entitled even to this instrumentalist ‘ought.’<sup>3</sup> Instrumentalism holds that all normativity derives from the *instrumental norm*: approximately, the principle that one ought to take the means to one’s ends, or

$$(\forall \text{ agents } x, \forall \text{ ends } y, \forall \text{ actions } z) ((x \text{ wills } y \ \& \ z \text{ is the means to } y) \rightarrow x \text{ ought to do } z).$$
<sup>4</sup>

---

<sup>1</sup> The publication of this paper is unfortunately untimely. Since it was originally written (2000-2003), there has been an explosion of scholarly work on instrumental normativity. But while the progress in the dialectic makes some of this paper’s emphases and omissions seem odd, none of the recent literature duplicates my viewpoint or weakens my confidence in it. I owe thanks to Gary Ebbs, Gideon Yaffe, David Copp, the participants of the ‘Hume, Is and Ought’ conference at the University of Otago in February 2003, some anonymous referees, and others whose identities I regret that I have forgotten.

<sup>2</sup> The existence of motivation-external ‘shoulds’ seems indubitable, but Humeans maintain they are also normatively external.

<sup>3</sup> See also Wallace 2003.

<sup>4</sup> Individual ‘hypothetical imperatives’ are obtained by instantiating *y* with a particular end *e*, and *z* with a particular action *a*; if my end is *e*, and *a* is the means to *e*, then I ought to do *a*.

The normativity of the instrumental norm, rationalists observe, cannot itself be dependent on any particular desires. Whatever I desire, I should acknowledge its normative authority and take the means to my ends.<sup>5</sup> Rationalists therefore charge that instrumentalism itself assumes a practical principle with normativity not derivable from any desire, but commanded purely by reason itself.<sup>6</sup> This provides leverage against resistance to other alleged rational requirements such as Kant's Categorical Imperative: once we acknowledge one command of pure reason, we have no principled ground for skepticism about the possibility of others.<sup>7</sup> In any case, once we concede the desire-independent normativity of an instrumental norm of reason, the desire-dependence thesis is compromised.

Is the desire-dependence view of normativity therefore hopelessly inconsistent, denying the possibility of rational norms while assuming the rational authority of the instrumental norm? Hume himself, unlike those who claim his inheritance, is said not to accept the instrumental norm.<sup>8</sup> While I am unaware that he ever explicitly denies it to be contrary to reason to fail to pursue means to our ends,<sup>9</sup> this is clearly entailed by his infamous rejection of practical irrationality *per se*:

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.... Now 'tis evident our passions, volitions, and actions, are not susceptible of any such agreement or disagreement.... 'Tis impossible, therefore, they can be pronounced either true or false, and be either contrary or conformable to reason. (T458)

---

<sup>5</sup> A universal desire (perhaps explicable in evolutionary terms) to take the means to our ends is sometimes proposed in explanation. But any such desire, and hence consequent normativity, could only be contingent, while the rational requirement to take the means to our ends clearly is not.

<sup>6</sup> 'However contingent the hypothetical 'ought' is on a desire, it is still *not* the same as a desire; to say, therefore, that its objective normative authority is what moves us to act rationally is to analyze the 'prescriptive force' of hypothetical imperatives such that it is identical to the prescriptive force of categorical imperatives. If naturalists reject the plausibility of this analysis for categorical imperatives, they must do so for hypothetical imperatives' (Hampton 1998: 162-3). See also Wallace 2003.

<sup>7</sup> Some, such as Korsgaard and Hampton, argue more ambitiously that if we are to formulate a coherent instrumental norm then we must assume the existence of further rational norms. This is a case of their *modus ponens* being my *modus tollens*.

<sup>8</sup> Hampton 1995, 1998; Millgram 1995, Wallace 2003.

<sup>9</sup> Hampton (1995) claims he does with his infamous claim, "'Tis not contrary to reason...to prefer even my own acknowledg'd lesser good to my greater," (1888: 416) interpreting 'greater good' as meaning *means to the more preferred end*, and 'interest' as meaning *end*. This leads her uncharitably to find Hume guilty of direct self-contradiction about whether we can fail to take the means to our ends.

Hume is a radical and *consistent* champion of the desire-dependence thesis, who denies the existence of any rational norm. This skepticism is claimed to be *absurdly* radical; the existence of the instrumental norm is a dogma in contemporary ethical theory, and many declare its denial simply unthinkable.<sup>10</sup> R. Jay Wallace writes that ‘In the modern era, this form of rationality has been widely viewed as the single unproblematic requirement of practical reason’ (2003). Hampton claims that Hume’s view of reason “defies common sense” (1995: 57). To champion the desire-dependence thesis consistently, one must adopt the heresy of Hume, which no right-minded person could.

It is therefore apparently against all reason that I seek to defend Hume and the desire-dependence thesis, by arguing that the instrumental norm is a mere phantasm of our confusion. There is no desire-independent normative principle commanded by reason (or anything else) that we ought to take the means to our ends, and all the intuitions and phenomena that seem to lend the idea credence are explicable in other, Humean ways. My argument will be that no formulation of the alleged norm can satisfy two basic conditions for a genuine normative command of reason. First, there is the *possible violation* criterion. As Kant saw, for there to be a principle that can tell me what I should do and thereby counsel, guide, or command me, it must be (at least logically) possible for me to act otherwise than as it directs. Second, there is the *reasonable expectation* criterion. Unlike a putative normative requirement springing from the will of a deity, or from the metaphysical nature of things, a requirement arising from pure *reason* (hence, as Kant claims, *a priori*) must be evidently reasonable. I must, in exercising my reason, be able to recognize the legitimacy of such an expectation, and it must be impossible to sustain an intelligible challenge to it. This is implicit already in the

---

<sup>10</sup> For example Dreier 1997: 90-4; Korsgaard 1986: 319, 1997: 215. Nicholas Sturgeon writes of Hume’s skepticism concerning practical reason, ‘His “strict and philosophical” account of reasonableness and unreasonableness...is not worth taking seriously, and properly receives almost no attention at all....It is a measure of my respect for Hume’s intellect that I find it hard to believe that he took it seriously either’ (quoted by Elijah Millgram [1995: 87] from an unpublished paper).

view that conformity to the instrumental norm is so self-evidently required that to reject it is absurd.<sup>11</sup>

Section I aims to clarify the dictate of the alleged norm. Section II presents my argument debunking it. Section III critically examines alternative rationalist formulations, and Section IV reinterprets the phenomena that have been taken as evidence for the norm.

## I

What does the instrumental norm command, and why should we obey? It is surprisingly difficult to formulate, given its alleged firm foundation in common sense, but it is important that we do so painstakingly, as the rationalist case for its existence trades upon ambiguity and indeterminacy. My method is to start with the simplest, most commonplace formulation, and progressively refine it. We commonly utter statements of the form ‘If you want  $y$ , you should do  $z$ ,’ where doing  $z$  is a means to the end  $y$ .<sup>12</sup> I shall mean by ‘ends’ sets of possible states of affairs, and by a ‘means’ to an end a possible course of action, or necessary part thereof, that may result in a member of that set obtaining. My first approximation is then

$$\text{IN}(1): (\forall \text{ agents } x, \forall \text{ ends } y, \forall \text{ actions } z) ((x \text{ desires } y \ \& \ z \text{ is a means to } y) \rightarrow x \text{ ought to do } z).$$

Three elements need refinement. I will first address the nature of the means-relation between action and end, then the proper formulation of the consequent imperative (what is it one ought to do if the conditions specified in the antecedent obtain?), and finally the proper relation between agent and end.

---

<sup>11</sup> Satisfying this criterion is not *sufficient* for normativity, as the requirement that one act in a way logically or psychologically necessary would certainly qualify as reasonable, and would be invulnerable to intelligible challenge.

