



## Subjective Accounts of Reasons for Action

David Sobel

*Ethics*, Volume 111, Issue 3 (Apr., 2001), 461-492.

Stable URL:

<http://links.jstor.org/sici?sici=0014-1704%28200104%29111%3A3%3C461%3ASAORFA%3E2.0.CO%3B2-%23>

---

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

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

*Ethics* is published by The University of Chicago Press. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ucpress.html>.

---

*Ethics*

©2001 The University of Chicago Press

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2002 JSTOR

# Subjective Accounts of Reasons for Action\*

*David Sobel*

Here is a familiar thought: consequentialism is an account of the moral rightness of acts (rules, etc.), it is not a decision procedure. That is, consequentialist ethical theories are best understood as accounts of what makes an act right or wrong rather than accounts of what should enter our heads when we decide what to do. They do not in the first instance offer a blueprint for how we ought to reflect about ethical predicaments in everyday life.

Frequently however, even still, consequentialism is attacked on the grounds that there would be bad consequences if people explicitly thought in terms of how to maximize the good when they deliberated about how to act. Such complaints are not wrong so much as misdirected. It may be that, as some have claimed, it makes no sense to so sharply separate truth-makers from decision procedures in practical matters.<sup>1</sup> Or it may be that consequentialism as an account of the truth-maker is sorely lacking. But criticizing consequentialism as a decision

\* I gratefully thank Rachel Cohon, David Copp, Justin D'Arms, Janice Dowell, Geoff Himes, Ted Hinchman, Dan Jacobson, Doug Lavin, Mark LeBar, Loren Lomasky, David Schmidtz, David Velleman, Mike Weber, and Matt Weiner for their help with this article. I am also grateful for helpful comments from anonymous associate editors of this journal. I want to express special thanks to Connie Rosati for her extraordinary and frequent help. I presented an earlier version of this article in 1998 at Bowling Green State University's Well-Being conference, in 1999 at Kenyon College and Bowling Green's WIMP reading group and in 2000 at Swarthmore College and the University of Cincinnati. I am grateful to these audiences for many useful suggestions.

1. James Griffin, "Incommensurability: What's the Problem?" in *Incommensurability, Incomparability, and Practical Reason*, ed. Ruth Chang (Cambridge, Mass.: Harvard University Press, 1997), p. 47, in speaking of indirect forms of utilitarianism writes that "although criterion and decision procedure can diverge, they are kept in the same general neighborhood by our capacities. Our decision procedures will, of course, be restricted by our capacities. But if a criterion becomes too remote from our capacities, it will cease serving as a criterion." Bernard Williams earlier had suggested that a radical split between truth-maker and decision procedure can create problems for utilitarianism. If one accepts such a split then it seems that it might happen that "utilitarianism has to vanish from making any distinctive mark in the world, being left only with the total assessment from the transcendental standpoint" (J. J. C. Smart and Bernard Williams, *Utilitarianism: For and against* [Cambridge:

*Ethics* 111 (April 2001): 461–492

© 2001 by The University of Chicago. All rights reserved. 0014-1704/2001/11103-0013 \$02.00

procedure simply fails to make contact with the view that most leading consequentialists have been busily advocating.<sup>2</sup>

In this article I will argue that the most prominent subjectivist accounts of reasons for action are, like consequentialism, best understood as accounts of the truth-maker in a certain domain. They are not in the first instance recommendations about the kind of reasoning that ought to be going on in people's heads. Further, as with the case of consequentialism, the failure to appreciate that such accounts are accounts of truth-makers and not decision procedures has misled some of our best philosophers, most notably Christine Korsgaard, into criticisms that similarly fail to make contact with the views they mean to address.

Thus my primary goal in this article is not to defend subjectivist accounts of reasons for action. Rather, my goal is to offer a persuasive interpretation of what such accounts are accounts of. This interpretation, if correct, reveals some criticisms of such views to misunderstand the views being criticized. Indeed, if my interpretation is correct, it seems that some proponents of the view also have not appreciated properly their own view. The failure to understand what subjectivist accounts of reasons for action are accounts of has impeded a fair assessment of the strengths and weaknesses of such views.

Many contemporary proponents of subjectivist accounts of well-being have followed the consequentialist's lead, advancing their views as accounts of the truth-maker of individual well-being rather than as methods of discovering it. As with consequentialism, such an approach provides the best interpretation of what subjectivist theories of well-being are theories of.

For reasons that are not clear, proponents of subjectivist accounts of reasons for action have not explicitly adopted this approach. Nonetheless, I will argue that just as consequentialism and subjectivist accounts of well-being are best interpreted as offering truth-makers rather than methods of discovery, so subjectivist accounts of reasons are best under-

---

Cambridge University Press, 1973], p. 135). This is apparently held to be bad news for the thesis of utilitarianism. As we will see later, if Williams's argument here was telling against utilitarianism (as I think it is not), it might be that Williams's own view of reasons would fall prey to the same sort of argument.

2. Few influential consequentialists have failed to notice that consequentialism is much better fitted for the role of truth-maker than decision procedure. See, among others, J. S. Mill, *Utilitarianism* (Indianapolis: Hackett, 1979), chap. 2; Henry Sidgwick, *The Methods of Ethics*, 7th ed. (Indianapolis: Hackett, 1981), pp. 405–6, 413, 489–90; G. E. Moore, *Principia Ethica* (Amherst, Mass.: Prometheus, 1988), pp. 162–64; R. E. Bales, "Act Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure," *American Philosophical Quarterly* 8 (1971); Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), pp. 24–29, 31–45, 98–100; R. M. Hare, *Moral Thinking* (Oxford: Oxford University Press, 1981), chaps. 2 and 3; Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," in *Consequentialism and Its Critics*, ed. Samuel Scheffler (Oxford: Oxford University Press, 1988), pp. 257–65.

stood as offering truth-makers for claims about what we have reason to do, rather than accounts of how to discover what we have reason to do.

To make out this case, I will at times draw parallels between subjectivist accounts of well-being and reasons for action. I do so because the most influential views of these sorts share a common crucial feature. They both require that we be factually informed, or at least not factually misinformed, before they allow that our proattitudes determine our well-being or reasons for action. This is true of Hume's and Bernard Williams's accounts of reasons for action and for a long line of subjectivist accounts of well-being. These subjective accounts share the important point that the factual information they require may be impossible for actual agents to get. Further, even when the required information is possible to get, it may yet be that the agent had no reasonable grounds to suspect it was worth the getting. Thus proponents of such accounts are not claiming that sensible deliberation would involve gathering all the information that the epistemically idealized agent would have.

This by itself shows that these accounts are not about how we ought to proceed in everyday life when we have to make a practical judgment.<sup>3</sup> They offer an account of what makes it the case that one has more reason to  $\phi$  rather than  $\psi$  or that it is better for you to  $\phi$  rather than  $\psi$ . But they are not claiming that a person need be irrational or imprudent for failing to learn such facts or for failing to act as she would had she known such facts. A person might do the best that could reasonably be asked of her by way of investigating such matters and still fail to learn the truth about her well-being or reasons for action. Such a person should not be thought irrational or imprudent for acting on such justified but mistaken beliefs. Thus, I will be claiming, the famous Humean instrumental account is not best understood as an account of rationality at all. It is an account of the truth-maker of claims about an agent's reasons for action.

This difference between an account of rationality and an account of reasons for action may sound trivial, but it is not. Subjectivist accounts of well-being or reasons for action are not relativized to our epistemic

3. It might seem that I am treating the fact that an account is epistemically relativized as criterial of its being a decision procedure. And this would be problematic since it seems that versions of consequentialism that morally require us to maximize expected (rather than actual) value could nonetheless sensibly claim to be accounts of the truth-maker of moral claims despite being epistemically relativized. But notice that my case hinges on claiming that the fact that an account is not epistemically relativized shows that it cannot sensibly be thought to be an account of the proper decision procedure. I can allow that some epistemic relativization is possible in some accounts of truth-makers and merely insist that the lack of epistemic relativization in a theory is incompatible with interpreting the theory to be an account of proper decision procedures for finitely rational creatures like ourselves. I must here resist the temptation to argue that the seeming plausibility of the version of consequentialism that recommends the maximization of expected utility is itself partly the result of the confusion between truth-makers and decision procedures. I thank an anonymous referee for urging me to be clearer about all this.

predicament. Rather they specify an ideal epistemic vantage point and claim that our subjective reactions from that vantage point determine our well-being or reasons for action. Accounts of rationality or prudence should be epistemically relativized. What counts as rational or prudent action depends on what information we have or could reasonably get. What options are best for me or that I have most reason to choose does not similarly depend upon my epistemic situation.<sup>4</sup> Whether or not I have a reason to  $\phi$  and whether or not  $\phi$ -ing will benefit me does not, except in special cases, change as my state of information changes. Because of these differences, claims about rationality are typically best fitted to be claims about appropriate decision procedures and claims about reasons are typically best fitted to be claims about truth-makers.

This article will have two sections. In the first, I distinguish an account of reasons for action from an account of rationality. I claim that prominent subjectivist accounts are best understood as accounts of reasons and not of rationality. In the second section, I will examine how accepting the claims of Section I would upset influential arguments against Humean instrumentalism offered by Korsgaard.

## I. SUBJECTIVISM: DECISION PROCEDURE OR TRUTH-MAKER

This section of the article will have five subsections. In subsection A, I distinguish accounts of rationality from accounts of reasons. In B, I consider the accounts of Hume and Bernard Williams and argue that they are best read as accounts of reasons and not accounts of rationality. In C, I briefly argue that the failure to attend adequately to this distinction between reasons and rationality can also be found in the work of prominent subjectivists such as Richard Brandt. In D, I consider the parallel development of subjectivist accounts of well-being. I try to show that speaking of subjectivism about well-being and reasons for action in one breath, as I have been doing, is justified. Finally, in subsection E, I consider what the subjectivist should have to say about the relationship between reasons and rationality.

