

## *The Simple Desire-Fulfillment Theory*

MARK C. MURPHY  
Georgetown University

An account of well-being that Parfit labels the ‘desire-fulfillment’ theory (1984, 493) has gained a great deal of support as the most plausible account of what makes a subject well-off. According to the desire-fulfillment, or DF, theory, an agent’s well-being is constituted by the obtaining of states of affairs that are desired by that agent.<sup>1</sup> Importantly, though, while all DF theorists affirm that an account of what makes an agent well-off must ultimately refer to desire, there now appears to be a consensus among those defending DF theories that it is not the satisfaction of the agent’s *actual* desires that constitutes the agent’s well-being, but rather the satisfaction of those desires that the agent would have in what I will call a ‘hypothetical desire situation.’ Just as Rawls holds (1971, 12) that the principles of right are those that would be unanimously chosen in a hypothetical choice situation, that is, a setting optimal for choosing such principles, defenders of DF theory hold that an agent’s good is what he or she would desire in a hypothetical desire situation, that is, a setting optimal for desiring.<sup>2</sup> While the precise nature of the hypothetical desire situation is a matter of debate among DF theorists, all of them seem to agree that any adequate DF theory will incorporate a strong information condition into the hypothetical desire situation. In treating of the concept of an individual’s good, Sidgwick writes:

It would seem. . . that if we interpret the notion ‘good’ in relation to ‘desire,’ we must identify it not with the actually desired, but rather with the desirable:—meaning by ‘desirable’ not necessarily ‘what ought to be desired’ but what would be desired. . . if it were judged attainable by voluntary action, supposing the desirer to possess a perfect forecast, emotional as well as intellectual, of the state of attainment or fruition (1981, 110–111).

Brandt writes that a state of affairs belongs to an agent’s welfare only if it is such that “that person would want it if he were fully rational” (1979, 268); an agent’s desire is rational, on Brandt’s view,

if it would survive or be produced by careful ‘cognitive psychotherapy’ [where cognitive psychotherapy is the ‘whole process of confronting desires with relevant information.’]... I shall call a desire ‘irrational’ if it cannot survive compatibly with clear and repeated judgments about established facts. What this means is that rational desire... can confront, or will even be produced by, awareness of the truth (1979, 113).

And Railton has argued that we should consider an agent’s good to be “what he would want himself to want... were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality” (1986a, 16).

Now, it is clear even from these brief quotations from a small sample of DF theories that DF theorists differ in their accounts of the relevant hypothetical desire situation. Only Railton appeals to a hypothetical desire situation in which second-order desires are at stake. Sidgwick and Railton appeal to ‘perfect’ or ‘full’ information, while on Brandt’s view only knowledge of ‘established facts’ need be included in the hypothetical desire situation. But these differences should not distract us from the remarkable consensus reached among DF theorists both that DF theory should appeal not to actual desires but to desires had in a hypothetical desire situation and that the idealization of the information available to the agent will be a feature of that hypothetical desire situation.

The idea of employing a full information condition as part of the hypothetical desire situation has generated some of the same sorts of worry, though, that Rawls’ employment of a hypothetical choice situation has generated. Just as some (e.g. Sandel, 1982, 177–178) have argued that the person as appearing in Rawls’ hypothetical choice situation is too abstract, too ‘unencumbered’ by the particular contexts that make decision-making possible, it has been argued recently by David Sobel that full information accounts of well-being suffer from the fact that “the limitations which are idealized away by the full information account play a fundamental role in shaping our capacity to value in the ways that we do” (1994, 808–809). Just as some (e.g. Dworkin, 1989, 16–19) have questioned whether the decisions reached by imaginary contractors in Rawls’ original position bind actual flesh-and-blood people like us,<sup>3</sup> it has been argued that the desires of our selves in the hypothetical desire situation would lack authority over us: Connie Rosati holds that what is required for the agent to possess adequate information to assess possible ways of life would transform the agent in the hypothetical desire situation to the extent that his or her desires would not be normative for the actual agent (1995).

I do not mention these worries as anything like decisive arguments against DF theories. Rather, I bring them up only as difficulties for DF theory that can provide us with motivation to reconsider the consensus among advocates of that view that the appeal to hypothetical desire situations is essential to its defensi-

bility as an account of well-being. Call a 'Knowledge-Modified' DF theory any DF theory that affirms either of the following theses: that the satisfaction of certain of an agent's desires fails to contribute to that agent's well-being because that agent would lack those desires in some hypothetical desire situation in which he or she is better informed, or that the satisfaction of certain desires that the agent has in some hypothetical desire situation in which he or she is better informed (yet actually lacks) contributes to that agent's well-being. If the appeal to a hypothetical desire situation involving better information is essential to the defensibility of DF theory, then the success of arguments like Sobel's and Rosati's would imply the indefensibility of DF theory.<sup>4</sup> But there is, of course, the option of rejecting the presupposition that the information requirement is really necessary for DF theory. One could instead defend the 'Simple' DF theory: the theses that only the satisfaction of an agent's actual desires contributes to that agent's well-being and that no actual desire is to be excluded from relevance to an agent's well-being on the ground that the agent would lack that desire in a hypothetical desire situation.<sup>5</sup>

The Simple DF theory asserts a view stronger than merely the denial of the Knowledge-Modified DF theory: it asserts not only the irrelevance to well-being of hypothetical desire situations involving improved information but also the irrelevance to well-being of *any* hypothetical desire situation. A DF theorist might, therefore, reject both, opting for some sort of Modified view where the hypothetical desire situation does not include the agent's being better informed. Nevertheless, a successful argument against the Knowledge-Modified DF theory is a good argument for the view that if one is a DF theorist, one should affirm the Simple view. Of all the conditions that DF theorists have incorporated within their hypothetical desire situations, the information condition has been the most common and has been thought to be the weakest and most in the spirit of DF theory. A successful argument against the Knowledge-Modified form of DF theory is thus an excellent, if still *prima facie*, case against all Modified versions of DF theory, and, in turn, a strong case in favor of the Simple account—if, that is, one is to affirm DF theory at all.