<sup>12</sup> Here, ‘You should do  $z$ ,’ which I provisionally shall follow rationalists in supposing to report a command of reason, occupies the position of the consequent following the hypothetical antecedent, ‘If you want  $y$ ’: hence, a ‘hypothetical imperative’.

According to IN(1), if an action is a means to my end, I ought to perform it. But there are many logically possible means to accomplishing some ends that are physically impossible for a given agent. Are there superhuman feats I ought nonetheless to perform? Requiring this for rationality clearly violates the reasonable expectation criterion. Reasonable criticism must be sensitive to agents' situational abilities, hence the instrumental norm must command only means that are *performable* by the agent. Epistemic circumstances are also significant. On IN(1) if I will an end to which, unbeknownst to me,  $\phi$ -ing is the means, and I fail to  $\phi$ , I violate the instrumental norm. But I would more appropriately be called irrational if I were to  $\phi$ , not believing that I thereby further my ends.<sup>13</sup> The norm must only command me to action that I *believe* to be performable means to my ends.

IN(1) tells me that if some action is a means to my end, I ought to perform it. But there can be many alternative means to attaining one and the same end, and it is unreasonable to demand that I perform them all (and often counterproductive to try). The instrumental norm must only direct me to take *one* sufficient means to my end. One suggestion is that the norm directs me to take the *best* means, but are we to maximize moral acceptability, aesthetic appeal, likelihood of success, or ease of execution, etc.? The most plausible candidate would be efficiency (*instrumentally* best). But a requirement of efficiency can be accommodated by IN(1) since we never have just one end, so that other ends affect reason's counsel in regard to means. If a more substantive account of efficiency is offered, we may observe that efficiency is not the only virtue of action, and it would be unreasonable to require that we always act in the most efficient manner. I here presume that I can only be justly accused of instrumental irrationality if I fail to take any

---

<sup>13</sup> Korsgaard, however, discusses at one point "people acting irrationally only because they do not know about the relevant means/end connection" (1986: 380).

means sufficient to my end<sup>14</sup> (this is certainly suggested by the standard examples), so we should reject evaluative criteria in favour of modal ones. I must take (what I believe to be) the performable means *necessary* for attaining my ends. Isn't it common, however, for an end to have several possible, hence no necessary means? In this case there is a necessary *disjunctive* means (e.g. *d*, where *d* = *a or b or c*, and *a*, *b*, and *c* are the only alternatives); here the instrumental norm commands performance of *d*. Combining these refinements we get

IN(2):  $(\forall \text{ agents } x, \forall \text{ ends } y, \forall \text{ actions } z) ((x \text{ desires } y \ \& \ \mathbf{x \text{ believes } z \text{ to be the necessary performable means for achieving } y}) \rightarrow x \text{ ought to do } z).$

Let us now consider what is commanded if the antecedent conditions are met. First, note that IN(2) commands *performance* of the action *z*, hence if the antecedent conditions hold of me and I fail to perform *z*, I am instrumentally irrational. But this is unreasonable. Suppose I try to perform *z*, but my efforts go awry: my shoelaces being tied together, I instead trip and fall. I may then be guilty of clumsiness, incompetence, or ignorance, but not of irrationality. Let us say for now that the instrumental norm commands not that I perform, but that I *try* to perform *z*.

Time introduces further difficulties. I can desire an end without desiring that it obtains at that moment. I may desire it for some future time, or for the nonspecific future. The appropriate means to our ends therefore need not be *presently* performable. I need not choose among presently performable means if I anticipate that another (perhaps preferable) means will become available later. IN(2) does not say when I ought to try to perform *z*, but it cannot be at *any* time. If I will the end on Monday, but the

---

<sup>14</sup> This may be mistaken; arguably it is instrumentally irrational to take an inferior means over a superior means without good reason. This indicates an avenue for rationalists to escape the arguments of this paper, although given space I would extend my case to argue that even this form of behavior is not possible. Even the more limited result is significant, however; defenders of the principle usually formulate it in terms of necessity (e.g. Wallace 2003), some defences of it cannot easily be extended beyond necessary means (e.g. Setiya forthcoming), and the classic examples of instrumental irrationality on offer all involve failure to take any sufficient means.

means  $z$  is the action of  $\phi$ -ing on the following Thursday, it would be absurd to counsel me on Monday to attempt to perform  $z$  on Monday. The ‘when’ of the imperative may then appear to be the time of the chosen means. But suppose I desire the end on Monday and not thereafter, and the means is an action on the following Thursday; ought I on Thursday to perform the means to an end that I no longer desire? This is an unreasonable expectation: I am at a time  $t$  subject to instrumental requirements only in virtue of ends that I still have at  $t$ . I provisionally amend the norm as commanding me to try to perform at  $t$  an action  $z$  if I believe that performing  $z$  at  $t$  is possible, and necessary for realizing an end I desire at  $t$  (an unsatisfactory solution, as we shall soon see).

There is, further, a problem with the very suggestion that this norm is ‘hypothetically imperative’ in form. With a genuine conditional, *If A then B*, satisfaction of the antecedent condition  $A$  entails the unqualified truth of the consequent. If my end is to give my wife a sapphire ring for our anniversary, and the only performable means is to steal one, then on IN(2) it is unqualifiedly the case that I should steal one. *Reason* commands that I do so. This can’t be right.<sup>15</sup> We often set ends without appreciating what their attainment will require, only later realizing that performing the necessary means is unthinkable. The reasonable thing to do is to give up the end. Of course, by surrendering the end we make the antecedent false: don’t we thereby escape the imperative? But in doing so, on IN(2), we are violating the command of reason, which doesn’t give the option of surrendering the end. The reasonable expectation criterion is not satisfied.

---

<sup>15</sup> See also Broome 2000: 89-90.

Some rationalists therefore rightly hold that the instrumental norm counsels us *either* to perform the means *z*, *or* to give up the end *y*.<sup>16</sup> This disjunctive imperative yields our third formulation:

IN(3): ( $\forall$  agents *x*,  $\forall$  ends *y*,  $\forall$  actions *z*,  $\forall$  **times t**) ((*x* **at t** desires *y* & *x* **at t** believes *z* to be the necessary performable means **at t** for achieving *y*)  $\rightarrow$  (*x* ought **at t to try** to do *z* **or cease desiring y**)).

This solves the preceding problems. Reason never demands that I take immoral or intolerable means to my ends; rather, that I pursue what I believe to be the necessary means to my end or else give up that end. Instrumental irrationality is the offense of not adopting what one believes to be the necessary means to an end while maintaining that end.

This brings us to the most difficult question: the character of the relation between agent and end. So far, taking my cue from an ordinary turn of speech, I have represented the relation as *desiring – being desired*. Am I indeed subject to hypothetical imperatives of reason as a consequence of my desires? Humeans who admit an instrumental norm often favour this formulation, and even Kant sometimes wrote this way.<sup>17</sup> But I have desires of which I myself disapprove, and believe should *not* be satisfied. It is unreasonable to expect that I should pursue the means to an end I desire, when I should not desire the end. (Of course IN(3) also gives me the option of ceasing to desire the end, but desires cannot be eliminated simply by *choice*. It would be unreasonable to call me instrumentally irrational because I continue to have such desires.) Perhaps an improved formulation would counsel pursuing means to the ends of only a *certain kind* of desire: endorsed desires, second-order desires, fundamental valuing, etc. Donald Hubin writes,

---

<sup>16</sup> Korsgaard 1997: 237, Hill 1973: 24, Broome 2001: 180. Schroeder (2004) reveals further options: eliminating the efficacy of the means, or ceasing to believe that the action is a means. I ignore these here, because no-one suggests that these are rational courses of action.

<sup>17</sup> For example 1997: 18.

[A means/ends irrational] agent has ends that he values intrinsically, recognizes that certain actions are necessary means to these ends but never feels any motivation to perform the acts in question (1996: 40).