### A. A Parallel with Theoretical Rationality

A person who is theoretically rational, I presume, adjusts her degree of confidence in a proposition to reasonably reflect the evidence for and against it. In some cases being theoretically rational also involves seeking out data for or against beliefs and perhaps even creating new data. The

4. Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, Mass.: Harvard University Press, 1990), p. 18, makes a comparable distinction. He writes that "one problem for any 'full awareness' account such as Brandt's is that rationality, in the ordinary sense, often consists not of using full information, but of making the best use of limited information. Acting in full awareness of all relevant facts suggests not rationality, but something more like 'advisability'. Whereas rationality is a matter of making use of the information one has, advice can draw on information the advisee lacks." See also pp. 89–92.

important point here is that a person could be theoretically rational to believe proposition P, and yet P might be false. Sometimes the evidence that even conscientious striving allows us to gather can be misleading. To be theoretically rational in one's belief formation, therefore, does not preclude fallibility in one's beliefs.

Further, we cannot simply discover who is theoretically rational by seeing who holds the greatest number or proportion of true beliefs. A theoretically irrational person might nonetheless get lucky in that many of her unjustified beliefs turn out to be true. There are two things an agent might accomplish: (1) processing information and gathering information in sensible ways and (2) believing the truth. We evaluate people as theoretically rational or not in the ways they satisfy the former requirement rather than how they fare with respect to the latter requirement. If the world cooperates then people who do well at the former will often do better at the latter. But the world failing to so cooperate does not make reasonable decision procedures unreasonable.

Rationality (or prudence) in practical matters seems quite similar to theoretical rationality in the above respects. A practically rational person is someone who processes the available information in reasonable ways in deciding what to do, not the person who happens, perhaps luckily, to act in the way that she had most reason to act.<sup>5</sup> There are two things an agent might accomplish: (1) processing information and gathering information in sensible ways and (2) acting in the way that she has most reason to act. We call people practically rational to the extent that they are successful in the former rather than the latter. The former accomplishment concerning invoking sensible decision procedures best corresponds to the way we want to use the term rationality, as we will see in the following example.

Imagine that Sally has carefully solicited advice from financial experts and decided to plan for retirement by investing in a diversified portfolio. Let me stipulate that Sally has admirably investigated how best to invest her money in the way most likely, given the information available at the time, to result in a comfortable retirement. Fred, on the other hand, has another retirement plan. He, like Sally, very much cares to have enough to live comfortably after retirement. But Fred decides to invest only a dollar in his retirement plan and buy a powerball lottery ticket with it. The lottery works like this. Each person, whether they buy the ticket or not, is assigned a number. Then a number is randomly drawn so that the winning number is determined prior to anyone investing in the lottery. If a person purchases her number for a dollar then, if her number has been drawn, she receives enough money at retirement to be comfortable. Fred buys his number, puts the ticket in a drawer, and

5. Of course we might also call someone practically irrational on the grounds that she does not act in accord with her sensible deliberation.

confidently awaits a comfortable retirement despite his knowing that the odds are 800,000,000 to 1 against his winning.

Suppose further that Fred's number has been drawn prior to his buying the ticket, although he has no way of knowing this. Suppose also that Sally's carefully thought out investment scheme will lose her money as a result of unforeseeable and unlikely changes in the world economy. What should we say about the rationality of Sally's and Fred's retirement schemes? We could say that Fred had a genuine reason to play the powerball (after all this will get him the money he so covets) and that since his action is in fact effective in realizing this important goal, his action is rational. Sally has a genuine reason to not invest her money. So she acted for a reason that was not a real reason and hence, we could say, is irrational. We could say all this, tying the notion of rationality to acting in ways that in fact promote one's goals.

But this is not the most natural way to use the term 'rationality.' Most of us would say that Fred was irrational in counting on a comfortable retirement yet got lucky and that Sally made rational decisions with her money but got unlucky. In speaking this way we tie rationality to the following of a reasonable decision procedure rather than acting in ways that in fact best further our goals. In this sense a person could be rational to engage in activities that do not turn out for the best, perhaps because bad luck sometimes keeps the best plan from working better than a bad plan.

If we want to be able to say that Sally invested rationally and Fred irrationally, then the Humean account of what agents have a reason to do is not an account of rationality. What is frequently called the Humean instrumental account of rationality is no account of rationality at all.

We must further distinguish between choosing the act that a rational decision procedure would recommend and actually invoking a rational decision procedure. When we merely act in conformity with the conclusions we would have reached had we deliberated rationally, let us say one makes a rational choice. Thus one might make a rational choice by sheer luck. When one makes a rational choice because one engaged in sensible deliberation, let us say one acted rationally. Thus only evaluating another as having acted rationally commits us to a positive evaluation of the actor's decision procedure.

Although we do not want to call Fred rational, we do want to allow that there is something to be said for his buying the ticket. Fred would certainly want to buy the ticket if he fully appreciated the truth of the situation. That he does not fully appreciate the truth does not mean that these attractive features disappear. This is clearest when a person requires inaccessible information to see the best causal path to achieve a given end. Assuming the end to be uncriticizable, surely we want to say that the agent has a kind of reason to take the path that in fact will best further her aim, even when the agent herself lacks good grounds for

suspecting that this action is the true means to her end. It is this kind of reason for action that is the topic of the theories of Hume and Williams. Or so I will argue in the following section.

### B. Hume and Williams

David Hume wrote that

a fruit, for instance, that is really disagreeable, appears to me at a distance, and thro' mistake I fancy it to be pleasant and delicious. Here is one error. I choose certain means of reaching this fruit, which are not proper to my end. Here is a second error; nor is there any third one that can ever possibly enter into our reasoning concerning actions.<sup>6</sup>

For Hume it would seem that our actual desires could mislead us into action that we have no reason to take. This is the case with the person who wants the fruit and takes effective means to it, yet finds it disagreeable. Hume wants to say that “a passion must be accompany'd with some false judgment, in order to its being unreasonable; and even then 'tis not the passion, properly speaking, which is unreasonable, but the judgement” (p. 416). According to Hume, passions are original existences, not attempts to accurately map or represent aspects of the world. Passions, on this view, cannot be true or false.

For Hume, “Reason is the discovery of truth or falsehood” (p. 458). Thus, passions cannot directly be contrary to reason. Rather, “passions can be contrary to reason only so far as they are accompany'd with some judgement or opinion” (p. 416). Thus, if there is, as Hume claims there is, the possibility of “errors” in “our reasonings concerning action” (p. 460), they will have to result from errors of fact or inference that affect our wants.<sup>7</sup>

6. David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge (Oxford: Oxford University Press, 1967), p. 460. This formulation is clearly better than the comparable one Hume offers at *Treatise*, p. 416, where he writes, “Tis only in two senses that any affection can be call'd unreasonable. First, When a passion, such as hope or fear, grief or joy, despair or security, is founded on the supposition of the existence of objects, which really do not exist. Secondly, When in exerting any passion in action, we chuse means insufficient for the desig'd end, and deceive ourselves in our judgment of causes and effects.” This formulation of the former error is inferior because it fails to allow room for cases in which the object that is the foundation of our attitude actually exists, but, if we brought it nearer or had first-hand familiarity with it, we would alter our attitude. This is exactly the possibility that the later formulation allows. Further page numbers in the text refer to Hume's *Treatise*.

7. Some modern interpreters of Hume have claimed that his view is that no action can be contrary to reason. Hence rather than really being the founder of Humean instrumentalist accounts, he was a full-fledged skeptic about the powers of reason to have anything at all to say about action. See Rachel Cohon, “Hume and Humeanism in Ethics,” *Pacific Philosophical Quarterly* (1988): 99–116; and Elijah Millgram, “Was Hume a Humean?” *Hume Studies* 21 (1995): 75–93. Korsgaard concurs with the picture of Hume as a radical skeptic about practical reason and, for that matter, theoretical reason, in “The Normativity



One can read Hume as claiming that an action nonculpably based on a false factual judgment is still contrary to reason. Evidence for reading Hume this way is that Hume claims that the errors that lead us to act contrary to reason might well be “unavoidable” (p. 460) and “draw no manner of guilt upon the person who is so unfortunate as to fall into them” (p. 459). He thus implies that there may well be no reliable method of learning the truth in the situation in which the agent made her mistake. Thus, on this way of reading Hume, he must think that we can do the best that we can in deliberation, given our epistemic circumstances, yet fail to discover our true reasons for action. This can happen when facts that would matter to us, if we knew them, are inaccessible.

We suppose, in wanting the fruit, that it will be pleasant to taste or that it will have a particular familiar taste. The belief is false. Thus we have a criticism of the agent’s action. It rested on a false belief and without that false belief there would be no desire for the fruit. Hume’s example makes clear that it is reason’s job to discover phenomenological facts as well as propositional facts.

If we appreciate what the fruit really looks like, tastes like, and so on, and still want it, and furthermore if we take steps that are effective in satisfying our wants, then our action cannot be contrary to reason. Acts, for Hume, are contrary to reason, albeit indirectly and in an “improper” way of speaking, when the act is motivated by a passion which would not exist (or, presumably, would be significantly altered) except for misinformation. On this reading of Hume it would be natural for him to say that one acts in accord with reason or one has a genuine reason to  $\phi$  when one’s motivation to  $\phi$  is not based on some such misinformation.

The above reading of Hume, it will surely be noted, develops a criticism of acts and passions that Hume himself found “figurative and improper.” He wrote that “these false judgments may be thought to affect the passions and actions, which are connected with them, and may be said to render them unreasonable in a figurative and improper way of speaking” (p. 459). This warning about improper ways of speaking, it seems to me, is meant to remind us that desires and actions that are misdirected by false judgments still lack representative value and so cannot themselves be true or false and therefore cannot literally, for Hume, be contrary to reason.<sup>8</sup>

Nonetheless, Hume clearly understood and felt the temptation to call acts or passions based on false beliefs contrary to reason. Suppose we want to understand the sort of criticism that Hume seems to level against

---

of Instrumental Reason,” in *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut (Oxford: Oxford University Press, 1997), p. 223, n. 23, and p. 229, n. 36.