My argument is this. If one is a DF theorist, holding that there is a tight connection between an agent's desires and that agent's good, one should prefer the Simple view, which appeals only to an agent's actual desires, unless there is a good reason for moving to a Knowledge-Modified view. But there are two possible reasons that the DF theorist could have for moving to the Knowledge-Modified view: either that possessing inaccurate information can lead agents to have desires whose satisfaction is irrelevant to well-being or that lacking accurate information can cause agents to lack desires whose satisfaction would be relevant to well-being.<sup>6</sup> The introduction of a hypothetical desire situation incorporating an information condition is supposed to be justified by its role in remedying these deficiencies. But, as I will show, neither of these grounds in fact gives the DF theorist reason to affirm a Knowledge-Modified view. Thus, the DF theorist should affirm the Simple view.

**The first rationale for the Knowledge-Modified DF view:  
that desires can be based on false beliefs**

One rationale that might be offered for moving from a Simple DF view to a Knowledge-Modified DF view is that an agent might have desires whose satisfaction is irrelevant to that agent's well-being because those desires are based on false beliefs. If it is true that a Knowledge-Modified view has the resources to explain why such desires are irrelevant to the agent's well-being while a Simple view has not, then there would be grounds for moving from a Simple to a Knowledge-Modified view. I will consider two senses in which a desire can be based on a false belief. In one sense in which a desire can be based on a false belief, it is true that the desire is irrelevant to the content of the agent's well-being, but the Knowledge-Modified view does not possess an account of the irrelevance of that desire superior to the account that the Simple DF theorist can provide. In another sense in which a desire can be based on a false belief, it is true that the Knowledge-Modified view implies the irrelevance of that desire to the agent's well-being while the Simple view allows its relevance. But in this case, the differences in implication favor the Simple view, for in this sense of 'based on' there is nothing objectionable about the idea that one's well-being is determined by desires that are based on his or her false beliefs. Since these are the only two ways, I shall argue, that desires can be based on false beliefs, it follows that the existence of desires based on false beliefs provides no rationale for moving from a Simple to a Knowledge-Modified DF view.

*Desires that are resultant upon false specificatory or instrumental beliefs.* Let us begin by considering a desire of very limited scope; in so doing we will avoid unnecessary complications without, I think, distorting any of the main issues involved. Suppose, for example, that I have a desire to own this particular baseball that is based on a false belief that it was autographed by Will Clark.<sup>7</sup> One way in which I may be said to have a desire for this baseball because of my false belief that Will Clark has signed it is the following: I desire a baseball with 'WC' signed on it; I falsely believe that this baseball has 'WC' signed on it; thus, because of my false belief that this baseball satisfies the description of the object that I desire, I am motivated to obtain this baseball; in virtue of my being motivated to obtain this baseball, it is the case that I have a desire for this baseball. Now, there is a way to construct an argument for the superiority of a Knowledge-Modified view on the basis of desires of this sort. Since it is obvious that my obtaining this particular baseball would not contribute to my well-being, the DF theorist wants to say that we need some way to rule out the desire's relevance to well-being. One way to fulfill this task is by appeal to a Knowledge-Modified view: since I would not desire this particular baseball were I to lack the false belief that it was signed by Will Clark, the Knowledge-Modified DF theory does not imply that the satisfaction of this desire contributes to my well-being. But it seems, by contrast, that the unmodified, Simple DF theory has the implausible implication that my owning this baseball as such

makes me better-off. Thus one might hold that there are grounds to prefer the Knowledge-Modified DF theory to the Simple view.

Even those writers that affirm some version of Knowledge-Modified DF theory are not likely to find this a very persuasive argument on behalf of that view. That a Knowledge-Modified DF theory implies that my desire for this baseball is not relevant to my well-being counts in favor of that view only if there is not some other, at least equally plausible explanation for the irrelevance of this desire available to the Simple DF theorist. But there is, according to many DF theorists, such an alternative explanation available: call it the 'basic desires' response. According to this view, DF theory should claim that only desires that have a certain place in the agent's motivational structure should count in determining that agent's good; it is only the satisfaction of an agent's *fundamental* or *basic* desires that makes that agent well-off. While I do not know precisely how to spell out what defenders of this view mean by 'basic,' we might offer the following examples of the sorts of desires that they mean to rule out as non-basic: those states of affairs desired by an agent merely because they are believed to be either *instances* of states of affairs already desired by the agent or *instruments* to the promotion of states of affairs already desired by the agent are not to be counted constituents of an agent's good. (I will call these desires, and the beliefs on which they are based, 'specificatory' and 'instrumental,' respectively.) Given the appropriateness of the basic desires response, it is clear that the Simple DF theory would have a way to explain why my desire for this baseball, which is based on my false belief that this ball was signed by Will Clark, is not relevant to the constitution of my well-being: it is that no desires that are based in this way on *any* beliefs, true or false, are relevant to the constitution of my well-being. Thus the Simple DF theorist can account for the irrelevance of such desires without requiring the machinery of a hypothetical desire situation.