But there is a general problem with formulating the norm in terms of any kind of desire. Suppose I intrinsically value world peace, but believe there is no means to it except the slaughter of half the world's population. I therefore despair of realizing my ideal; must I also cease valuing it? IN(3) says I must, on pain of irrationality. But this is not a reasonable expectation! (It may even be impossible).

It does not follow from my having a desire that I rationally ought to seek its satisfaction.<sup>18</sup> As Kant intended, it is when I *will an end*, not when I desire something, that there are actions I instrumentally ought to perform. The notion of 'willing an end' is however significantly vague. We know at least that it contrasts with desiring something, which doesn't obviously require anything of us. Desires can be idle, unendorsed, and merely wishful, while having an end involves resolve and commitment. This difference needs greater articulation. What is 'willing'?

Rationalists believe that willing is one critical element of action neglected by Hume and those who would reduce deliberation to the interplay of rival desires. Voluntary action requires states and actions of intellect, as well as (perhaps even in place of) conative states, and willing is to be identified with some such intellectual state or action – such as *decision*, *volition*, or *intention*. So which, if any, of these modes of intellectual activity is 'willing'?

When we decide to act on an inclination – to do a desired action or seek a desired end – then its object becomes an object of volition. (Korsgaard 1997: 234)

Here decision plays a pivotal role: should we identify willing with deciding? But to *decide* (or resolve or commit oneself) to realize an end is to perform a momentary, temporally discrete action. The end that I decide upon and the means I believe necessary

for attaining that end may be temporally distant, so deciding upon some end does not always precipitate nor require immediate action. It cannot therefore be merely *when* I decide upon an end that I will it, and the antecedent of the instrumental norm should not specify the *act of deciding* as the pivotal relation between agent and end, or the norm will fail to command me after my decision is made. Korsgaard's point is more plausibly that willing is a product of decision, which is the process of arriving at the intellectual state of willing. We could, then, identify willing as the state of *being decided* (resolved, determined, committed) upon an end – but here this would be circular and uninformative. The concept of decision does not help to isolate the meaning of 'willing.'

Korsgaard clearly suggests *volition* as the product of decision – perhaps this is our quarry. (Indeed, etymology favours identification of will with volition.) The concept of volition is that of aiming at literal or figurative motion by an exercise of intellect. To 'will' an end would then be to exert causal influence towards self-activity by power of thought – it is to actually *aim at, seek, or pursue* that end. (Note this is not equivalent to being causally determined to activity by thought, since volition can fail.)<sup>19</sup> Clearly rationalists, including Kant, often mean by 'will' precisely this. But so interpreted Korsgaard's claim is false; decision does not always result in volition. Deciding tonight to drive to work tomorrow morning does not generate volition tonight (and could even fail to do so tomorrow). Commitment to my decision requires volition only when morning arrives. Evidently another mental state is at work, one acquired when a decision is made, and able to produce volition when conditions are right. Shouldn't we call this, also,

---

<sup>18</sup> As Ramon Das observed, I haven't shown that there isn't a rational requirement to seek the satisfaction of my desire *in the absence of opposing reasons* – perhaps IN(3) could be fixed with a *ceteris paribus* clause. My rationalist opponents, however, will concede my claim.

<sup>19</sup> Volition can fail to produce action; a creature unaware of the amputation of a limb may still attempt to move it, and psychological states like terror may immobilize our muscles so that volition is ineffective. These phenomena do not constitute instrumental irrationality, as here the means *are* willed, and it is unreasonable to demand action where there is physical incapacity.

willing an end? Since rationalists insist that one can will an end while being irrationally unmoved, they cannot mean to identify willing with volition.

The product of decision and cause of volition is the mental state of *intending* to realize the end. Intention is the most plausible candidate for a state to identify with ‘willing’: it incorporates the commitment to the end that makes plausible the claim of normative requirement of the means, and intending an end does not entail present volition towards that end.<sup>20</sup> This also resolves a problem regarding what the instrumental norm counsels when the conditions of its antecedent are met. Any formulation that counsels me to *perform*, *attempt to perform*, or *seek* the means to my ends has trouble with temporally dislocated means and ends, as previously discussed. I provisionally suggested we interpret the norm as counseling merely that I attempt to perform any means to my ends that are both *presently performable and necessary*, but this is incompatible with disjunctive means (performance of which cannot be required at a specific time if the disjuncts specify different times), and entails that the instrumental norm gives no counsel about pursuing our ends until all but one of the possible means have been declined. Those who refuse to put off today what can be done tomorrow are then neither following nor flouting the counsel of instrumental rationality, whereas they appear rather to be following one disjunct of its advice.

The dictum, ‘whoever wills the end, *wills* the means’ invites us rather to consider intention for the imperative as well as the antecedent. My ends require that I *intend* to perform the means.<sup>21</sup> This can be required when I will the end on Monday, even though the means are not performable until the following Thursday, and I am not required to pursue on Thursday the means to ends erstwhile abandoned. We can also eliminate talk

---

<sup>20</sup> Henry Sidgwick writes, “the ‘hypothetical imperative’...prescribes the fittest means to any end that we may have determined to aim at.... The adoption of an end as paramount... is quite a distinct psychical phenomenon from desire: it is a kind of volition, though it is, of course, specifically different from a volition initiating a particular immediate action. As a species intermediate between the two, we may place resolutions to act in a certain way at some future time.” (1907:37)

of *attempting*. Since intention is a matter of attitude rather than performance, it is not a condition on my intending to  $\phi$  that I successfully  $\phi$ . We then get:

IN(4):  $(\forall$  agents  $x, \forall$  ends  $y, \forall$  actions  $z, \forall$  times  $t)$   $((x$  **intends** at  $t$  **to achieve**  $y$  &  $x$  at  $t$  believes  $z$  to be the necessary performable means for achieving  $y) \rightarrow (x$  ought at  $t$  **to intend** to do  $z$  or cease **intending to achieve**  $y))$ .

A noteworthy objection is that a norm commanding only intention and not attempted action fails to proscribe an important form of instrumental irrationality. Suppose I believe the necessary means to my end is to  $\phi$  at  $t_i$ . On IN(4), to comply with the counsel of reason at  $t_i$  it is sufficient that at  $t_i$  I intend to  $\phi$  at  $t_i$ , and not necessary that I attempt or have volition to  $\phi$  at  $t_i$ . The strange consequence is that failing to pursue what I believe to be the necessary means to my end is not instrumentally irrational. The solution is to recognize the intimate connection between intention and volition. To intend in the present to  $\phi$  in the present *is* to have volition to  $\phi$  in the present. A gap between intention and volition only opens through temporal distance between intending and what is intended. Therefore, if in the present I do not attempt to  $\phi$ , it follows that in the present I do not intend to  $\phi$  in the present.

## II

With IN(4) we have at last met the reasonable expectation criterion. Obviously I instrumentally should intend to do what I believe necessary for achieving the result I intend to bring about – and it is *prima facie* plausible to say that reason itself demands this of me. Indeed, I believe IN(4) successfully captures what we have in mind when we think of an instrumental norm of reason. To be a genuine normative principle, however, it must also satisfy the possible violation criterion. Instrumental irrationality, as defined by IN(4), is the failure to intend to do what one believes necessary for achieving some end, *while* intending that very end. Is such irrationality possible? Kant himself seems at

---

<sup>21</sup> See also Broome 2000, 2001 who concurs with this formulation of the instrumental norm.

times to maintain that the principle, whosoever wills the end wills the necessary means, is an *analytic* hence necessary truth. Failure to comply would then be impossible – but a normative principle that cannot be transgressed is no normative principle at all, *per* Kant and the possible violation criterion. This interpretation of Kant is mistaken: elsewhere (e.g. 1911: 417) he clearly explains that the connection is analytic only insofar as we are rational.<sup>22</sup> Since we are capable of irrationality we can violate the command of reason and fail to pursue the necessary means to our ends.