8. It would have seemed open to Hume to claim that acts or passions that would only result from such false judgments are themselves defective on grounds other than that they are false.

acts or passions based on false beliefs—that is, the criticism that Hume himself finds “figurative and improper.” This criticism, I argue, is closer to the criticism of being something one has no reason to do than the criticism of being irrational.

Further, it is this “figurative and improper” criticism of actions that Williams means to be building on in constructing his Humean account of reasons for action.<sup>9</sup> Modern Humeans have tended to fix on and develop Hume’s “figurative and improper” criticism of acts and passions not because they think that passions or actions can be true or false but because they think that reason can do more than discover truths and falsehoods. Modern Humeans suppose that practical reason can also recommend an act on the grounds that it would serve one’s authentic ends. When I speak of Hume’s account of reasons for action, I am assuming this more expansive understanding of reason than Hume himself allowed.

Williams explicitly adopts and extends Hume’s account of reasons for action. Williams claims that one only has a practical reason to  $\phi$  if one would arrive at a proattitude towards  $\phi$  by “proper deliberation.” The boundaries of proper deliberation, according to Williams, are vague but not hopelessly so. Williams, like Hume, holds that it is not the case that just any of an agent’s concerns provide that agent with reasons for action “because of elements in S [the agent’s subjective motivational set] based on false beliefs.”<sup>10</sup> Thus, for Williams, proper deliberation involves being disabused of false beliefs and, one would suspect, being supplied with unknown truths.<sup>11</sup> Williams also allows further and more creative roles for proper deliberation as well.<sup>12</sup>

Williams does not explicitly require that the agent know all truths, but rather seems more concerned that the agent not have falsehoods affect her choice of actions. In a wide array of cases the latter requirement really encompasses the former. For in many cases a person might

9. It is worth noting that it must also be this “improper” understanding of Hume that Korsgaard is working with when she writes, in “The Normativity of Instrumental Reason,” p. 228, that “Hume . . . does explicitly allow that actions can be irrational in two *derivative* ways: we act ‘irrationally’ when our passions are provoked by non-existent objects, or when we act on the basis of false causal judgements.” I take it that Korsgaard’s use of the word “derivative” and her putting in scare quotes “irrationally” signals that she is trying to understand, as I am, what Hume’s “improper” criticism of acts and passions amounts to.

10. Bernard Williams, “Internal and External Reasons,” in his *Moral Luck* (Oxford: Oxford University Press, 1981), p. 102. Michael Smith’s notion of proper deliberation in *The Moral Problem* (Oxford: Blackwell, 1994) is quite similar and explicitly builds on Williams’s account. However, Smith claims that desiring to get X after proper deliberation only provides one with a reason to get X if all rational agents would converge in their desires. Smith also argues that we have good reasons to expect such a convergence. I argue against both of Smith’s claims in “Do the Desires of Rational Agents Converge?” *Analysis* 59 (July 1999): 137–47.

11. Williams, “Internal and External Reasons,” p. 104.

12. See *ibid.*

choose to  $\phi$ , but she would not have done so had she known fact Y. In many such cases the agent must have at least tacitly assumed that Y was false, otherwise she would not have chosen to  $\phi$ . Of course the agent might only have assumed that Y was unlikely, which could be true. However, it is hard to believe that Williams means to make much of this latter sort of possibility. It would be odd, to say the least, to require that the agent know that Y will occur when the agent believes that it will not, but to not require the agent to know that Y will occur when the agent only believes that Y is unlikely.

Williams offers the following example: "The agent believes that this stuff is gin, when it is in fact petrol. He wants a gin and tonic. Has he reason, or a reason, to mix this stuff with tonic and drink it?" Williams acknowledges that we speak both as if the agent did and did not have a reason to drink. Williams tells us that in the explanatory sense, which is concerned with understanding why the agent acted as he did, the agent surely did have an intelligible reason to do what he did. But in the justificatory sense, which is his concern here, Williams suggests the agent did not have a reason to mix the stuff with tonic and drink it.<sup>13</sup> I think Williams's understanding of this case is somewhat faulty. He wants to allow that the agent that drinks is acting rationally relative to his false beliefs. Thus there is not only an explanation of his drinking but a kind of justification as well. I think Williams should say that the agent is justified in drinking in that it is rational for him to do so but not justified in that there is no genuine reason for him to do so.

An agent can fail to act for a genuine reason in this sense even though she did not engage in sloppy or stupid methods of deliberation and information gathering. Hume's fruit-fancier and Williams's petrol-drinker acted contrary to reason. But for all we know these characters might have engaged in perfectly reasonable methods of deliberation.

Hume and Williams offer, in the fruit and petrol examples, cases in which people take actions that they do not have genuine reason to take. The point of these examples is to provide a clear case of an agent acting contrary to reason in the relevant sense and show why such actions count as contrary to reason from the point of view of their theory. Yet Hume and Williams give us no grounds to suspect that the agent's decision procedures in these cases were improper; no reason to suspect that they acted unwisely given their situation.<sup>14</sup> Why is all such information about the gullibility or carelessness in the way these individuals formed their

13. *Ibid.*, pp. 102–3.

14. Williams, in rehashing the petrol example, remarks that "we are allowed to change—that is, improve or correct—his beliefs of fact and his reasoning in saying what it is he has reason to do." Our license for doing so does not stem from the agent's poor use of available information but rather from the thought that "what he wants is a drink of gin and tonic." See "Internal Reasons and the Obscurity of Blame," in his *Making Sense of Humanity* (Cambridge: Cambridge University Press, 1995), p. 36.

beliefs left completely unspecified? Because, I submit, such information is irrelevant to the question that Hume and Williams are interested in, which is whether or not these people really had a reason to do what they did. The issue of the reasonableness of the decision procedure employed in these cases is not under examination and irrelevant to the question at hand.

Hume simply says that the choosing of means which are not efficient in producing the wanted end is an error that can lead us to act contrary to reason. Williams tells us that elements in our subjective motivational set which are there thanks to false beliefs do not give rise to real reasons. Neither Hume nor Williams finds any reason to investigate into the grounds of these false beliefs. Some false beliefs are justified. Others are not. But this distinction is treated by both philosophers as irrelevant.

Assume for a moment that I am wrong and Hume and Williams are offering a theory of rationality. Why might Hume and Williams think that the fruit-fancier and the petrol-drinker are acting irrationally? Obviously the only sensible answer would have to involve the unjustified nature of the beliefs that the agents had. But then our attention in the above examples should crucially be directed to the distinction between justified and unjustified beliefs. Acts based on justified belief, even if the belief is false, are not thereby irrational. Hume and Williams, however, far from focusing on the unjustifiability of the agents' beliefs, give us no information about the justifiability of the agents' assumptions. Can we imagine that the only thing in these cases which is relevant to the rationality of the agent's action, namely, the justifiability of the agent's beliefs, was nonetheless forgotten when these philosophers offered canonical statements of their accounts of rationality?

Williams allows that the agent that drank the petrol and tonic would be "relative to his false belief, acting rationally."<sup>15</sup> Thus Williams thinks there is a kind of commendation of the agent that drinks the petrol. But it is also clear that Williams thinks that there is something to be said against the action. Williams's attention is on the attempt to understand the sense in which the agent in the petrol case has no reason to drink the petrol. The fact that we cannot criticize the agent's deliberation as irrational drops from sight in Williams's discussion. All the attention is on understanding what counts as ideal deliberation (which obviously can involve knowing facts that we were excusably unaware of) and the connection between one's motivations after ideal deliberation and one's reasons for action.

The deliberation that Williams claims can close the gap between our current motivations and our genuine reasons is deliberation that, in many cases, we are unable to carry out in the actual world. Often, for example, the relevant facts that one would have in ideal deliberation

15. Williams, "Internal and External Reasons," p. 103.

have not yet been discovered. Additionally, the deliberation that Williams thinks can close the gap will in many cases be deliberation that it would be impractical for actual people to pursue. To think otherwise is to require rational agents never to trust that previously reliable sources of gin contain gin but instead to thoroughly test each beverage for authenticity prior to drinking.

Whatever the merits of their proposals, Hume and Williams are hoping to capture the sense of having a reason to  $\phi$  in which one might retrospectively say of oneself, "I had a reason all along to  $\phi$  and didn't realize it or have any reason to suspect it until now." We do speak as if we could have had a reason to  $\phi$  all along even though we did not have any information that would have made  $\phi$  a sensible choice at the time. Their question is not "is it reasonable for A to believe that she has a reason to  $\phi$ " but rather "what makes it the case that A has a reason to  $\phi$ ."

There is room for subjectivists to disagree about the vantage point from which one's proattitudes determine one's reasons. Thus it is not necessary that a subjectivist accept a "full information" account of reasons or well-being in which the deliberating agent knows all of the facts. However, I do suppose throughout that the most plausible subjectivist account would have to require that the agent deliberates in the light of information to which she may lack epistemic access.

I do not claim that Hume and Williams always clearly had in mind the interpretation I have offered of their writings or that there are no parts of their thinking that are in tension with my interpretation.<sup>16</sup> Rather, I want to claim that this interpretation is better able to capture and express the bulk of their concerns than an interpretation that took them to be offering accounts of decision procedures. I claim that their accounts are best understood as accounts of an agent's reasons for action and not accounts of practical rationality.