Since the aim of this paper is to argue that there is no need to move from a Simple to a Knowledge-Modified DF theory, it may seem overscrupulous to worry about a response that is so uniformly endorsed and which provides a means of explaining why desires based in a certain way on false beliefs can be denied relevance to well-being in the absence of an appeal to the agent's hypothetical desires. But it does not seem that the propriety of the basic desires response is self-evident: why, precisely, is it that only basic desires count? Suppose an agent to have a basic desire that X obtain and to have a specificatory or instrumental desire that Y obtain, where Y's obtaining is believed by that agent to be either an instance of or an instrument to X's obtaining. What justification can the DF theorist give for holding that it is only the agent's desire for X whose fulfillment makes him or her better-off? Why isn't the fulfillment of the desire for Y a constituent of the agent's well-being also?

One argument that is clearly unsuccessful appeals to the idea of dependence: the reason that the desire for X is relevant and the desire for Y is not is that the desire for Y is dependent on the desire for X; the agent would not want Y unless that agent wanted X, and if the agent ceased to want X, the agent would no longer

want Y. But the mere fact of counterfactual dependence is surely not enough to warrant the normative conclusion that the satisfaction of one desire makes the agent better off while the satisfaction of the other does not. For, after all, this counterfactual dependence might be exhibited in the case of two unquestionably basic desires, where brute natural law might determine in the case of a particular agent that there is this sort of dependence of one desire upon the other.

Sometimes it is said that the source of the difference is that the agent wants X ‘for its own sake’ whereas the agent wants Y only for the sake of X. While there is no further ‘in order to’ explanation for the agent’s wanting X, there is a further ‘in order to’ explanation for the agent’s wanting Y: the agent wants Y in order to bring about X. But even if we were to grant the success of this argument with respect to instrumental desires, the argument is far less powerful with respect to specificatory desires: for in the most natural sense of ‘wanting something for its own sake,’ states of affairs that are the objects of specificatory desires are usually thought to be wanted for their own sakes, for such desires are not for states of affairs that are sought only in order to bring about further, distinct ends. It thus seems that this sort of argument fails to explain why the DF theorist should hold that specificatory desires are not relevant to agents’ well-being.

We still lack a reason to think that the basic desires response provides a justifiable way for the Simple DF theorist to account for the irrelevance of my desire for this baseball, given that it is a specificatory desire based on a false belief. But to press the point a bit further, it still seems to me to be far from clear why even instrumental desires are to be denied relevance. Given that a desire for X and an instrumental desire for Y are both desires, that for an agent X’s obtaining and Y’s obtaining both matter to that agent, that the agent is motivated both to promote X and to promote Y, why is the obtaining of X a part of the agent’s well-being while the obtaining of Y is not? One might say, as Parfit does in dealing with a desire-based theory of rationality, that

We should ignore *derived* desires. These are desires for what are mere means to the fulfilment of other desires. Suppose that I want to go to some library merely so that I can meet some beautiful librarian. If you introduce me to the librarian, I have no desire that is unfulfilled. It is irrelevant that you have not fulfilled my desire to join this library (1984, 117).

Parfit seems to be suggesting the following. If one has a basic desire and an instrumental desire, then the satisfaction of the basic desire—even if it is achieved in a way that does not involve the obtaining of the state of affairs desired instrumentally—will leave that agent with no unsatisfied desires. Thus the satisfaction of the instrumental desire is not itself worthy of consideration merely as such. But this argument also fails. Once again, it could be true with respect to two of an agent’s basic desires that, as a matter of sheer, brute natural fact, if one of these desires were satisfied, the other would be extinguished. But this sort of dependence alone is surely not enough to call into question the place of the

counterfactually-dependent desire with respect to one's well-being. If one is tempted by the view that this sort of dependence is relevant, the likely source is a confusion about what DF theories of well-being claim: they do not claim that being well-off consists in being a person with no unsatisfied desires, but rather that being well-off consists in the satisfaction of the desires that one has. Perhaps one might claim that once the desire to join the library is extinguished, joining the library is not an aspect of the agent's well-being. But the mere fact that this desire will be extinguished if one meets the librarian by some other means cannot be sufficient to call the relevance of that desire into question.

Here is another argument. One might appeal to the idea that to allow the relevance of instrumental desires in the consideration of an agent's well-being is to engage in illegitimate *double-counting* of desires. But this is clearly incomplete: what needs explaining is why this double-counting is illegitimate. If what fundamentally matters with respect to an agent's well-being is the satisfaction of his or her desires, and there are two distinct desires—one basic, one instrumental—present, why shouldn't both be counted in determining what makes that agent well-off?

Even if the basic desires response yields intuitively correct implications with respect to the constitution of an agent's well-being, what is missing is a principled account of why the DF theorist is entitled to this response. One way of providing such a response is to take literally a common manner of speaking: an agent that wants Y solely in order to satisfy his or her desire for X doesn't *really* want Y; what that agent *really* wants is X. To take this idea literally is, of course, just to hold that there are, in fact, no specificatory or instrumental desires; the only desires that exist are basic desires. If this thesis can be defended, then it is obvious why the DF theorist's basic desires response is appropriate: if there exist nothing but basic desires, then surely only basic desires are relevant to an agent's well-being.