I believe Kant's principle is, nonetheless, analytic without this qualification.<sup>23</sup> Observe that the expression, 'to intend the end *y*,' is significantly incomplete. Intention has to do with action: what is intended is always action (or inaction) of some sort. To intend the end *y*, therefore, is really to intend *to act* so as to bring the end *y* about.<sup>24</sup> Now observe that the expression, 'to intend the necessary means *z*,' is similarly incomplete. Talk of the 'necessary means' implicitly invokes the end, so that to intend the necessary means is to intend to act *so as to bring the end z about*. Intending the end, therefore, simply is intending the necessary means to that end – the two locutions give partial and complementary descriptions of the same intention, so one logically cannot intend an end without intending the necessary means.<sup>25</sup> This result is significant, but I'm seeking something more. Suppose I believe that the necessary means to my intended end *y* is to do *z*. We have established that my intending the end entails that I intend the necessary means. Doesn't it therefore follow that I intend to do *z*?

---

<sup>22</sup> Korsgaard 1997: 236, Hill 1989: 19.

<sup>23</sup> See also von Wright 1978.

<sup>24</sup> It is has been objected that an intention (say) to kill Bob is not identical to an intention to act to bring it about that I kill Bob; the former intention need not include a commitment to perform any merely instrumental means. But such an intention is really something quite different: a *conditional* intention, to kill Bob if a certain kind of situation arises. The unconditional intention to kill Bob involves a commitment to bringing it about that I kill Bob.

<sup>25</sup> This much is admitted by Broome, who nonetheless maintains the possibility of such irrationality (without argument, that I am aware). He writes, "When you intend to  $\phi$ , a causal process is in train that you believe will result in your  $\phi$ -ing...To form the intention of  $\phi$ -ing, you must set such a process in train. If there is no process you can set in train that you believe will result in your  $\phi$ -ing, you cannot form the intention." (2001: 187)

Actually it doesn't – not straightforwardly, anyhow – because of the opacity ('intensionality') of intentional contexts. The content of intention, like other propositional attitudes, does not permit the substitution of identicals *salva veritate*. Suppose I aim to spot the kitten on the floor so as to avoid stepping on it in the dark, and believe that turning on the light is the necessary means. I also believe that the action of turning on the light is at the same time the action of startling the fish. I intend to turn on the light – do I also intend to startle the fish? Clearly not, even though I believe the actions to be one and the same. Intending A and believing that A=B does not entail intending B. Therefore what follows from my intending the end *y* is merely that I intend the necessary means to *y de dicto*, and not that I intend the necessary means *de re* (i.e. to do *z*). Surely (one says) I rationally *ought* to intend to do *z* – but the claim I am defending is rather that I metaphysically *must*. Isn't this implausibly to insist upon the impossibility of failing to acquire a new intention on identifying the means?

It might be enough for my purposes to point out that we've established an analytic requirement of intending the necessary means. If I intend the end *y* at *t*, and believe that doing *z* at *t* is the necessary means, then (given the identification of volition with present intention for the present) it follows that I will knowingly do *z* (barring failure) even if I don't intend to do *z*.<sup>26</sup> But I believe we can, further, show an analytic requirement to intend to do *z*. To do so, we have to look beneath the principle of the opacity of intention, at its explanatory ground.

The opacity of intention is often glossed as a matter of being 'description-sensitive.' We must note, however, that it is not strictly speaking *descriptions* – i.e. sentences – to which opacity is sensitive. My intention to turn on the light is also the intention to illuminate the room by flipping the light switch. The description reports

---

<sup>26</sup> In a weaker sense I can even be said to do it intentionally. Unlike thrashing in an epileptic fit, which I may do knowingly, my action is still intentional.

some aspect or element of the action, which permits of multiple descriptions, and opacity is sensitive to this *content* of description. Every action can be characterized in indefinitely many ways, according to the impact that action has upon different parts of the world (e.g. illuminating the room, startling the fish). Intentions can be thought of as agents' automatic targeting – i.e. they involve dispositions for one's deliberative endeavours to *track* a certain content. This explains the general opacity of intention: my target is illuminating the room, not startling the fish. As circumstances change I will deliberately track the goal of illuminating the room, but not the goal of startling the fish. Hence we say that I do not intend to startle the fish, although I do perform that action. An intention, then, will entail some further intention *provided* the content of the further intention is appropriately connected to the content of the first.

The aspect of being a necessary means to an end (e.g. to spotting the kitten) is very different from the aspect figuring in the description of the means itself (turning on the light). In this example, the former characterizes the action in terms of its impact on my awareness of a part of my environment, the latter characterizes the action in terms of an impact directly on the environment itself. There is no connection of constitutivity here, therefore – the connection is rather causal. The impact of illuminating the room is causally responsible for the impact of my spotting the kitten. To defeat opacity we need more than just a causal connection; illuminating the room is causally responsible for startling the fish, but my intention for the former doesn't translate to an intention for the latter. However, consider the significance of my belief that turning on the light is the necessary means to my seeing the kitten. A belief identifying an action as a means is not *simply* an identification of an action with itself under two separate aspects. We would *not* say that my startling the fish is the necessary means to seeing the kitten, even though my startling the fish *is* the same action as turning on the light. Rather, a means-identification indicates the aspect of an action which is causally significant *as* a means.

So long as I have this belief in the necessity of going through the event of my turning on the light in order to reach the event of my spotting the kitten, my deliberative efforts to perform the necessary means to spotting the kitten will of necessity track the goal of turning on the light – which simply means (given the identification of intending with tracking) that so long as I have this means-belief, my intention to perform the necessary means is an intention to turn on the light. So long as I believe that the necessary means to my intended end is *z*, I do indeed intend to do *z*.

The connection between intending the end and intending what I believe to be the necessary means to the end is analytic. On IN(4), if I am not willing what I believe to be the necessary means, then I am not willing the end, hence the possible violation criterion is not met. We have a logical rather than a normative law (*per* Hume, what else should we expect from pure reason?) Nonetheless, I believe that IN(4) does indeed articulate our intuitive notion of a rationally required instrumental norm. It is precisely when – and only when – I aim at or intend an end that the expectation that I aim at or intend the means appears incontrovertibly reasonable: because it is a requirement with which I cannot fail to comply! There is no such thing as an instrumental norm of reason, because Hume is right: pure reason never commands. The illusion that the idea has substance, even that its denial is absurd, is a conjuring trick, rationalist sleight-of-hand. Rationalists offer a vague formula (‘willing an end’) admitting of differing interpretations. Some satisfy the possible violation criterion, others the reasonable expectation criterion, but none satisfy both. On one interpretation, IN(4), it is analytic that if I will the end, I will what I believe to be the necessary means to the end: our grasp of this yields our impression that the principle derives purely from reason. But the world cannot violate the laws of logic, and so to give the impression that we have a *normative* principle, rationalism substitutes different interpretations.

### III

Rationalists must resist this claim of analyticity: it is possible to violate the instrumental norm – indeed, it is a common occurrence. There are two strategies I must address: first, one might accept IN(4) as an adequate interpretation of the instrumental norm, but claim that my arguments have failed to prove that it cannot accommodate possible violation: one can indeed fail to intend what one knows to be the necessary means to one's intended ends. Second, the rationalist can deny that IN(4) adequately captures the instrumental norm. I shall attend to these objections in turn.