16. Williams, in "Internal and External Reasons," p. 103, writes that "the internal reasons conception is concerned with the agent's rationality." Additionally, Williams confusingly writes that "A may be ignorant of some fact such that if he did know it he would, in virtue of some element in S, be disposed to  $\phi$ : we can say that he has a reason to  $\phi$ , though he does not know it. For it to be the case that he actually has such a reason, however, it seems that the relevance of the unknown fact to his actions has to be fairly close and immediate; otherwise one merely says that A would have a reason to  $\phi$  if he knew the fact" (p. 103). I think the last claim here shows that Williams is sometimes incoherently trying to straddle the line between an account of reasons for action and an account of rationality. But even here note that Williams does not require that the unknown fact be one that the agent was in any sense culpable for not knowing. Hume writes, "A person may also take false measures for the attaining his end, and may retard, by his foolish conduct, instead of forwarding the execution of any project" (p. 459). The word 'foolish' here is troubling for my interpretation. It is also very awkward in Hume's context. In this and surrounding passages, Hume stresses merely the truth or falsity of the judgment, not its wisdom or stupidity. It is not at all clear why Hume thinks he has earned the right to call such "false measures" foolish rather than merely mistaken. Not all mistakes are foolish, and he tells us nothing about this false measure that helps us see why it counts as foolish.

Williams, it must be conceded, is arguing in the first instance for internalism rather than subjectivism. “The internalist view of reasons for action is that this formulation provides at least a necessary condition of its being true that A has a reason to  $\phi$ : A has a reason to  $\phi$  only if he could reach the conclusion to  $\phi$  by a sound deliberative route from the motivations he already has.”<sup>17</sup> This formulation leads to two ambiguities I need to briefly address.

First, internalism of this sort is compatible with either a tracking or a truth-making interpretation.<sup>18</sup> The former position claims that our informedly concluding to  $\phi$  is necessary to our having a reason to  $\phi$  but not what makes it the case that we have a reason to  $\phi$ . Groundhogs returning to their holes on Groundhog’s Day, upon being scared by their shadow as I hear it, is said to mean that there will be six more weeks of winter. But it is reasonably clear that the groundhogs’ behavior is not thought to make the winter linger. We cannot blame the cold on them for this would be to blame the messenger. Rather, the groundhogs’ behavior is claimed to be a reliable guide to the weather.

Tracking internalism holds that one’s informed proattitude toward  $\phi$ -ing is similarly just a reliable guide to one’s reasons, not what makes it the case that one has a reason to  $\phi$ .<sup>19</sup> It is thus compatible with objectivism rather than subjectivism about reasons for action. Objectivism, in

17. In “Internal Reasons and the Obscurity of Blame,” p. 35, Williams announces that he thinks this constitutes a sufficient condition as well. I argue against the stability of the necessary but not sufficient view in the context of well-being, in “On the Subjectivity of Welfare,” *Ethics* 107 (1997): 501–8. David Copp, in his *Morality Society, and Normativity* (Oxford: Oxford University Press, 1995), chap. 9, offers an interesting position in which one’s subjectively valuing 0 is sufficient but not necessary to the existence of reasons to achieve 0. Copp claims that one’s nonsubjectively determined needs provide an independent source of reasons that can conflict with one’s values.

18. Stephen Darwall’s formulation of “existence internalism” in *Impartial Reason* (Ithaca, N.Y.: Cornell University Press, 1983), p. 55, and his “metaphysical internalism” in “Reasons, Motives, and the Demands of Morality: An Introduction,” in *Moral Discourse and Practice*, ed. Stephen Darwall, Allan Gibbard, and Peter Railton (Oxford: Oxford University Press, 1997), pp. 308–9, are both, like Williams’s view, formulated in terms of necessary conditions for being a reason. Thus these versions of internalism are also subject to the importantly different interpretations mentioned in the text. Darwall briefly notes this ambiguity in the latter discussion.

19. Michael Smith’s account of reasons for action in *The Moral Problem* is best understood as a version of tracking internalism. He thinks that the desires of all ideally rational agents converging on certain things is necessary and sufficient for our having reasons, and, in particular, reason to do what we would so converge on. But, according to Smith, the best explanation for such a convergence, if it occurred, would be that there are “extremely unobvious a priori moral truths” (p. 187). On his view, it is these truths that make it the case that we have reasons to do certain things and our ideally informed deliberations simply gets our motivations to track these truths. The view Peter Railton offered in 1986 also looks to be tracking internalism rather than truth-making internalism. See his “Facts and Values,” *Philosophical Topics* 14 (1986): 5–29, p. 25, and “Moral Realism,” *Philosophical Review* 96 (1986): 163–207, pp. 175–76, n.17.

this context, is a thesis about what makes it the case that one has a genuine reason to  $\phi$ . If an account claims that the answer to this question is not to be found in the agent's contingent proattitudes, it counts as a version of objectivism. On the other hand, truth-making internalism embraces the subjectivist's claim that what makes it the case that one has a reason to  $\phi$  is that one has the relevant proattitude toward  $\phi$ -ing. Although Williams's defense of internalism is compatible with either the subjectivist or objectivist interpretation, he is clearly much more inclined to embrace the subjectivist interpretation.

Second, Williams writes, "A has a reason to  $\phi$  only if he could reach the conclusion to  $\phi$  by a sound deliberative route from the motivations he already has."<sup>20</sup> Suppose one started out without a certain bit of true information and without a proattitude toward  $\phi$ -ing. Suppose further one gained that bit of information and as a result one had a proattitude toward  $\phi$ -ing. Is this sufficient to say that the agent reached the conclusion to  $\phi$  by a sound deliberative route? I think clearly not and that Williams is committed to thinking so as well. Williams is interested in claims about one's reasons that are not epistemically relativized. This is why he announces that one has no reason in his sense to drink the gin and petrol without adding an epistemic qualifier.

Rather we should read Williams to mean, by the requirement that one could reach the conclusion to  $\phi$  via a sound deliberative route, that further, more ideal deliberation would not upset one's conclusion to  $\phi$ . Only in such a case does one's conclusion to  $\phi$  count as being such that it could have been reached via a sound deliberative route. Cases in which adding some true, but misleading, beliefs would create a motivation to  $\phi$  that would be undone by further acquisition of truths would not count as being such as could have been reached via a sound deliberative route.

Finally, there is one more problem with Williams's account I must briefly mention. It is clear that the idealization process of having our mistaken beliefs undone can turn an agent into someone whose well-being and reasons for action differ from those of our actual misguided selves. Presumably the fact that I have gained a bit of information will typically make it the case that it is no longer good for me to have that bit of information brought to my attention and that I have lost my reason to seek out that bit of information. But the fact that my factually informed self has bits of information does not take away my reason to learn that fact, despite its being the case that presumably my idealized self does lose any motivation to again be presented with or research that fact. Alternatively we could focus on cases in which the idealized agent's broad experience has caused her to develop a delicacy of taste that the actual agent lacks.

20. Williams, "Internal Reasons and the Obscurity of Blame," p. 35.

If we are to look to the idealized version of ourselves to determine our well-being or reasons for action we must take care lest aspects of our well-being or reasons that are due to our being nonidealized agents get lost. We are not interested in the well-being or reasons of the fully informed version of the agent. Rather, it is no doubt best to think of the idealized agent as an advisor.<sup>21</sup> This keeps front and center the thought that it is what the informed agent has to say about desirable ways for the uninformed agent to live that we are interested in.<sup>22</sup>

This points to a problem with Williams's account of reasons for action. Williams seems to suggest that it is the desires that the informed agent would have for herself, rather than the advice that the informed agent would have for us mere mortals, that determines the agent's reasons. And this is bound to lead to a mistaken view of the reasons for action of nonidealized agents such as ourselves. In subsection D we will see that in the well-being literature a plausible method of allaying this problem has been developed.

### C. Brandt

In an attempt further to illustrate that there has been confusion in the literature between reasons and rationality, even amongst chief advocates of subjectivist accounts, I will briefly point to the case of Richard Brandt. He does not distinguish adequately an account of reasons from an account of rationality and this creates problems for his view. However, my chief example of a case in which this confusion has created real trouble will be the case of Korsgaard that will concern us in Section II.

Brandt writes, "I shall pre-empt the term 'rational' to refer to actions, desires, or moral systems which survive maximal criticism and correction by facts and logic," and later, "this whole process of confronting desires with relevant information, by repeatedly representing it, in an ideally vivid way, and at the appropriate time, I call *cognitive psychotherapy*."<sup>23</sup> Even with this minimal description of cognitive psychotherapy it is plain that one could be blameless from the standpoint of rationality and yet still fail to act as one would want to act after it. It is not irrational to fail to read all the latest health journals. Thus it must not be irrational to act in light of the best information one can arrive at with-

21. Railton seems to have initiated this way of thinking of the ideal deliberator. See his "Facts and Values," p. 16.

22. Although such a picture would produce a more adequate account of reasons for action, it would falsify Williams's version of internalism and his explanation condition. Or so I argue in my "Explanation, Internalism, and Reasons for Action," *Social Philosophy and Policy*, vol. 18 (Summer 2001). There I argue that the best subjectivist account of reasons for action must reject Williams's understanding of internalism as well as his claim that "if it is true A has a reason to  $\phi$ , then it must be possible that he should  $\phi$  for that reason" (Williams, "Internalism and the Obscurity of Blame," p. 39).

23. Richard Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon, 1979), pp. 10 and 113.



out reading all the latest health journals. Nonetheless, one's failure to read all the journals might result in one making a choice that one would not have made had one read all the journals. Cognitive psychotherapy is insensitive to the costs of getting information and deliberation. In this sense it seems best suited to be an account of the truth-maker rather than a decision procedure. Thus it would seem that the topic of Brandt's account was an account of reasons rather than rationality.