The rationale for holding that the only desires that exist are basic desires derives from a certain plausible view on what the best principle of individuation for desires would entail about the possibility of specificatory or instrumental desires. It seems to me that a principle of individuation for desires must fulfill the following desiderata. First, it must be adequately responsive to the fact that desires are ascribed to an agent as explaining why that agent acts as he or she does. Secondly, it should be parsimonious: it should allow one to ascribe two desires to an agent rather than one only if the ascription of more than one desire will make a difference in the capacity to explain an agent's actions. Given these desiderata, I suggest the following principle of individuation for desires: for all putative desires A and B, A and B are distinct if and only if A and B together would have motivational force in addition to that which either of them alone would have.<sup>8</sup>

While this principle may lack the self-evidence of a Sidgwickian intuition, it is extremely plausible, at least once the following clarifications are made. First, with regard to the idea of 'additional motivational force': what I have in mind, roughly, is that the presence of both desires, rather than one or the other alone,

would add to either the *scope* or the *power* of the agent's motivation, where the scope of motivation is the set of states of affairs toward which the agent is motivated and the power of motivation is the degree to which the agent is motivated to promote some state of affairs. So, with respect to scope: if there is a single state of affairs toward which A and B together motivate yet toward which A alone does not, and a single state of affairs toward which A and B together motivate yet toward which B alone does not, then we will know that we are dealing with distinct desires. With respect to power: if there is a single state of affairs toward which A and B motivate the agent to a greater degree than A alone does, and there is a single state of affairs toward which A and B motivate to a greater degree than B alone does, then we will know that we are dealing with distinct desires. If A and B together would motivate the agent with the same scope and power as A alone would, then B is superfluous, and should not be ascribed to the agent along with A; if A and B together would motivate the agent with the same scope and power as B alone would, then A is superfluous, and should not be ascribed to the agent along with B.

Secondly, when one attempts to compare the motivational force of these desires and sets of desires, the background conditions that might affect the scope and power of the agent's motivation must be kept constant. Such background conditions include, most prominently, the agent's belief set.

Given these clarifications, this principle for individuating desires seems *prima facie* plausible. After all, desires are ascribed to agents as explaining why they act as they do. If a putative desire does not add at all to the motivational force of another desire, either by extending to further states of affairs or more powerfully moving the agent to a particular state of affairs, then it is hard to see what sense there is in the ascription of that additional, distinct desire to the agent, given the aim of explaining agents' actions.<sup>9</sup>

Now, consider the case of my desiring a particular baseball because of my false belief that Will Clark has signed it. Given the principle of individuation for desires that I have suggested, it is clear what the DF theorist should say: the DF theorist should deny that in this case I have a desire for this particular baseball. The reason is that my desire for a baseball with 'WC' signed on it and my alleged desire for this particular baseball fail the plausible test for individuating desires. As this case is described, the desire for a baseball with 'WC' on it and the desire for this particular baseball together have no motivational force beyond that which the desire for a baseball with 'WC' on it has on its own. The positing of this additional desire neither adds to the extent that I am motivated to promote any particular state of affairs nor adds to the set of states of affairs that I am motivated to pursue.

Why is this? Consider two sets of background conditions: one in which I believe that this particular baseball has 'WC' on it, and one in which I lack this belief. In the former case, my desire for a baseball with 'WC' on it along with my belief that this baseball has 'WC' on it explains my seeking to own this baseball; nothing is added by positing an additional desire for this particular baseball.



(Surely I am not motivated more strongly to have *this* ball than to have *a* ball with 'WC' on it?<sup>10</sup>) In the latter case, one might think that these desires pass the test for individuating desires; for, lacking the belief that this ball has 'WC' on it, my desire to own a baseball with 'WC' on it will not motivate me toward the state of affairs of my owning this particular baseball, but my alleged desire to own this baseball must motivate me toward the obtaining of that state of affairs. The problem, though, is that the existence of my desire for that particular baseball is inconsistent with the situation as described: as the situation is characterized, the existence of my desire for that particular baseball depends on my having the belief that this particular baseball answers to the description 'is signed by Will Clark.' Thus, there are not distinct desires present: while I have a desire for a baseball with Will Clark's signature on it, I have no desire for this particular baseball.<sup>11</sup>

One might be tempted to reject this conclusion on the basis of the sheer obviousness of the fact that I do desire this baseball. After all, I'm trying to get it; I'm motivated to obtain it. But what should be rejected is the inference from *being motivated to get X* to *having a desire for X*. Being motivated to get X is an event; the desire for X is a certain functional state. One's desire enters into an explanation for one's being motivated to act a certain way. While it is true (we may grant) that one's being motivated to get X implies that there is *some* desire that gives rise to that motivation, it does not imply that the desire responsible is a desire *to get X*. While I am motivated to get this baseball, then, it does not follow from my being thus motivated that I have a desire for this baseball. All that follows is that I have some desire that gives rise to this motivation. That desire is the desire for a baseball with Will Clark's signature on it, which motivates me to try to get this baseball in conjunction with, and only because of, the fact that I believe that this baseball was signed by Will Clark.

This account of the individuation of desires, and its implications with respect to those putative desires based on specificatory or instrumental beliefs, provides the DF theorist with the sought principled rationale for the basic desires response. Consider first the case of putative specificatory desires. The rationale for rejecting the relevance of any specificatory desire is clear from the treatment of the case in which I seem to desire a baseball because of my false belief that it was signed by Will Clark. For there is nothing in that argument that relied on the belief's being *false*; rather, the argument relied only on that belief's being *specificatory*. If, therefore, that argument was successful in establishing that there are no desires based in this way on false beliefs, then it was successful in establishing that there are no desires based in this way on true beliefs, either.