Surely, it may be insisted, such irrationality as proscribed by IN(4) *is* possible. The rationally required means/end connection can be broken by a number of familiar influences, such as “rage, passion, depression, distraction, grief, physical or mental illness” (Korsgaard 1986: 378). To support the claim, rationalists appeal to a wide variety of alleged examples of instrumental irrationality: Korsgaard offers the case of a woman who wills that she overcomes her fear of riding roller-coasters, but walks away at her moment of truth, overcome by terror (1997: 228-9). I shall instead consider the modern paradigm of instrumental irrationality, the addicted smoker who, despite being resolved to kick the habit, succumbs to addiction and smokes another cigarette. I ought not deny that such things occur, nor that these people deserve criticism. But are they genuine cases of violation of IN(4)? For that to be the case, they must both be aware of the necessity of the means, and continue to will the end.

The problem for the rationalist case is that the relevant behaviour calls for interpretation, and there are alternative, arguably more plausible analyses of such weak-willed actions. Is the smoker passing up what he believes to be a necessary means to his end? The means under consideration involves abstaining from a cigarette on this particular occasion. But he could give up *tomorrow* instead of today and still achieve his end – smoking *one more* cigarette is compatible with his end. It is partly this knowledge,

that the means at hand is not *necessary*, that makes his addiction so hard to escape. The smoker doesn't violate IN(4), therefore, unless we modify the example. Suppose he resolves *never* to smoke another cigarette, effective immediately (at *t*). When later tempted to light up, he knows that doing so is incompatible with attaining his end. I ask: can we conceive of his succumbing to temptation *while still intending* never to smoke again after *t*?

I find myself incapable of imagining this. However other, plausible diagnoses come readily to mind: in choosing to smoke, he abandons his end, perhaps forming a new one ("*this* will be the last, I swear"), but perhaps resigning himself in self-disgust never to escape his addiction. He may try to deceive himself that he hasn't abandoned his end ("it's still before *t* in the Pacific time zone"), which would be to self-deceptively change his end. He may well continue to *wish* for the end, but this is a different matter, and involves no violation of IN(4).

What about the usual suspects – "rage, passion, depression" etc? One common effect of both violent passion and depression is the abandonment (even if only temporary) of previously stable ends. When, in Korsgaard's example, the woman turns in terror away from the roller coaster, she seems in no way still resolved upon riding it (in the present). Another common effect of violent passion is to make us oblivious to things we generally know. Indeed, four out of the five ways Korsgaard suggests these mental perturbances make us irrational involve causing us to be unaware of our reasons for acting! (1986: 378n) But, most plausibly, this is to say that in those circumstances *we are unaware that our actions are incompatible with our ends!* This would therefore be, at best, a form of *theoretical* irrationality, but not of practical irrationality, which must be assessed relative to a agent's epistemic limitations. The case for genuine violations of IN(4) depends upon the possibility of *knowingly* failing to take necessary means to our ends – and I submit that there is no anecdotal evidence that clearly supports this.

The very paradigms of instrumental irrationality fail to demonstrate violations of IN(4). However, this may reasonably be thought to evince rather that IN(4) is not the true instrumental norm. Rationalists offer various distinct formulations of the norm: might any of these meet both criteria? Perhaps I have not yet captured what is meant by ‘willing an end’ in this context: is there a viable interpretation without the same close link to action? In particular, my observations regarding the failure to meet the possible violation criterion have been based on constitutive claims about what it is to have *present intention*. But suppose what it is to will an end is free of any conditions on a agent’s *present* state of mind? Some rationalists identify instrumental irrationality with a form of *diachronic inconsistency*:

We continually make [resolutions to act in a certain way at some future time], and sometimes when the time comes for carrying them out, we do in fact act otherwise under the influence of passion or mere habit, *without consciously cancelling our previous resolve*. This inconsistency of will our practical reason condemns as irrational... (Sidgwick 1907: 37, my emphasis).

When I will an end, I must *ipso facto* will that even on another occasion, even when I am tempted not to, I will stay on the track of that end.... So when you will an end, the form of the act of your will is general: you will a kind of law for yourself, a law that applies not only now, but on other possible occasions. (Korsgaard 1996: 230-1)

If what we will is determined by our past resolutions rather than our present mental states, then the possible violation criterion is easily met. We cannot, of course, be rationally beholden to every end we have ever set for ourselves, and Sidgwick acknowledges our entitlement to change our minds: we are normatively bound only where “we do not consciously retract our adoption of the end” (1907: 38). Why, however, take past resolutions as normatively binding? Suppose that as a child I resolved never to force *my* future children to eat vegetables – surely I am not now irrational when I insist that my daughter eat her vegetables, never having consciously disowned that resolution.

Korsgaard explains that “the claim to generality, to universality, is essential to an act’s being an act of the will” (1996: 232) – to will at *t* is to commit yourself at *t* to a law that is to apply to you at all times, so that “when I follow a hypothetical imperative, one

part of me – say my will at one moment – governs another part of me – say at another moment” (1996: 230).<sup>27</sup> This may appear incompatible, however, with our entitlement to change our minds, and hence an unreasonable expectation. Korsgaard’s response is that while we may indeed change our minds, we may not do so *all* the time:

The reason I must follow hypothetical imperatives in general is that if I don’t follow them, if I *always* allow myself to be derailed by difficulty or dread or dullness, then I never really *will* an end. The *desire* to pursue the end and the desires that draw me away from it each hold sway in their turn, but *my will* is never active. (1996: 230)

Why does it matter that I never will an end? Korsgaard explains,

The distinction between my will and the operation of the desires and impulses in me does not exist, and that means that I, considered as an agent, do not exist. (1996: 230)

This claim is extended to normativity generally:

The function of the normative principles of the will...is to bring integrity and therefore unity – and therefore, really, existence – to the acting self. (1996: 229)

She explains this as follows. Strictly speaking, I am my reflective self. In order to see my actions as arising from my *self* (hence to see myself as an agent), rather than the mere products of various desires and impulses ‘in me,’ I must be able to see myself as something other than my particular, momentary states of mind. But since a ‘self’ is never an immediate object of experience, this is only possible if I can infer some persisting reflective character as the cause of my particular actions. As causation can only be discerned by regular, lawlike connections, I must act according to *some* universal, constant principles (my practical identity, as law), in order to exist at all as a temporally extended, reflective agent.

Supposing this to be a true story, there remain several avenues of intelligible challenge to this ‘rational requirement’. First, if nonexistence is a consequence merely of

---

<sup>27</sup> Korsgaard insists that instrumental irrationality does not require diachronic inconsistency: I can flout hypothetical imperatives while simultaneously willing the end, because I have two distinct *parts* that can separately determine my actions; “one that wills, and one that is capable of resisting my will” (1996: 231). The part that wills is my ‘*self*’, a reflective agent or ‘mind’ (1996: 228), the other consists of the impulses and desires “causally effective in or through my body”. I violate the instrumental norm when my mind is resolved upon an end, and my impulses and desires cause me to neglect required means. However, behaviour motivated by desire is *voluntary*, unlike reflexive behaviour, hence it cannot be external to or in conflict with the active will.

*never* (or only seldom) pursuing means to our ends, it would seem quite reasonable to flout the norm's counsel on many occasions (like the counsel to brush one's teeth). Second, why can't we reasonably opt for nonexistence of this kind? Indeed, repudiation of the myth of the 'self'<sup>28</sup> is a central teaching in many philosophical and religious systems, including existentialism and Buddhism. One could appeal to a desire for existence – but this would be to concede the desire-dependence of instrumental normativity! One could appeal to a further rational norm commanding us to maintain our existence – but this would again compromise the rationalist case, as the normative authority of such a principle would be less self-evident than that the instrumental norm.

Korsgaard's answer to this question is unclear, but she seems to offer the following: it is not that we have reason to obey the instrumental norm because we have reason to maintain our existence. Rather, obeying the instrumental norm (and other normative principles) is *constitutive* of agency:

the reason that [the reflecting self] has to unify itself into an agent who can persist through a series of relevant occasions is *not* that it has some reason to want or anticipate that it will persist into the *future*...The reason is rather that the view of itself as active *now* essentially involves a projection of itself into other possible occasions (1996: 229).