But then, surprisingly, Brandt limits the information that cognitive psychotherapy would give us to "available information." The information that counts as available for Brandt is "the propositions accepted by the science of the agent's day, plus factual propositions justified by publicly accessible evidence (including testimony of others about themselves) and the principles of logic."<sup>24</sup> Brandt explicitly notes that available information in this sense "need not be true. It could also be information it would be intolerably expensive to get, so that trying to get it might itself be irrational."<sup>25</sup>

Now this is confusing. If Brandt's targeted notion was our notion of rationality, then cognitive psychotherapy should not involve knowledge that it makes no sense to try to collect. But if his targeted notion is our notion of a reason, then he should not limit the information to the possibly false beliefs of the best science of the day. The fact that the science of the day has not discovered that smoking causes cancer does not affect the fact that there is a reason for me not to smoke based in the fact that smoking has this effect. Thus Brandt's cognitive psychotherapy does not capture our notion of a reason or rationality. Sometimes he was tempted to construct his theory as if it were an account of reasons (as when he requires that we deliberate in light of information that it might be prohibitively expensive to get). In other cases he constructed his account as if it were an account of rationality (as when he relativizes the information used in cognitive psychotherapy to what is known in one's time).<sup>26</sup> The best explanation for the awkward combination is that Brandt did not keep clear the distinction between rationality and having a reason.

#### *D. Subjective Accounts of Well-Being*

In the well-being literature one finds a subjectivist account that is remarkably similar to the subjectivist accounts of reasons for action considered above. Call it the full information account of well-being. Mill's

24. *Ibid.*, p. 13.

25. *Ibid.*

26. In a conversation late in his life, Brandt told me that he regarded Railton's discussion of such issues as an improvement on his own because Railton's version of cognitive psychotherapy required all information rather than merely available information. I think this suggests that we do best to understand Brandt to have really been trying to capture the notion of a reason rather than rationality.

competent judges test offered an early model, Sidgwick offered perhaps the first explicit formulation of the account, and Brandt, Hare, Rawls, Gauthier, Griffin, Darwall, Lewis, Railton, and Harsanyi have developed and/or endorsed the view.<sup>27</sup> Roughly the picture is this: getting what one wants after one has full knowledge of the options available is what makes one's life go best.

Sidgwick's early version went like this:

A man's future good on the whole is what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were accurately foreseen and adequately realized in imagination at the present point in time.<sup>28</sup>

But Sidgwick's version above quickly runs into difficulties. Consider that our fully informed and rational self would never want more information for herself but we are firmly convinced that it can sometimes be intrinsically in our interests to gain information. Further our better selves no doubt have a refined palate and may well highly value expensive complex wines that taste just like the cheaper stuff to us. It is implausible that one wine is much better for me than another when I cannot tell the difference (assuming that it is only the taste of the expensive wine which causes our idealized self to prefer it over the cheaper stuff).<sup>29</sup> The idealization process turns us into such different creatures that it would be surprising if the well-being of the two of us, my informed self and my ordinary self, consisted in the same things.

In response to problems such as these, Peter Railton has revised the account proposing that

an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation

27. Mill, chap. 2; Sidgwick, pp. 111–12; Brandt, pp. 10, 113, 329; Hare, pp. 101–5 and 214–16. See also Douglas Senor and N. Fotion, eds., *Hare and Critics: Essays on Moral Thinking* (Oxford: Clarendon, 1990), pp. 217–18; James Griffin, in *Well-Being* (Oxford: Oxford University Press, 1996), pp. 11–17; John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), pp. 407–24; Gauthier, *Morals by Agreement* (Oxford: Clarendon, 1986), chap. 2; Stephen Darwall, *Impartial Reason* (Ithaca, N.Y.: Cornell University Press, 1983), pt. II; Railton, "Facts and Values"; David Lewis, "Dispositional Theories of Value," *Proceedings of the Aristotelian Society*, suppl. ser., 63 (1989): 113–37; John Harsanyi, "Morality and the Theory of Rational Behavior," in *Utilitarianism and Beyond*, ed. Sen and Williams (Cambridge: Cambridge University Press, 1973), p. 55. Several important caveats apply to some of the above authors' commitments to subjectivism and some would decline the label.

28. Sidgwick, pp. 111–12. Robert Shaver, "Sidgwick's False Friends," *Ethics* (1997): 314–20, helpfully reminds us that Sidgwick ultimately qualified this view in a way that limits its claim to being a subjective account. I dispute Shaver's further claim that Sidgwick had good reason to so qualify his view in my "Reply to Shaver," in the e-journal BEARS, <http://www.brown.edu/Departments/Philosophy/bears/homepage.html> (1997).

29. I take this example from Griffin's *Well-Being*, p. 11.

from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality.<sup>30</sup>

The adoption of a “wanting to want” framework neatly eschews the implausible identification of interests between our informed and our ordinary self while retaining the insight that “the advice of someone who has this fuller information, and also has the deepest sort of identification with one’s fate, is bound to have some commending force.”<sup>31</sup> Thus the sort of improvement Railton makes on Sidgwick’s suggestion is just the sort of fix for the problem with Williams’s account of reasons for action discussed above.<sup>32</sup>

With an agent’s well-being it is clear that the best forms of deliberation available to us when making everyday decisions do not assure us of understanding which option would be best for us. Thus, even in cases in which an agent has done the best she could, there may still be important information relevant to the location of her well-being that eludes her. Perhaps the science of one’s day has not yet offered up reasons to think that decaffeinated coffee causes cancer. Nonetheless, it may be that decaffeinated coffee does cause cancer. The fact that one had no reasonable way of determining that decaffeinated coffee was bad for you does not keep it from being bad for you. What you don’t know can hurt you.

Thus, because it is clear that subjective accounts of well-being mean to be accounts of what actually makes one’s life go better rather than accounts of the best available methods of investigating the question of what makes one’s life go better, it is clear that subjective accounts of

30. Railton offers this account in “Facts and Values,” p. 16. Notice that Railton’s compelling claim that it would be “an intolerably alienated conception of someone’s good to imagine that it might fail in any way to engage him” (p. 9) is also compatible with a merely tracking claim. In his more recent work, Railton claims that the subjective reactions from the approved vantage point are indicators of the presence of a fit between an individual and an end. See his “Aesthetic Value, Moral Value, and the Ambitions of Naturalism,” in *Aesthetics and Ethics: Essays at the Intersection*, ed. Jarrold Levinson (Cambridge: Cambridge University Press, 1998).

31. Railton, “Facts and Values,” p. 14. But consider that our idealized self could want our ordinary self to want  $x$  because the idealized agent knows that our ordinary self’s doing so will be instrumentally effective in bringing about, albeit unintentionally,  $y$ , which is what the idealized agent finds to be best for our ordinary self. If we say that what is good for our ordinary self is what our idealized self wants our ordinary self to want, we seem to misdescribe these cases of indirection. Perhaps it would be better to focus on the kind of life the idealized agent wants the ordinary self to have.

32. Michael Smith, *The Moral Problem*, chap. 5, adopts the spirit of Railton’s move. Connie Rosati, “Persons, Perspectives, and Full Information Accounts of the Good,” *Ethics* 105 (1995): 296–325, extensively considers such views, which she helpfully labels “Ideal Advisor” accounts.

well-being are accounts that purport to determine the right answer about what benefits us rather than accounts of the proper decision procedure.<sup>33</sup>

In contrast to accounts of reasons for action, accounts of well-being are easily seen to be accounts of truth-makers rather than decision procedures because well-being is clearly not an epistemically relativized notion. Having a normative reason, on the other hand, seems to have both an epistemically relativized sense and a nonepistemically relativized sense. Thus confusions can and do arise. But when we see that Hume and Williams are best seen as pointing toward the latter sense we see that here too the account is of the truth-maker and not of a decision procedure.

#### *E. Trying to Put Reasons and Rationality Back Together*

No doubt it is sometimes the case that the account of the truth-maker in a domain can also plausibly serve as an account of the appropriate decision procedure for discovering the truth in that domain.<sup>34</sup> This might happen when the account of the truth-maker does not require extra human capacities. But the accounts that subjectivists offer do not suggest ideals of deliberation that it makes sense to try to approximate or that we might call someone prudent or rational to the extent that they approximated. It would make no sense to gather randomly as many facts as possible, no matter how esoteric, before making practical decisions. Actual agents such as ourselves often must make decisions quickly, without crucial bits of information, and in the heat of the moment. Further we

33. A recent criticism of such accounts is that there are serious conceptual difficulties in attempting to specify the fully informed vantage point. See my "Full Information Accounts of Well-Being," *Ethics* 104 (1994): 784–810; and Rosati, "Persons, Perspectives, and Full Information Accounts of the Good." Such complaints retain whatever force they have against full information accounts of well-being when deployed against full information accounts of reasons for action.

34. Kant claimed that "in studying the moral knowledge of ordinary human reason we have now arrived at its first principle. This principle it admittedly does not conceive thus abstractly in its universal form; but it does always have it actually before its eyes, and does use it as a norm of judgement" (*Groundwork of the Metaphysics of Morals*, trans. H. J. Paton [New York: Harper & Row, 1964], p. 71). This backs away from claiming that the categorical imperative is the decision procedure of "ordinary human reason." It is unclear what ordinary people do that counts as close enough to having the categorical imperative "before its eyes." I take it to be a problem if only a handful of academics have a shot at invoking an approved decision procedure. Yet the relationship between the categorical imperative and an approved decision procedure available to nonacademics seems obscure. Some have suggested to me that the common thought of "What if everyone did that?" might count as invoking a recognizably Kantian thought as part of one's decision procedure. Presumably the Kantian needs a merely instrumental understanding of the relationship between the truth-maker and the decision procedure.

must distinguish between information that is relevant or likely to be worth the getting and information that is not.

A good theory of practical rationality must come to terms with the scarcity of time, limited resources, unreliable information, and our tendency to be tempted to rig deliberation in favor of the nearer benefit. Taking such things into account will necessitate that the proper decision procedure involves heuristics and rules of thumb.<sup>35</sup> The subjectivist's account simply abstracts away from such problems. These added complexities in everyday deliberation make the subjectivist's account a silly model to try to approximate when making practical decisions.