This principle of individuation underwrites the same sort of explanation for why the satisfaction of instrumental desires is not constitutive of an agent's good. Consider by way of example one's desire to pass a test, and a (putative) desire to study that is had as a result of the instrumental belief that studying is a means to doing well on tests. I claim that there cannot be these two distinct desires. According to the principle of individuation for desires, they are distinct only if these

two desires in tandem have motivational force in addition to that which either of them alone has. For these two desires in tandem to have additional motivational force, one of the following must be true: either the desire to pass and the desire to study together motivate the agent toward states of affairs beyond those that the desire to pass alone does, or the desire to pass and the desire to study together motivate the agent more powerfully to some state of affairs than the desire to pass alone does. But it seems that neither of these is the case. The desire to pass the test, together with the belief that studying is a means to passing tests, is sufficient to explain the agent's motivation to study. And if the motivation toward passing the test extends to the act of studying, it would be strange if there were more motivation to perform an act of studying than that provided by the desire to pass the test: after all, in the case as described, the whole point of studying is to pass the test. Thus, it follows that the desire to study is not distinct from the desire to pass the test. In the case as described, the agent has no desire to study.

Since this argument is obviously generalizable to any instance of a putative instrumental desire, its success would show that there are no instrumental desires.<sup>12</sup> Apart from the principle of individuation for desires upon which the argument relies, the most controversial premise is its claim that the desire to pass the test together with the belief that studying is a means to passing the test is sufficient to explain the agent's motivation to study. This is not so hard to see in the case of specificatory beliefs: if one wants an *x*, it seems sufficient to explain that person's motivation for this particular object that he or she believes this particular object to be an *x*. Perhaps what occasions doubt in the case of instrumental beliefs is that the motivation extends to a state of affairs that is at a distance from the state of affairs originally desired, where there is no such distance in the case of specificatory beliefs.<sup>13</sup> But it seems to me that to insist on the positing of a distinct desire to fill this gap is a mistake. Think of desires as pushing agents from the presently obtaining state of affairs toward the obtaining of another state of affairs. It seems that there is already present in desires thus conceived a latent motivation to employ some means believed to transform the present situation into one in which the desired state of affairs is realized. All that is necessary to make that motivation manifest is a belief about those means that will effect that transformation. No ascription of an additional, distinct desire to promote the means to the desired end is necessary.<sup>14</sup>

*Desires based on false beliefs in a merely causal way.* There are, of course, some cases in which a putative desire based on a false belief does pass the test for individuating desires. In such cases, the motivation to acquire the object does not depend, as in the previously described case, on the agent's being motivated to pursue the object *qua* satisfying some other desire. Rather, the false belief simply occasions a basic desire for that object. My false belief that this baseball has 'WC' signed on it might, in conjunction with other features of my psychological constitution and the natural laws governing desire-formation, generate a desire in me for this particular baseball. Given the special features of my psychological make-up, this desire could depend on the false belief in either of the following ways. It

could remain dependent on it, in that if I were to lose the false belief I would no longer desire the object. It could, on the other hand, be the case that while the desire was brought about by the false belief, it does not depend for its persistence on my continuing to hold the false belief.

Now, it seems perfectly likely that there could be desires of this sort, and that they could pass the test for individuating desires that I have offered. The Simple DF theorist could not hold that the satisfaction of such a desire does not contribute to my well-being on the grounds that it is not really a desire of mine at all. But this seems to me to be no problem for the Simple DF theory: in this sense of 'being based on a false belief,' it does not seem objectionable that one might have a desire whose satisfaction contributes to his or her well-being yet part of whose causal history includes one's believing something false. Why would one find a desire of this sort problematic? Why would the mere fact that the set of causal origins of a desire includes a false belief call into question the contribution of the satisfaction of that desire to one's well-being?

That DF theories should not be concerned to rule out desires generated by false beliefs in this sense is clearest in those cases where the desire persists even once the falsity of one's belief comes to light. Suppose that just as a matter of the laws of nature my falsely believing that Will Clark has signed this baseball generates in me a desire to own this baseball—or that any other state of affairs obtain (that I get a drink, that I write a philosophy paper, etc.). Later I come to hold the true belief that Clark did not sign this ball, but I continue to desire to own this ball (or to take a drink, or to write a paper, etc.) Is it not clear that the DF theorist lacks grounds for denying that owning that ball would contribute to my well-being? And while this may not be as clear in the case where my desire is, merely as a matter of fact, conditional on my holding the false belief—in part, I think, because these cases exhibit the same counterfactual dependence of desire on belief that holds when one desires something *qua* fitting a certain description—it seems that the DF theorist lacks any reason for holding that there is anything intrinsically disreputable about this sort of desire.

The defender of the Knowledge-Modified view owes the defender of the Simple view an explanation of why the fact that a false belief figures causally in the generation of a desire is sufficient to call that desire's relevance to well-being into question. Suppose that my previous arguments were mistaken, and that where I am motivated to seek this baseball *qua* signed by Will Clark due to my false belief that it was signed by Will Clark, there is a genuine desire for this baseball present. If such were the case, then there would be a natural explanation available for why the DF theorist holds that the satisfaction of my desire for this baseball does not contribute to my well-being. If I had a desire to own a baseball signed by Will Clark, and I formed a desire for this particular baseball *qua* signed by Will Clark, I would justify my having this latter desire on the basis of there being a logical relationship between the contents of those desires: 'my owning a baseball signed by Will Clark,' I would say, is an existential generalization of 'my owning this baseball.' But my assertion of this logical relationship between the contents of

my desires has its warrant only from the truth of my belief that this baseball was signed by Will Clark. Thus, it is clearly important to the justification of my desire to own this particular baseball that my belief that it is signed by Will Clark be true; and so, if my belief is false, the DF theorist has grounds for denying that the satisfaction of this desire contributes to my well-being.