This *may* mean that I am simply not acting if I am not willing ends and pursuing means, hence it is impossible for me to act otherwise. If this is the account (and a number of inconsistencies suggest otherwise), it is an attempt to meet the reasonable expectation criterion by denying the possibility of violation. As Korsgaard acknowledges (1996: 231), constitutive rules that do not admit the possibility of deviation cannot be normative. Either violation of the instrumental norm is impossible, or else we are yet to receive an adequate explanation as to why we ought to obey it.

---

<sup>28</sup> Korsgaard's argument has a serious flaw; at most, absence of perceived lawlike connection between self and action would entail that I cannot *recognize* that I exist. This evinces that Korsgaard's transcendent 'self' is a nonexistent fiction; her normative principles command behaviour contrived to create the *appearance* of a unified cause. Existentialists condemn this as 'Bad Faith,' the fabrication of a fictional self standing behind our individual actions.

The diachronic inconsistency strategy, however, may significantly illuminate that my argument errs in assuming people to be time-slice agents. It must be observed that we frequently attribute to people as genuine constituents of their psychology *nonoccurrent* mental states. I can be said to believe some proposition *p* even at moments when *p* is ‘the farthest thing from my mind,’ and to intend some end *y* similarly, providing my dispositions meet certain conditions. It follows that I may nonoccurrently believe and intend, and hence that there may be no occurrent mental presence of something I nonetheless intend. One effect of violent passion is, indeed, to cause us to overlook nonoccurrent beliefs and intentions. While the failure to recall nonoccurrent beliefs can be chalked up once more to *theoretical* irrationality, the failure to recall nonoccurrent intentions may seem rather a practical failure. Clearly it is possible to fail to act in ways that I believe to be necessary means to my nonoccurrent and overlooked ends – and this may even appear a genuine violation of IN(4), avoidance of which is reasonably required.

To this problem I offer four responses. First: this too might turn out to be a case of epistemic rather than practical failure. Such lapses in intention (unlike lapses in desire) seem to be failures of memory, which suggests that intentions, if not reducible to beliefs, have at least a significant belief component, which is what fails in these cases. Second: it still seems appropriate to say that I did not intend the end *at that moment*. Attribution of nonoccurrent mental states might be a pragmatic turn of speech that imposes false but useful simplicity on complex psychological phenomena: in other words, I don’t *really* intend *y* at such moments, hence I don’t *really* violate IN(4). Third: any victory the rationalist may win here is hollow, as the paradigms of alleged instrumental irrationality are not manifestations of this particular form of psychological failure. Fourth: supposing both possible violation and reasonable expectation criteria are met, this is not yet *sufficient* to prove a normative command of reason, and rationalism

remains vulnerable (as I shall argue in Section IV) to an argument from an inference to the best explanation.

Perhaps, it might be thought, a rationalist can concede the analyticity of the means-end connection, and formulate a genuine instrumental norm of reason consistent with this. On Hampton's interpretation, the norm rather commands us to will certain *ends* (those we reflectively judge to be the *best*), and if we do, then necessarily we will the means, as we cannot will an end without being instrumentally rational: "to be instrumentally rational is to *will* the end that is in some way best or strongest in the circumstances, where willing means 'sustaining a commitment to that end through action'" (1998: 163).<sup>29</sup> Willing is a "norm-responsive psychological state," so that I am not willing an end unless I am committed to it on the basis of judging it to be best. Instrumentally irrational agents are thus those who "fail to will the end that they acknowledge is the best end in the circumstances (they want it, but they don't will it)" (1998: 165). Two problems are evident here. First, this distorts the intuitive notion of an instrumental norm – as a principle that gives normative guidance when we will some end – to the point of unrecognizability. Second, no true Humean will concede the existence of rationally mandated ends, or the self-evidence of this alleged normative principle (which Hampton herself describes as "metaphysically occult" [1998: 6] and as having 'mysterious objective authority' [1998: 166]) – hence her 'instrumental norm' offers no leverage from common sense against the thesis of the desire-dependence of normativity.

These rationalist attempts to find a satisfactory interpretation of the instrumental norm yield results that are alien and implausible. The notion of willing an end invoked in

---

<sup>29</sup> "I am trying...to make a conceptual point here about the nature of rational willing – that all and only instrumentally rational agents can be said to will ends, as opposed to merely 'wanting' them... Whereas we may merely 'want' an end, but refuse to take what we regard as hateful means to achieving it...if we are instrumentally rational agents then, via our reason, we will commit ourselves to that end – that is, 'will' it,

intuitive appeals to an instrumental norm is adequately captured in IN(4)'s terms of intention, and not in Sidgwick's talk of past commitments, Korsgaard's talk of lawlike conceptions of practical identity, or Hampton's talk of commitments to ends we judge to be best. Although I have not examined every possible rationalist account, I believe these failures strongly suggest that the rationalist project can only be fruitless.<sup>30</sup>

#### IV

There is no instrumental norm of reason. This conclusion will be too radical for many, who will point to familiar phenomena in everyday thought and discourse appearing to show that common sense supports rationalism here, and hence that skepticism must embrace an implausible error theory of practical rationality. Two ordinary practices in particular may appear to furnish proof of this everyday wisdom. First, there is ascription of 'irrationality.' We make judgments about what agents 'rationally' ought to do, and we frequently ascribe 'irrationality' to those who act otherwise, furnishing countless concrete examples of irrational action. Second, utterance of hypothetical imperatives is not the preserve of rationalist philosophers, but similarly a familiar part of everyday discourse and judgement. I will now gesture at how these phenomena can be accommodated by a desire-dependence account of normativity. My goal in this section is merely to present an alternative, anti-rationalist point of view.

I have argued that even the paradigms of instrumental irrationality fail to violate any plausible command of reason. Must we then decide against ordinary wisdom that irrationality is impossible? No. We may, like Hume, deny merely that irrationality *as defined by the rationalist* is possible. Rationalism deceives us, by speaking in the

---

insofar as we regard it as the best or the only end in the circumstances – and thus will not become derailed by the temptation for lesser goods" (Hampton 1998: 164).

<sup>30</sup> Joseph Raz writes, "The fact that the pursuit of the means is part of what makes one have the ends misled some into thinking that instrumental irrationality is impossible. If one does not pursue what one believes to be the necessary and available means, then one does not have the end... However, even in [simple] cases it is not failure to pursue the means alone, but also the absence of feelings of conflict... guilt and remorse that

language of common sense, into believing that it speaks *for* common sense. But it misinterprets these practices and utterances. Hume infamously writes,

What we commonly understand by *passion* is a violent and sensible emotion of mind... By *reason* we mean affections of the very same kind with the former; but such as operate more calmly, and cause no disorder in the temper... (T437)<sup>31</sup>

We need not endorse Hume's exact translations: perhaps 'reason' and 'rationality' are ordinarily used to refer to behaviour controlled by stable or persistent rather than calm desires, 'unreason' and 'irrationality' referring to that behaviour prompted by unstable or fleeting rather than violent passions (particularly where they are in conflict with 'reason') – or perhaps these terms are used in a variety of distinct ways. A football player who, overcome with rage, commits a personal foul depriving his team of the victory that is the object of his stable desires is acting 'irrationally' (hereafter I use scare-quotes to distinguish the ordinary, as opposed to the rationalists' use). Indeed, this Humean account can, unlike rationalism, accommodate the rationalists' own paradigms of irrational behaviour. The addicted smoker acts on the prompting of a temporary nicotine-induced urge, rather than on his stable, reflective desire that he not smoke. Everyday wisdom seems here to vindicate Hume.