What the Humean should say about how we should deal with imperfect information, time pressure, and our predictable human weakness in a rational way is an interesting question. In fact it is as interesting as the question of what actual policies the consequentialist ought to recommend after she convinces herself that morality requires us to maximize the good. In both cases, the fact that much of the work of figuring out what we must do if we are to act for our true reasons or to act morally remains to be done even after we accept such theories does not, by itself, constitute an objection to such theories. Indeed I would say that it is no criticism of an account of a truth-maker qua truth-maker that it offers no useful guidance for helping us to actually discover the truth of the matter.

These considerations so far highlight the lack of connection between a subjectivist account of the truth-maker of reasons and an account of rationality. But, it might be thought, surely there must be some interesting connection between reasons and rationality. I will here canvass three attempts to forge such an interesting connection. I will be doubtful that any of these links can be sustained.

First, one could offer a connection of the sort that Michael Smith champions.<sup>36</sup> Smith offers an interesting proposal that attempts to forge a formal link between rationality and reasons. He offers a vantage point alleged to be ideal for deliberation about one's reasons. He then claims that if from that vantage point one would want X for one's nonidealized self, then one has reason to get X. He also claims that if one believes that one would so want X from his vantage point, then one is irrational to not desire X. Thus the connection between reasons and rationality would

35. Daniel Kahneman and Amos Tversky have written much about actual human heuristics and biases. See esp. Kahneman, Slovic, and Tversky, *Judgment under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982). Such heuristics and biases will, of course, sometimes lead us to act contrary to our genuine reasons but this is not sufficient to show that adopting them as part of one's decision procedure is irrational. It would be a drastic mistake to reject all such heuristics and biases on the grounds that even the best are fallible.

36. Smith, *The Moral Problem*, pp. 177–80. Gauthier (p. 31), in passing, also supports such a view.

be that if O is the case one has a reason to act in a certain way. If one believes that O is the case one would be irrational to not act in the way one would have reason to act if O were the case.

This seems to involve the tacit assumption that one cannot rationally reject the thought that one has reason to  $\phi$  if one would want to  $\phi$  from Smith's vantage point. Smith writes, "Suppose we believe that we would desire to  $\phi$  if we were fully rational and yet fail to desire to  $\phi$ . Are we irrational? We most certainly are. And by our own lights. For we fail to have a desire that we believe it is rational to have."<sup>37</sup> But why is the agent irrational by her own lights in this scenario? She would be irrational by her own lights here only if she accepts the thought that it is rationally required for her to desire as her fully informed counterpart would have her desire. Smith tacitly assumes that she accepts this thought. If she rejects this thought, however, then she is not irrational by her own lights to fail to desire to  $\phi$  in the above scenario. Further, an agent clearly can sensibly reject the thought that she is rationally required to desire as her fully informed counterpart would have her desire.

Indeed it seems to me that Smith's view here confronts a dilemma. Smith could either say that no one can rationally dispute that one has a reason to  $\phi$  if one would want it in the relevant way from the vantage point Smith describes. Once we remember that one can rationally dispute the truth of a claim even in a wide array of cases in which that claim is true, any plausibility to this thought evaporates. In order to disagree, Smith would have to argue that he has some sort of guarantee that no thought that disputes his above claim can even meet the standards of rationality. This would be bold stuff indeed.<sup>38</sup>

Alternatively, Smith could claim that although an agent can rationally dispute that her wanting to  $\phi$  from Smith's vantage point gives her reason to want  $\phi$ , she is nonetheless irrational if she believes that she would so want  $\phi$  and does not therefore want  $\phi$ . But how can this be? If she is indeed rational in doubting that her wanting to  $\phi$  from Smith's vantage point signals that she has a reason to  $\phi$ , how can it be thought to be necessarily irrational for her to combine a belief that she would want to  $\phi$  from Smith's vantage point with a failure to be motivated to  $\phi$ ? If

37. Smith, *The Moral Problem*, p. 177.

38. David Copp, in "Belief, Reason, and Motivation: Michael Smith's *The Moral Problem*," *Ethics* 108 (1997): 33–54, offers comparable considerations against Smith. Copp considers Smith's attempt to insist on an identity between (1) thinking one has a reason to  $\phi$  and (2) thinking one would be motivated from Smith's vantage point to  $\phi$ . Copp persuasively argues that even if there were an identity between 1 and 2 above, this would not solve Smith's problem. Even if Clark Kent is Superman we are not rationally required to treat them as the same person unless available evidence rationally requires us to believe that Clark Kent is Superman. Smith needs the identity between 1 and 2 to not only be true but to play a role in the agent's mental economy. Smith at least needs the claim that one cannot rationally reject the thought that 1 is identical to 2 and not merely the claim that they are identical. See Copp, pp. 38–43.

one can rationally reject the claim that a certain consideration gives one a reason, one cannot be held to be irrational for failing to give that consideration weight in one's deliberation.

Smith's case for the claim that "if we believe that we would desire to  $\phi$  if we were fully rational then we rationally should desire to  $\phi$ " trades on an ambiguity.<sup>39</sup> If we suppose that the agent whose reasons are in question supplies the content of "fully rational," then the thesis is plausible but uninformative. If an agent accepts that she would want to  $\phi$  if she deliberated in a way that she herself accepts to be ideal, then plausibly she is irrational to fail to be motivated to  $\phi$ . However, on Smith's view, Smith's theory, not the agent, supplies this content. But if the content of "fully rational" is supplied by Smith's account of what it is to be fully rational, then the account is informative but implausible.<sup>40</sup> This looks to be bad news indeed for Smith's claim to have solved "the moral problem" as the above equation between belief and rationality is explicitly held by Smith to be the key to solving the problem.

Second, one might think that necessarily there is some reliable tendency for rational deliberators to arrive at the truth about their reasons. As yet I see no reason to believe this. Of course we could just define rationality as the method of deliberation that tends to lead one to the right answer about one's reasons. This will be the third method of connecting reasons and rationality that I will dispute below. For now understand rational deliberation to have no conceptual connection to one's true reasons. One still might claim that apparently rational practices such as avoiding intransitive preferences or avoiding weakness of will necessarily have some tendency to lead one in the direction of one's reasons. But consider that logically valid reasoning need not have any general tendency to lead one to true beliefs if the true premises are epistemically inaccessible. And in the case of the subjectivist's account of reasons the true premises certainly do look to be epistemically inaccessible.

Third, one could simply define rationality or the best decision procedure as the one that best leads to one acting in accord with one's true

39. Smith, *The Moral Problem*, p. 177.

40. For a fascinating discussion of the extent to which an internalist account of the truth-maker must not only specify an idealization process and claim that being motivated to  $\phi$  from that vantage point is necessary for having a reason to  $\phi$  but must also ensure that the idealization procedure itself connects up with the agent's concerns in the right way, see Connie Rosati's "Internalism and the Good for a Person," *Ethics* 106 (1996): 297–326. Roughly, Rosati's view can be seen as an attempt to ensure that a rational agent would be motivated by her beliefs about what she would want after ideal deliberation by allowing the agent's current concerns to inform what counts as ideal deliberation for her. Rosati calls such a view "two-tier" internalism. If Smith accepted Rosati's two-tier internalism he might be able to get the connection between reasons and rationality he was looking for but perhaps at the price of losing the cognitivism of the view. Rosati herself tells us that her view "makes room for what might be a quite plausible antirealism" (p. 325).

reasons. I do not want to be fighting definitions. However, one should notice that on such a view any epistemic inaccessibility of one's true reasons will simply translate into epistemic inaccessibility in one's account of rationality. If the truth-maker is epistemically inaccessible on this view, then the proper decision procedure will also be. One could again claim to be offering merely an account of what makes it true that one is rational. But eventually surely we will want a term for saying that someone has acted sensibly or not *given* her epistemic situation. But this third proposal for linking reasons and rationality deprives us of the notions of rationality and irrationality for this purpose.

Some have supposed that Williams offered a fourth possible connection between reasons and rationality. On this reading, Williams is claiming that internalism and only internalism can ensure that one's reasons are rationally available to one. Thus it is supposed that Williams holds that on an externalist view it is possible for an agent to deliberate rationally and still be unmoved by her reasons whereas this is alleged by him to be impossible on an internalist view.<sup>41</sup> But this confuses Williams's argument. Williams and other subjectivists are crucially committed to the thought that an agent can deliberate rationally and yet fail to be moved by her true reasons when, for example, she excusably lacks important factual information. Korsgaard writes that "practical-reason claims, if they are really to present us with reasons for action, must be capable of motivating rational persons. I will call this the *internalism requirement*." Although Korsgaard does not directly attribute this form of internalism to Hume or Williams, she writes that "the internalism requirement is correct, but there is probably no moral theory that excludes it."<sup>42</sup> But real reasons need not motivate rational agents (when, e.g., they excusably lack important information) and hence Korsgaard's version of internalism is false. Further, if I am right about the best understanding of the Humean program, it claims that there are reasons for action that do not respect Korsgaard's version of internalism. The sensible subjectivist should argue that there is a crucial connection between being motivated after *ideal* deliberation and one's reasons.<sup>43</sup> I do not claim that the subjectivist cannot champion an interesting connection between reasons

41. See Rachel Cohon, "Internalism about Reasons for Action," *Pacific Philosophical Quarterly* 74 (1993): 265–88. Although Cohon (wrongly, I would say) supposes that this is Williams's argument, she (rightly, I would say) sees that this purported disanalogy with external views cannot be sustained.

42. Christine Korsgaard, "Skepticism about Practical Reason," *Journal of Philosophy* 83 (1986), quotations on 11 and 23, respectively.