This sort of account is clearly unavailable, though, in cases where false beliefs play a merely causal rather than a justificatory role in the generation of desires. Where a desire is justified by appeal to a prior desire and a belief that there is some logical relationship between the content of the two desires, it is apparent why the truth or falsity of the belief that played a role in generating the desire would be relevant. In the absence of such an attempt at justification, though, it is unclear why the fact that a desire was occasioned by a false belief would provide any reason to discount the place of that desire in an agent's well-being. After all, on a DF theory of an agent's good, what is good for an agent boils down, ultimately, to what that agent just wants. And on any DF theory, Modified or not, what agents want depends on all sorts of quirky facts about them—their physical constitution, their environments, the (apparently, anyway) whimsical arbitrariness of the content of the laws of nature—what is special about desires that have false beliefs as part of their causal history? Since these are not cases in which justification of the desire is called for by appeal to some other desire, it appears that the role of a false belief in the generation of the desire has no normative relevance above and beyond that of any other contingent fact in the absence of which the agent would lack that particular want.

Consider, for example, Brandt's attempt to rule out desires based on false beliefs where the sense of 'based on' is merely causal. Brandt, who holds that one's good consists in what one would desire for oneself if one were to undergo 'cognitive psychotherapy,' suggests that it is a matter of psychological law that some desires caused by false beliefs will fade away if placed in the light of truth. He offers an example of a person who decided to work in an academic profession because he believed that such a life is what his parents wished of him; but, as it turned out, they didn't wish such a life for him. Nevertheless, he has now come to desire an academic life for its own sake. Brandt claims that this desire will be extinguished if "the person repeats to himself the fact that he will not achieve the goals involved in instituting the desire [e.g. pleasing his parents] by doing a certain thing [e.g. becoming an academic]"; thus, the satisfaction of this desire is not to be accounted part of his good, for it would not survive cognitive psychotherapy (1979, 116).

It seems unlikely that a DF theorist can provide an adequate rationale for holding that the satisfaction of this desire does not constitute part of the academic's good. I do not deny that such a rationale could be found during the time in which he wanted an academic life only because he believed that such a life fit the description 'doing what my parents want me to do': during that time, the academic would justify that desire by appeal to his belief that his parents wanted such a life for him, and the falsity of that belief would provide grounds for de-

nying that the satisfaction of the desire for a scholarly career has a place in the academic's well-being. But once the academic desires that life for its own sake, the place of the false belief in the generation of the desire is merely causal; and it is unclear why the truth or falsity of the belief would have any bearing on whether the satisfaction of the desire would make the academic well-off. Why would the fact that, as a matter of brute psychological law, the repeating of 'I first sought an academic career to please my parents, but I was wrong to think it the way to please my parents' would extinguish the intrinsic desire for an academic career give us any reason to doubt that the satisfaction of that desire contributes to that agent's well-being?<sup>15</sup>

I thus conclude that the Knowledge-Modified DF theory lacks a rationale for its hypothetical desire machinery in dealing with desires based on false beliefs. With regard to one way that desires are said to be based on false beliefs, there are in fact no such desires, and thus no reason to try to rule out such desires by invoking a hypothetical desire situation. With regard to the other way that desires can be based on false beliefs, there are such desires, but the DF theorist offers no grounds for doubting that the satisfaction of such desires does contribute to the agent's well-being.

**The second rationale for the Knowledge-Modified DF view:  
that desires can be absent due to a lack of true beliefs**

The fact that some desires are based on false beliefs gives the DF theorist no reason to move from the Simple view to the Knowledge-Modified view. It is sometimes suggested, though, that since agents often fail to have certain desires because they lack important pieces of information, some kind of idealization is necessary: the states of affairs whose obtaining constitutes one's well-being are those that satisfy the desires that one would have if one possessed a more adequate stock of true beliefs. Suppose that I lack a desire to own this particular baseball, but if I possessed the true belief that this baseball has Will Clark's signature on it, I would desire it. One might reject Simple DF views on the basis that idealization of desires in a hypothetical desire situation is necessary to capture the fact that possessing this baseball would contribute to my well-being. The Simple DF theorist can reject the necessity of this idealization by using only slightly altered formulations of the arguments used against the first rationale for the Knowledge-Modified view. For, once again, there are two sorts of case that fit the description of the situation that I have described. In one sort of case, there is no need to appeal to idealization to capture the plausibility of the view that owning that baseball contributes to my well-being; on the other sort, it is not implausible that owning that baseball is, in this situation, not relevant to my well-being at all.

Consider the most straightforward version of the case that I describe. Suppose that I have a desire for a baseball with 'WC' on it, and this particular baseball has 'WC' on it. Since I do not believe that this ball has 'WC' signed on it, however,

I lack a desire for this particular ball. But it might be thought that since I lack a desire for this ball, the contribution to my well-being of my owning that ball cannot be explained unless my desire-set is idealized to what I would desire if I had true beliefs, including the belief that this ball has 'WC' signed on it. This line of reasoning is mistaken, though. First, this argument supposes that, in the hypothetical desire situation of more complete information, the agent would have a distinct desire for this particular baseball. But, once again, since the desire for a baseball with 'WC' on it and the alleged desire for that particular baseball would together have no more motivational force than the desire for a baseball with 'WC' on it alone, these desires could not be distinct. Secondly, there is no need to appeal to idealized desires in this type of case. If one wishes to explain the contribution of owning this baseball to my well-being, one need appeal only to the facts that I have a desire for a baseball with 'WC' on it and that this baseball really does have 'WC' on it. Reliance on a desire formed only in a hypothetical desire situation is completely otiose.