Rationalists will challenge the claim that stability in reflection (for example) is *definitive* of 'rationality' in desire and action. That a desire survives reflection is rather a consequence of its being such as to be endorsed by dispassionate practical reason. Here I firmly dissent. But what, then, is to be made of the *normativity* of 'rational' criticism, if it simply amounts to a conflict of calm/stable against violent/fleeting passions? The answer is that 'rational' criticism is endorsement or censure *from the point of view* of the 'rational' self – i.e. our calm/stable concerns. Judgement of the 'rationality' of others'

---

establish I do not have the end.' (2005: 27). This argument seems without merit; surely we can feel conflict, guilt, and remorse over ends that we have not *willed*, but merely desire or believe we ought to have willed.  
<sup>31</sup> Also, 'When any of these passions are calm, and cause no disorder in the soul, they are very readily taken for the determinations of reason.... What we call strength of mind, implies the prevalence of the calm passions above the violent' (T417-8).

behaviour involves stepping into their shoes, and viewing their actions from the perspective of their '*rational*' desires together with their beliefs.<sup>32</sup> It guides us insofar as we are motivated by such desires, which is not always the case. The judgement that we 'ought' to be so guided is *itself* made from this point of view.

So far, I have assumed with rationalism that 'ought' is to be interpreted as expressing a command of reason (as logically consistent thought) – but we need a different interpretation. 'Ought' is rather teleological or *end-relational*.<sup>33</sup> Every 'ought' presupposes some particular end as framing a normative perspective, and recommends a course of action from that particular point of view as serving that end. Normative judgements do not express commands of dispassionate reason, but endorsement from particular motivated perspectives (hence the frequent conflict of moral and prudential 'oughts.')

There is, then, room for 'rational' criticism in the desire-dependence theory of normativity. But this does not extend to criticism of *instrumental* 'rationality' – as I cannot but intend what I believe to be the necessary means to my intended ends.

This may sound like rejection of any kind of instrumental normativity whatsoever. But what about the judgements we utter in the form of 'hypothetical imperatives' – are they not evidence of normativity in instrumental matters? Rationalism misinterprets the function of these judgements, which fall into two categories: propositions of *instrumental necessity*, and *practical directives*. First, there are statements of the form, 'In order to realize *y*, you must do *z*.' This 'must' is an alethic modal: it reports the instrumental necessity of doing *z* if it is to be the case that *y* obtains. It is a constitutive principle, that realizing *y* is impossible without doing *z*, which (as Korsgaard observes) *can* be used normatively in case an agent intends to realize *y*. In

---

<sup>32</sup> A similar, albeit non-Humean, view of rational criticism, as observation of what is wrong with an agent's behaviour from her own point of view, has since been persuasively advanced by Niko Kolodny (2005). Kolodny does not, however, question the possibility of instrumental irrationality.

<sup>33</sup> This suggestion can also be found in Foot 1972 and Mackie 1977. I argue for this view in my 'Oughts and Ends' (unpublished), and more generally for an end-relational view of normativity in my 2004 and 2006.

this circumstance, such propositions are used as practical directives, which I must now address.

It seems that not all apparent ‘hypothetical imperatives’ can be mere statements of alethic necessity, and that some involve deontic modals instead: particularly those incorporating ‘ought’ or ‘should’ rather than ‘must’.<sup>34</sup> These practical directives are *not* uttered in order to tell people that reason commands they take the means to their ends, but rather to inform them about the means to their ends. Should I say to someone, “If you want to see a modern defense of rationalism, you should read Korsgaard’s *The Sources of Normativity*,” I am aiming to inform her *not* of a rational requirement operating on facts we both recognize, but that reading this work is an excellent means to that end. There is never any need to prescribe to anybody that they take what they know to be the necessary means to their ends – although reminding or informing about means and prodding to induce lacking resolve or decision are often called for and easily mistaken for the former.<sup>35</sup>

This instrumental ‘ought’ still needs explanation. Haven’t I denied that failure to take the means to our ends is even possible? Since I endorse the possible violation criterion, in what sense am I entitled to this ‘ought’? All I have denied is that we can fail to will *what we believe* to be the necessary means to our ends. Ignorance, false belief, and confusion can lead us to neglect required means, and hence there is indeed room for an instrumental ‘ought.’ But this cannot be the *rationalists’* ‘ought.’ First, it is insensitive to the epistemic circumstances of the agent; if the necessary means to my end is to  $\phi$ ,

---

<sup>34</sup> Elsewhere (unpublished) I reject this appearance, and argue that the normative ‘must’ is nothing other than the alethic ‘must’ in an end-relational (in order that...) context, and further that the normative ‘ought’ is nothing other than the so-called *predictive* ‘ought’ in an end-relational context. But the argument of this paper doesn’t depend on these radical theses.

<sup>35</sup> This is now recognized by a number of philosophers (e.g. Scanlon 2003, Raz 2005), who deny that rational principles are *normative* in the strict sense of giving guidance. But these philosophers still accept the possibility of instrumental irrationality, viewing the principle as *evaluative* or as underlying criticism. Kolodny (2005: 554-6) defends the normative status of rational principles.

then, instrumentally speaking, I ought to  $\phi$  – even if unaware of this fact.<sup>36</sup> Second, it is not limited to what is directly within my power: it is actually true that I (instrumentally) ought to  $\phi$ , not simply that I ought to *try* or *intend* to  $\phi$ .

These features of instrumental normativity fit the model of the end-relational ‘ought’ – a thesis that allows reintroduction of further intuitive features we had to strip from everyday ‘hypothetical imperatives’ when seeking to formulate a satisfactory rational norm. Third, therefore, the antecedent of a practical directive need not specify the willing of an end; it can specify a desire, so that ‘If you want  $y$ , you ought to do  $z$ ’ legitimately recommends some course of action  $z$  from the perspective of a desire for  $y$ .<sup>37</sup> Fourth, the means need not be *necessary*. We can recommend a means as the *best*, from the perspective of some desires. We can say “Well, if you want  $y$ , then you *ought to do*  $z$ ...but you *may* also do  $u$ .” Rationalism cannot account for this ‘may’, since it interprets ‘ought’ as expressing a command of reason, so that not doing  $y$  is contrary to reason, hence impermissible.

In being insensitive to an agent’s epistemic and physical limitations, judgements of instrumental normativity are *not* judgements of ‘rational’ criticism. I cannot justly be condemned as ‘irrational’ for failing to do what I instrumentally ought. Instrumental ‘oughts’ and ‘rational oughts’ fall into separate categories of normative propositions. The desire-dependence thesis makes room for instrumental normativity and ‘rational’ criticism, but not for criticism of instrumental ‘rationality.’

Given this affirmation of instrumental normativity, can we after all formulate a legitimate instrumental norm – of *desire*? The rationalists’ response to this proposal is to

---

<sup>36</sup> Suppose you ask me how you should accomplish some end. My answer, if cooperative, will not be to recommend you pursue what I take *you* to believe to be the best means, but rather to recommend to you what *I* believe to be the best means. The imperative I utter is not constrained by your epistemic situation.

<sup>37</sup> Note, however, that the hypothetical form is misleading – hence my rejection of the label ‘hypothetical imperative’. The preceding clause does not actually function as the antecedent of a conditional, but rather indicates the perspective of the ‘should’. (However, ‘If you have pedophilic desires, then you should stay away from kindergartens,’ is (ironically) legitimately described as a hypothetical imperative).

challenge, “which desire?” Taking the means to my ends is instrumentally required of me *whatever* my desires, but desires are contingent. Rationalists conclude that the requirement must derive from an authority independent of all desire. But how satisfactory as a solution is appeal to the command of reason? Why would reason command such a thing? Rationalists often maintain that the justification of the instrumental norm is simply *brute*.<sup>38</sup> There is no further explanation why we should take the means to our ends; it is a fact about reason we must simply accept. But what harm could there be in disobeying a gratuitous command without justification or sanction?<sup>39</sup>

The explanation why I ought to pursue the means really isn’t at all mysterious. Failure to take the means signifies failure to attain the end – this is the sanction and reason. To what does it matter that we fail to attain our ends, regardless of whether they are evil or misguided? Objective reason? But why would it – and how could it – care? If we are seeking that patron that always cares whether or not we attain the object of our desire, whatever the desire, there is one obvious candidate. Failure to satisfy a desire always matters – analytically! – to that desire itself. It is from the perspective of the desire for *y* that we judge we *ought to do z* when we believe doing *z* to be the means to *y*. To the question, “which desire counsels us to pursue the means?” I answer, whichever desire would have us seek the end.