43. Obviously the subjectivist does not want to suggest that there are substantive standards for desire that are specifiable independently from ideal deliberation (in the way that many think that truth does supply deliberation independent standards for belief). See David Velleman, "The Possibility of Practical Reason," *Ethics* 106 (1996): 694–726. The point is that what counts as ideal deliberation in this sense will be very different than what counts as rational deliberation.



and nonideal but rational deliberation. But I do not yet see how a persuasive version of this thought would go.

## II. KORSGAARD'S CRITIQUE OF HUMEAN INSTRUMENTALISM

In this final part of the article I will try to show that the failure to keep in mind the distinction insisted on above between an account of rationality and an account of an agent's reasons and/or failing to see that Humeans are offering an account of the latter has misled even our best philosophers into misunderstanding the subjectivist's account of reasons for action.

This section of the article will have three subsections. Subsection A considers objections to Korsgaard's critique of instrumentalism that are at some remove from the considerations raised above. The point of subsection A is to argue that Korsgaard does not have other good arguments against instrumentalism that are not subject to criticism for failing to appreciate what has been argued above. Subsection B resumes the main line of thought of the article by showing how what has been argued in Section I above tells against Korsgaard's critique of instrumentalism. I argue here that properly understanding Humean instrumentalism as an account of reasons rather than rationality helps us see that Korsgaard's critique simply fails to make contact with such views. Subsection C argues that, having seen that Korsgaard's arguments are misdirected against Humean accounts of reasons for action, we cannot successfully redeploy Korsgaard's arguments against Humean accounts of rationality.

### *A. Instrumentalism and Real Ends*

Christine Korsgaard's principal complaint against subjective accounts of reasons for action in "The Normativity of Instrumental Reason" is that they cannot serve as guides, that is, normative instructions that we might fail to heed. This objection, if it could be successfully made out, would clearly be devastating. She writes,

The problem is coming from the fact that Hume identifies a person's end as what he wants most, and the criterion of what the person wants most appears to be what he actually does. The person's ends are taken to be revealed in his conduct. . . . [This] problem would be solved if we could make a distinction between what a person's end is and what he actually pursues. Two ways suggest themselves: we could make a distinction between actual desire and rational desire, and say that a person's ends are not merely what he wants, but what he has reason to want. Or, we could make a more psychological distinction between what a person thinks he wants or locally wants and what he 'really wants.' . . . But in order to distinguish rational desires from actual desire, it looks as if we need to have some rational principles determining which ends are worthy of preference or pursuit. So the first option takes us beyond instru-

mental rationality. . . . But really the second option—the claim that these people are irrational because they do not promote the ends which they ‘really want’—also takes us beyond instrumental rationality, although this may not be immediately obvious. If we are going to appeal to ‘real’ desires as a basis for making claims about whether people are acting rationally or not, we will have to argue that a person ought to pursue what he really wants rather than what he is in fact going to pursue. That is, we will have to accord these ‘real’ desires some normative force. It must be something like a requirement of reason that you should do what you ‘really want,’ even when you are tempted not to. And then, again, we will have gone beyond instrumental rationality after all.<sup>44</sup>

Hume’s case of the unappetizing fruit is sufficient to show that he distinguishes between what one chooses and one’s real end. After all, the point of the example is that choosing the fruit is contrary to reason because doing so is not efficient in achieving one’s genuine ends. Hume does not say that because one chose the fruit it must be one’s real end. The accusation that Hume or the Humean is saddled with a “revealed” account of a person’s wants is ungenerous and undermotivated. Hume and many of his followers instead identify an agent’s genuine end with what one would want after one knows the facts. They accord these desires normative force. Korsgaard suggests that such a move takes us “beyond instrumental rationality” because it takes one’s informed wants to determine one’s genuine ends and then claims that furthering one’s genuine ends has a kind of normative priority over one’s uninformed wants. If this does take us beyond an account of reason as instrumental, then the accounts of Hume and Williams are not instrumentalist.

One of Korsgaard’s key goals in this paper is to show that “the instrumental principle cannot stand alone. Unless there are normative principles directing us to the adoption of certain ends, there can be no requirement to take the means to our ends.”<sup>45</sup> And this contains, I think, an important truth. If the instrumentalist had nothing to say about how to distinguish one’s real ends from one’s apparent ends, then there could be no requirement to take the means to our real ends rather than our apparent ends. But if the instrumentalist really was precluded from distinguishing real from apparent ends, then the revealed preference account that Korsgaard saddles Hume with would be just as precluded as any other understanding of an agent’s true ends.

Korsgaard seems to suppose that once the instrumentalist has something to say by way of distinguishing real from apparent ends, that she has gone beyond instrumental rationality because the instrumentalist

44. Korsgaard, “The Normativity of Instrumental Reason,” p. 230. Such claims are quickly summarized in Korsgaard’s *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), pp. 163–64.

45. Korsgaard, “The Normativity of Instrumental Reason,” p. 220.

would then be forced to recommend rational principles which do not flow from the instrumental principle itself. But the instrumentalist having something to say here is simply for her to interpret the instrumental principle, not to offer a second fundamental principle. The consequentialist does not add a second fundamental moral principle when she offers an interpretation of the good. The thought that the instrumental principle can stand alone should not be confused with the thought that it offers normative guidance even when left uninterpreted.

In any case, we should not think anything important hangs on how we count principles here. Perhaps some interpretations of one's true ends would violate the traditional subjectivist spirit of instrumentalism and so would interestingly take us "beyond instrumentalism." If someone added to instrumentalism an account of an agent's true ends such that necessarily all agents had the end of being moral, then the distinctive traditional subjectivism of instrumentalism would be lost. However, Hume and Williams identify one's ends with what one would want after one knows the facts. We can call this going beyond instrumentalism if we like, but doing so is misleading and the interesting differences with Kantian accounts of practical rationality remain.

Put another way, if Korsgaard is right that to be an instrumentalist one must forswear the project of distinguishing real from apparent ends, then instrumentalism is not a widely held view. The real interest would then be in Humean views rather than instrumentalist views. The former do allow themselves to distinguish real from apparent ends but deny that all agents share any interesting substantive ends.

Relatedly, it is sometimes suggested that the Humean cannot happily appeal to a differential authority between, for example, first and second order desires. Michael Smith writes, "For even if we assume that reason is on the side of a harmony between our first-order and second-order desires . . . there is simply no reason to assume that reason is on the side of achieving that *harmony* by changing our first-order desires to suit our second-order desires rather than *vice versa*. On the Humean's maximizing conception of rationality it all depends on which desire is stronger."<sup>46</sup> The idea seems to be that the Humean may only appeal to strength of desire but never to a measure of authority of desire that is not derived from strength. This is an odd attribution. For, as we have seen, a central Humean contention is that informed desires have more authority than uninformed desires. And this judgment of differential authority is clearly not derived from appeals to strength of desire. If the Humean is barred from appeals to differential authority between, for

46. Smith, *The Moral Problem*, p. 145. Smith goes on to consider the prospects of the Humean appealing to a differential authority between different kinds of desires. He seems unconvinced but merely concludes that the Humean would have to offer a rationale for the differential authority.

example, first and second order desires (as I say she is not), this cannot be because the Humean has forsworn the idea of their being authority to a desire that is independent of strength.

Korsgaard closes her long essay claiming that “only three positions are possible: either (i) the Kantian argument that autonomy commits us to certain substantive principles can be made to work; or (ii) we are left in the position of the heroic existentialist, who must ultimately define his will through acts of unconditional commitment that have no further ground; or (iii) complete practical normative skepticism is in order.”<sup>47</sup> But the position I have claimed Hume and Williams offer is none of these options. It supposes that there are principles that determine one’s ends, but that these principles do not require us to have any particular substantive ends.

### *B. Reasons not Rationality*

Korsgaard sometimes, as in the long paragraph quoted above, accuses Hume of identifying an agent’s end with whatever he actually chooses. But her considered view seems to be that Hume does not do this. She allows that one can have, according to Hume, an end different from what one chooses in at least two ways: (1) when one’s current wants are based on factual misinformation or (2) when one takes means insufficient to satisfy one’s ends.<sup>48</sup>

Korsgaard’s considered central argument against Hume and the Humeans is that neither of these ways of one’s actions coming apart from one’s ends can count as an instance of irrationality. That is, suppose for the sake of argument that it is always rational to take the means to one’s ends. Korsgaard’s claim is that the two ways for Hume that one’s act can count as not being the means to one’s end are such that they too cannot count as irrational. Thus Korsgaard claims that Hume and the Humeans lack the resources to label any action irrational.

Korsgaard claims that “both of these [ways in which Hume can make room for our action not being the means to our end] are cases of mistake; the actions that result from them are not, strictly speaking, irrational.”<sup>49</sup> Thus, Korsgaard is claiming that cases in which the agent is misled by false beliefs into choosing an action that she would not be tempted toward except for the misinformation are not, strictly speaking, cases of irrationality.<sup>50</sup>

47. Korsgaard, “The Normativity of Instrumental Reason,” p. 253.

48. *Ibid.*, p. 228. Again, I take it Korsgaard’s official view is that Hume thinks there is no such thing as practical reason. What she says here must be understood to be developing a Humean account that Hume himself found “improper.”

49. *Ibid.*

50. Korsgaard also claims in “The Normativity of Instrumental Reason,” p. 220, that “after all, a person may be conditioned to do the correct thing as well as the incorrect thing; but the correctness of what she is conditioned to do does not make *her* any more

Further, Korsgaard reminds us that Hume held that “the moment we perceive the falsehood of any supposition, or the insufficiency of any means our passions yield to our reason without opposition” (p. 416).<sup>51</sup> Korsgaard concludes that

This suggests that Hume thinks no one is ever guilty of violating the instrumental principle. Making a mistake, after all, is not a way of being irrational, and Hume thinks we do take the means to our ends as soon as mistakes are out of the way. But this is worrisome. How can there be rational action, in any sense, if there is no irrational action? . . . The problem is exacerbated when we see that Hume’s view is not just that people don’t in fact ever violate the instrumental principle. He is actually committed to the view that people cannot violate it.<sup>52</sup>

Thus, according to Korsgaard, Hume (and his followers) have no account of how an agent might act irrationally in light of the information that they are aware of, and this, according to Korsgaard, is to make no room for irrationality, strictly speaking.