There is another sense, though, in which lacking certain information might be said to cause one to lack a certain desire. In the case just described, what is missing is information about what will, in fact, satisfy one's desires; one is lacking true specificatory beliefs. In a different sort of case, possessing additional true beliefs would transform the desires that one has. Just as holding false beliefs might causally contribute to the possession of a certain distinct, basic desire, having true beliefs might do the same. It might be the case that even though I at present have no desire for a baseball with 'WC' on it, or a desire for this baseball, it is a matter of sheer psychological fact that if I were to come to believe that this baseball has 'WC' on it, I would come to desire this baseball. It seems obvious that in this type of case the Simple DF theorist has no way to explain how owning this baseball would contribute to my well-being in the absence of my actually coming to have that desire. But it also seems implausible that the Simple DF theorist needs to explain this contribution, for it does not seem that owning this baseball contributes to my well-being at all. Why is it relevant to the present constitution of my well-being that, as a matter of sheer psychological fact, if I were to form this true belief, then I would desire that ball?<sup>16</sup>

Even though it is a mistake to think that, in the case in which one lacks a specificatory belief, an appeal to a hypothetical desire situation of better information is needed (or even useful), the rationale behind that maneuver is clear. Just as one who affirms *p*, where *p* entails *q*, is committed to affirming *q*, it seems that one who desires *X*, where *Y* is an instance of *X*, is (in some analogous sense) committed to desiring *Y*. But no such rationale is available in the case of desires that are formed as a matter of brute causal fact upon coming to hold true beliefs. The Knowledge-Modified DF theorist, in my view, can offer no account of why the Simple DF theorist should be at all concerned to accommodate such hypothetical desires as relevant to anyone's good. Thus, the upshot is clear: just as the Knowledge-Modified DF theorist provides no rationale to reject the Simple DF view on account of desires possessed due to false beliefs, that theorist also pro-

vides no rationale to reject the Simple DF view on account of desires lacked due to an absence of true beliefs.

### **Does Railton's view support the move from a Simple to a Modified DF theory?**

We find, I claim, no rationale for the move from a Simple to a Knowledge-Modified DF theory in either desires present due to false beliefs or desires absent due to a lack of true beliefs. It might be claimed, though, that the argument as developed thus far has proceeded without sufficient attention to the sort of Modified DF view advanced by Railton, whose account of the role of desire in understanding the content of well-being differs in two important ways from the sort of DF view that I have considered up to this point. First, on Railton's view the role of desire with respect to the content of an agent's good is less constitutive than it is on the standard DF view.<sup>17</sup> Secondly, on Railton's view, second-order desires—what an agent wants him- or herself to want—have a privileged place that they are not accorded in the generic conception of DF theory that I have focused upon. It seems to me, though, that the considerations already raised suggest that neither of these distinctive features of Railton's view supports an appeal to a Modified over a Simple DF theory.

The first way that Railton's version of DF theory differs from more standard versions is that Railton wants to say that the satisfaction of an agent's desires is not what constitutes that agent's good; rather, the agent's good is determined by what Railton calls his or her "objective interests" (1986b, 175). Terminologically, at least, it seems that Railton is far from DF theory's guiding idea, but the account of objective interests offered by Railton makes clear that a tight connection between the agent's well-being and the agent's desires is preserved. As we have already seen, Railton finds the agent's actual desires—his or her "subjective interests" (1986b, 173)—an inadequate starting point for an account of the agent's good: we should, rather, appeal to the agent's "objectified subjective interests," that is, what the agent would want him- or herself to want if he or she had "unqualified cognitive and imaginative powers, and full factual and nomological information about his [or her] physical and psychological constitution, capacities, circumstances, history, and so on" (1986b, 173-174). If Railton held that the agent's well-being were constituted by the satisfaction of his or her objectified subjective interests, his view would be straightforwardly a version of the standard Knowledge-Modified DF theory. But Railton holds that it is the *reduction basis* for the agent's objectified subjective interests, rather than those interests themselves, that is the truth-maker for correct claims of the form 'such-and-such is an aspect of this agent's well-being.' The reduction basis for the agent's objectified subjective interests—roughly, those facts about the actual agent that the idealized agent would employ to determine what the idealized agent would want the actual agent to want—determines what the agent's objective interests are, and the ob-

taining of the ends of these objective interests Railton identifies with the agent's good (1986b, 176).

Now, this way of implicating desire in an account of well-being is undoubtedly distinct from the way of implicating desire in well-being taken by a standard DF view. On the standard DF view, well-being is constituted by the satisfaction of desire; on Railton's view, well-being is constituted by the satisfaction of objective interests. Those states of affairs that constitute an agent's good are fixed by certain of the agent's (hypothetical) desires, but the nature of well-being is not defined by the satisfaction of those desires. I will not concern myself here with the issue of whether Railton's or the standard DF account of the relationship between desire and well-being is more defensible, though, for it seems to me that the same arguments that I have offered against the move from a Simple to a Knowledge-Modified version of the standard DF theory can be employed against the move from a Simple to a Knowledge-Modified version of Railton's objective interest account.

Here is why. For Railton to motivate adequately the locating of the agent's objective interests with reference to the reduction basis of the agent's objectified subjective interests rather than with reference to the reduction basis of the agent's subjective interests, he must provide some rationale for doing so. But it seems that the argument of the previous two sections has cut off the most straightforward route: that the full information stipulated in the case of objectified subjective interests yet possibly not present in the case of subjective interests might cause an agent either to lose desires irrelevant to well-being that he or she otherwise would have had or to have desires relevant to well-being that he or she otherwise would have lacked.<sup>18</sup> While Railton's take on the relationship between well-being and desire clearly makes his view a distinctive version of DF theory, the fact that his Knowledge-Modified view offers the same rationale for rejecting the Simple view that is offered by more standard conceptions of Knowledge-Modified DF theory leaves it open to the same line of objection that I have pressed against more standard Knowledge-Modified DF views.