Every desire is instrumentally normative: whatever the object of my desire *y* might be, I ought<sub>*y*</sub> to take some means to satisfying it. We can even abstract a true general principle that looks like an instrumental norm; approximately,

$$\text{IN}(5): (\forall \text{ agents } x, \forall \text{ ends } y, \forall \text{ actions } z) ((x \text{ desires } y \ \& \ z \text{ is the means to } y) \rightarrow x \text{ ought}_y \text{ to do } z).$$

This is compatible with the rationalists’ claim that it is possible there is no desire that counsels taking the means in every circumstance, or, approximately,

---

<sup>38</sup> Hampton takes this stance, Korsgaard resists it. See also Beardman 2007.

Possibly,  $\sim\exists$  a desire  $u$  such that  $(\forall x, \forall y, \forall z) ((x \text{ desires } y \ \& \ z \text{ is the means to } y) \rightarrow x \text{ ought}_u \text{ to do } z)$ .

We can now see that it does not follow from this, as rationalists claim, that there is a desire-independent (unsubscripted) ‘ought.’ But might it nonetheless be true?

Rationalists argue for norms of reason over desire-dependent normativity on the grounds that the correctness of taking means to our ends is not internal to that particular motivated perspective. We criticize those who fail to take such means even where we do not share their desires. The desire-independence of instrumental criticism is alleged to prove the desire-independence of instrumental normativity. Instrumental and rational criticism is thereby seen as possible from a perspective devoid of all desires (the ‘view from nowhere.’) Observe, however, that when challenged to justify the claim that the subjects of criticism *ought to care*, rationalists insist that such agents are making a mistake *even within the context of their own desires and beliefs*. Rather, it is a ‘mistake’ simply *because* it is a mistake within the context of that perspective. Rationalism overlooks the possibility of imaginatively and conversationally borrowing another’s perspective of concern, and making judgements indexed to it (a phenomenon amenable to Hume’s doctrine of sympathy.) Judgements of instrumental normativity involve assessing people’s actions from the perspective of their own motivating desires, just as judgements of ‘rationality’ involve adopting the perspective of others’ ‘rational’ desires plus beliefs. Suppose while watching a crime documentary I declare, “The murderer should have used a gun!” I am not invoking a desire-independent rational norm, but making a normative judgement from the point of view of the murderer’s ends. The crucial rationalist error lies in the inference that because one always ‘ought’ to take the means to one’s ends,<sup>40</sup> therefore it must be commanded from outside all desires.

---

<sup>39</sup> Similar skepticism about the importance of rationality is pushed by Kolodny (2005).

<sup>40</sup> There is always a sense of ‘ought’ upon which this is true, but there might not be a sense on which it is always true.

Is there an instrumental norm of desire? IN(5) may be true, but it is not normatively authoritative. On the end-relational account offered here, 'ought' has only contingent authority. An 'ought<sub>e</sub>' has authority for an agent only insofar as the end *e* is integrated into the agent's desires and concerns. IN(5) is therefore a generalization from instrumental propositions, various of which are contingently authoritative for various agents, but does not itself possess any authority for those agents. Its subscript *y* is a variable ranging over all possible ends, only a few of which are important to any actual agent. I conclude that there is no authoritative instrumental norm at all.

## References

- Beardman, Stephanie (2007) 'The Special Status of Instrumental Reasons' *Philosophical Studies* 134: 255-287.
- Broome, John (2000) 'Normative Requirements.' Reprinted in Jonathan Dancy (ed.), *Normativity*. Oxford: Blackwell.
- Broome, John (2001) 'Normative Practical Reasoning.' *Proceedings of the Aristotelian Society*, Suppl. Vol. 75: 175-193.
- Dreier, James (1997) 'Humean Doubts about the Practical Justification of Morality', in Garrett Cullity & Berys Gaut (eds) *Ethics and Practical Reason*. Oxford: Clarendon Press.
- Finlay, Stephen (2004) 'The Conversational Practicality of Value Judgement' *Journal of Ethics* 8 (3): 205-223.
- Finlay, Stephen (2006) 'The Reasons that Matter' *Australasian Journal of Philosophy* 84 (1): 1-20.
- Finlay, Stephen (unpublished) 'Oughts and Ends.'
- Foot, Philippa (1972) 'Morality as a System of Hypothetical Imperatives' *Philosophical Review* 81: 305-316.
- Hampton, Jean E. (1995) 'Does Hume Have an Instrumental Conception of Practical Reason?' *Hume Studies* 21 (1): 57-74.
- Hampton, Jean E. (1998) *The Authority of Reason*. Cambridge: Cambridge University Press.
- Hill (Jr.), Thomas E. (1973) 'The Hypothetical Imperative' Reprinted in *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca & London: Cornell University Press, 1992.
- Hill (Jr.), Thomas E. (1989) 'Kant's Theory of Practical Reason' Reprinted in *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca & London: Cornell University Press, 1992.
- Hubin, Donald (1996) 'Hypothetical Motivation' *Nous* 30 (1): 31-54.
- Hume, David (1888) *A Treatise of Human Nature*. L. A. Selby-Bigge (ed.). Oxford: Clarendon Press, 1888.
- Kant, Immanuel (1911) *Grundlegung zur Metaphysik der Sitten*. Paul Menzer (ed.). Berlin: Koenigliche Preussische Akademie der Wissenschaften.
- Kant, Immanuel (1997) *Critique of Practical Reason*. Mary Gregory (transl., ed.). Cambridge: Cambridge University Press.
- Kolodny, Niko (2005) 'Why Be Rational?' *Mind* 114: 509-563.
- Korsgaard, Christine M. (1986) 'Skepticism about practical reason' Reprinted in *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.
- Korsgaard, Christine M. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Korsgaard, Christine M. (1997) 'The Normativity of Instrumental Reason' in Garrett Cullity & Berys Gaut (eds) *Ethics and Practical Reason*. Oxford: Clarendon Press.
- Mackie, J. L. (1977) *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin Books.
- Millgram, Elijah (1995) 'Was Hume a Humean?' *Hume Studies* 21 (1): 75-93.
- Nagel, Thomas (1970) *The Possibility of Altruism*. Oxford: Clarendon Press.
- Raz, Joseph (2005) 'The Myth of Instrumental Rationality' *Journal of Ethics & Social Philosophy* 1 (1): [http://www.jesp.org/PDF/6863\\_Raz-vol-1-no-1-rev.pdf](http://www.jesp.org/PDF/6863_Raz-vol-1-no-1-rev.pdf).

- Scanlon, T. M. (2003) 'Metaphysics and Morals' *Proceedings and Addresses of the American Philosophical Association* 77: 7-22.
- Schroeder, Mark (2004) 'The Scope of Instrumental Reason' *Philosophical Perspectives* 18: 337-364.
- Sidgwick, Henry (1907) *The Methods of Ethics*. Chicago: Chicago University Press.
- Setiya, Kieran (forthcoming) 'Cognitivism about Instrumental Reason' *Ethics*.
- Von Wright, G. H. (1978) 'On So-Called Practical Inference' Reprinted in Joseph Raz (ed.) *Practical Reasoning*. Oxford: Oxford University Press.
- Wallace, R. Jay (2003) 'Practical Reason' *Stanford Encyclopedia of Philosophy*: <http://plato.stanford.edu/entries/practical-reason/>