Most of this argument is prefigured in Korsgaard’s earlier paper, “Skepticism about Practical Reason.”<sup>53</sup> There she wrote that “judgments of irrationality, whether of belief or action, are, strictly speaking, relative to the subject’s beliefs. Conclusions drawn from mistaken premises are not *irrational*.”<sup>54</sup> She also pointed out in the earlier paper that the two ways of one’s action not being the means to one’s end, for Hume, “look to be of this sort,” that is, look to be cases of conclusions drawn from mistaken premises. “Hume might, and in fact does, mean simply that we base our action on a false belief about causal relations. So this is no more genuinely a case of irrationality than the other [case of belief in non-existent objects]. Relative to the (false) causal belief, the action is not irrational.” Thus, the familiar conclusion: “If the only possibility Hume means to be putting forward here is the possibility of action based on false belief about causes and effects, we get a curious result. Neither of

---

rational.” She also writes that “the rationality of action depends on the way in which the person’s own mental activity is involved in its production, not just on its accidental conformity to some external standard” (p. 236). These passages make clear that Korsgaard’s concern is with the agent’s decision procedure, or what I have been calling the agent’s rationality, rather than with her reasons.

51. This additional premise is permissible if Korsgaard’s argument is intended only against David Hume. But Korsgaard makes clear that she intends to be arguing with Humeans quite generally. Against this broader position, the use of this premise is contentious. Surely most Humeans admit that Hume made some mistakes and this looks like one of them. Further, Korsgaard has not made a case that Hume’s commitment to this sentence runs deep.

52. Korsgaard, “The Normativity of Instrumental Reason,” p. 228.

53. Korsgaard, “Skepticism about Practical Reason.”

54. *Ibid.*, pp. 11–12. Emphasis in original.

the cases that Hume considers is a case of true irrationality: relative to their beliefs, people never act irrationally.”<sup>55</sup>

Korsgaard’s claim that attributions of rationality and irrationality are relative to the agent’s beliefs seems to me essentially correct. To demonstrate that an agent is irrational it does not suffice to show that her act was based on a false belief. Attributions of rationality and irrationality are directed toward the agent’s decision procedure and an agent could be employing the best available decision procedure and still sometimes base her actions on false beliefs. I agree with Korsgaard that such cases are not best thought of as cases in which the misled but diligent agent is being irrational. I agree, indeed I have been insisting, that attributions of rationality and irrationality should be relativized to an agent’s epistemic situation. No action is ever simply rational or irrational until we know what the actor knew or should have known.

Korsgaard’s central mistake in these two papers is that she fails to appreciate that Hume’s account is an account of an agent’s true reasons for action and not an account of rationality. Thus Korsgaard’s criticisms of Hume and his followers are misdirected and leave unaddressed the Humean theory because she misunderstands what such theories are accounts of. She is correct when she writes of the sorts of cases in which Hume thinks we act contrary to reason that “these are cases of [factual] mistakes; the actions that result are not, strictly speaking, irrational.”<sup>56</sup> Yet such actions certainly do count as contrary to reason for the Humean.

Recall that Williams, in the context of the petrol example mentioned above, allowed that the agent that drank the petrol and tonic would be “relative to his false belief, acting rationally.”<sup>57</sup> This did not keep him from claiming that the agent had no reason for drinking the petrol. To make sense of this combination of claims we must understand Williams to be distinguishing rationality from an agent’s genuine reasons. His principal question concerns an agent’s reasons for action and not her rationality, as is made clear by his attention being riveted by what there is to say against drinking the petrol.

Once we understand the Humean claims to be about reasons for action we can see how it is possible to fail to comply with the normative recommendations of the view. Mistaken facts are, as Korsgaard says, not (at least typically) sources of irrationality, but they are all too familiar sources of failures to act as one has genuine reason to act. Korsgaard writes as if the Humean instrumental principle only sanctions taking the means to one’s ends when one has information that reveals what the true

55. *Ibid.* This general thrust of this paper leans less heavily on these thoughts than the later paper. Thus I think it significantly less hurt by my case. Indeed, for what it is worth, I find the earlier paper’s central points convincing.

56. Korsgaard, “The Normativity of Instrumental Reason,” p. 228.

57. Williams, “Internal and External Reasons,” p. 103.

means to one's ends are. I have argued that the Humean instrumental principle is better read as claiming that we have a reason to take the means to our ends whether we recognize what the true means are or not. Korsgaard's remarkably ungenerous assertion that "Kant, unlike followers of Hume, recognizes that we cannot be guided by an imperative unless we can also fail to be guided by it" only seems plausible if we mistakenly assume that the topic of discussion in Hume and Williams is rationality.<sup>58</sup> As we have seen, when we properly understand the topic of discussion it is clear that Hume and Williams leave plenty of room for people to fail to act as they have genuine reason to act.<sup>59</sup>

### *C. Problems for Humean Accounts of Rationality?*

One might concede that the Humean instrumental principle is an account of reasons for action and not an account of rationality, and hence concede that Korsgaard's argument leaves the Humean view of genuine reasons unaddressed, yet insist that Korsgaard's argument provides real trouble for a Humean understanding of rationality. If it were true that accepting a Humean understanding of reasons for action left one incapable of developing a compatible theory of rationality according to which people sometimes act irrationally, this would be a serious criticism of the Humean program. However, I am unconvinced that the Humean cannot develop such an account for two reasons.

First, before such a line could be compelling we would need to understand what would make an understanding of rationality Humean. I find this a challenging question. I have been arguing that Hume's own writings do not provide an account of rationality, and if this is right what counts as a Humean account of rationality will involve some subtlety of interpretation. Korsgaard does not help us in this task.

Second, it would seem that Korsgaard does help us see what the instrumentalist's first line of defense should be against the charge that she cannot make room for irrational action. This would be to point to the possibility of what Korsgaard calls "true irrationality." This is the "failure to be motivated by the consideration that the action is the means

58. Korsgaard, "The Normativity of Instrumental Reason," p. 238. Korsgaard, in *The Sources of Normativity*, pp. 163–64, reiterates that it is impossible to violate the hypothetical imperative as the instrumentalist must conceive it and concludes that therefore the instrumental principle cannot be a normative principle.

59. I suppose Korsgaard could continue to claim that although the Humean view understood as I describe it leaves plenty of room for us to act contrary to its recommendations, nonetheless it cannot serve as a useful practical guide in actual (rather than counterfactual) deliberation. My response to such a line is that this is not what the Humean view aspires to be. Perhaps it will be claimed that this very notion of a reason that might not be able to serve as a useful practical guide in actual deliberation is somehow bankrupt. I think, on the contrary, that this is a commonsense notion that we would have difficulty doing without. The comparisons I offer with the concept of well-being in Sec. I, subsection D, might be helpful here.

to your end.”<sup>60</sup> Korsgaard herself tells us that “one can admit the possibility of true irrationality and yet still believe that all practical reasoning is instrumental.”<sup>61</sup> This seems clearly correct, but it ruins the case that instrumentalists can make no room for irrational action.

Why did Korsgaard think that instrumentalism could make no room for true irrationality? It seems that there were two reasons. First, she claims that an instrumentalist can have nothing to say about how to distinguish real from apparent ends. I dealt with this above. But notice that if it were true, the instrumentalist could not label any action as irrational *or rational*. If we do not know what an end is, we do not know what it would be to take the means to it. It is thus unsustainable to suppose that instrumentalists held that the instrumental principle might serve as the sole principle of rationality without any understanding of what a real end is. Second, Korsgaard points to Hume’s own premise that “the moment we perceive the falsehood of any supposition, or the insufficiency of any means our passions yield to our reason without opposition” (p. 416). This does seem to rule out the possibility of true irrationality. But if a Humean abandoned this optional premise, the possibility of her finding cases of true irrationality is plain. Thus, Korsgaard must restrict her argument to David Hume and cannot, as was her aim, generalize to Humean accounts.<sup>62</sup>

### III. CONCLUSION

In conclusion, I have argued that we have good reason to reject the view that the famous Humean account is an account of rationality, but that it has it that no one can ever act irrationally. We can avoid this surprising conclusion by rejecting the assumption that the Humean account is best read as offering an account of rationality, and instead understand such accounts as accounts of an agent’s reasons for action.

When we do so interpret such accounts we can understand, as Korsgaard could not, how the instrumental principle can make recommendations that we might fail to follow. Further, we can then understand why Hume and Williams find such an interesting connection between an agent being factually misinformed and her acting contrary to her genuine reasons for action. Finally, I have argued that Korsgaard has not made a compelling case that the Humean lacks the resources to develop an account of rationality according to which people sometimes act irrationally.

Advocates and critics of Humean instrumentalism have misunderstood or not fully appreciated what the account is an account of. Prop-

60. Korsgaard, “Skepticism about Practical Reason,” p. 12.

61. *Ibid.*, p. 15.

62. Korsgaard tells us that Hume is not the real target of her critique but rather Humeans in “The Normativity of Instrumental Reason,” pp. 229–30, n. 36.



erly understanding the nature of the account helps us see that some criticisms of the view are misdirected. However, nothing I have argued for here has aimed at vindicating Humean accounts of reasons for action. I suspect that such accounts are importantly correct. Yet the burden of vindicating such an account seems to me no less heavy now that we better understand the view. Nonetheless, arguments for or against the view are more likely to be fruitful if we properly understand the nature of the view we are arguing about.