The other distinctive feature of Railton's view is its reliance upon second-order desires as being of controlling importance in fixing the content of the agent's well-being. An agent's good, on Railton's view, depends on what that agent would, from a fully informed standpoint, want him- or herself to want. This view is doubly Modified, for it involves appeal to two hypothetical desire situations. The first is that an agent's good is determined not by the agent's actual desires, but by the desires that the agent would have if he or she were to desire in accordance with his or her second-order desires. The second is that the relevant second-order desires are not the agent's actual second-order desires, but those that the agent would have in conditions of full information. But it seems to me, once again, that the earlier argument against the employment of knowledge conditions in DF theory suffices to undercut Railton's appeal to second-order desires in determining the content of an agent's good.



Consider Warren Quinn's remarks criticizing those subjectivist views that hold that while an agent's desires generate reasons for action, it is only those desires that are approved, or at least not disapproved, by second-order desires that should count as reasons. He imagines a pyromaniac who looks upon his own fire-setting tendencies with distaste. Does the subjectivist have an adequate rationale to hold that the pyromaniac's disapproval of his desires to burn things precludes those desires from constituting reasons to act?

I think that it is very doubtful that a subjectivist can legitimately attach this significance to the existence or nonexistence of opposing higher-level attitudes. Here, as elsewhere, he is presupposing a significance that depends not on level but on content. . . . The subjectivist. . . can see nothing in the higher-level disapproval except more complexly structured psychological opposition, and such opposition would seem to leave the lower-level attitude securely in place with its own proper force (1993, 238–239).

Now, Quinn's criticism concerns subjectivism about reasons for action, but it is clearly applicable—or at least transformable—into a concern about those DF theories that appeal to second-order desires. Why is it, we might ask, that in the case of the imagined pyromaniac we should say that the pyromaniac's good does not include setting fires? What grounds can the DF theorist offer for saying anything more than that there is just a tension in the imagined pyro's motivational structure, so that while he is made better off by setting things afire, he is made worse off by continuing to desire to set things afire? Quinn's own view is that subjectivists that appeal to higher-order desires are covertly appealing to content, looking for some way to exclude the pyromaniac's first-order desire from constituting a reason for action on the basis of its unsavory object. But even if we refrain from ascribing this motive to the DF theorist that appeals to second-order desires, Quinn's concern remains: how is the DF theorist to explain "the *authority* of the higher-level attitudes"? (1993, 239)

It seems to me that a second-order desire view like Railton's will not be able to explain this, if the arguments that I have already made against Knowledge-Modified views are sound. Railton's case for the authority of the second-order desires depends on those desires' being more subject to influence by the nature of our beliefs, so that they can have a claim to be better justified than the particular first-order desires that are their objects (1986a, 16). What guarantees that the second-order desires that determine the content of the agent's good will not be merely *potentially* more authoritative, but *actually* so, is his appeal to a hypothetical desire situation of full and vivid information in which the agent's second-order desires are formed. But if my earlier arguments against the Knowledge-Modified view were correct, then no differences between an agent's actual second-order desires and an agent's second-order desires in a hypothetical desire situation of more accurate and complete information would be of any normative importance. At the risk of being tedious: the only roles for a hypothetical

desire situation involving improved information would be either to rule out desires based on false beliefs or to bring into play desires that the agent would have if he or she possessed additional true beliefs. But neither role calls for the use of a hypothetical desire situation.

Suppose that a pyromaniac has a second-order desire to desire to set fires which is based on a false belief. On one hand, this desire might be possessed *qua* satisfying some other desire: one might desire to have desires whose satisfaction doesn't endanger others, and one might possess the false belief that setting fires doesn't endanger others. In this case, there really is no desire to desire to set fires present: the principle of individuation for desires implies that there is only one desire, a desire to have desires whose satisfaction does not endanger others. Thus, there is no need to appeal to a hypothetical desire situation to rule out this second-order desire, for this putative desire does not even exist. On the other hand, if the second-order desire were merely occasioned by some false belief, that would not seem to be sufficient grounds to rule it out: just as many of our first-order desires have strange causal histories, so too do many of our second-order desires. Again, there is no need to appeal to a hypothetical desire situation to rule out this second-order desire, for there are no principled grounds within DF theory for denying it a role in determining the content of the agent's well-being.

Suppose that a pyromaniac lacks a second-order desire not to want to set fires, but would have that desire if fully informed. If, on one hand, the agent already has a desire not to engage in activities that endanger others, and thus would desire not to want to set fires if he or she possessed the true belief that setting fires endangers others, we can say both that appeal to a hypothetical desire situation is unnecessary, because the agent already has a second-order desire that would condemn setting fires, and that this appeal is unhelpful, because there is in reality no additional desire generated by that true belief. If, on the other hand, some new true belief simply occasions a second-order desire not to want to set fires, it is unclear why we would think that what the agent would want in a situation in which he or she possesses this true belief would have any bearing on the agent's good while he or she lacks this true belief. Thus, the same considerations that render otiose appeal to a hypothetical desire situation to improve the agent's first-order desires render otiose appeal to a hypothetical desire situation to improve the agent's second-order desires.

How, though, does showing there to be no need to appeal to a hypothetical desire situation of improved information render dubious Railton's Modified DF theory? After all, one might still claim that only those first-order desires that would be held by an agent if he or she were to desire in accordance with his or her second-order desires determine the nature of the agent's well-being. But this leaves unanswered Quinn's question of the authority of those second-order desires. Without the higher level of justification of the second-order desires that is supposed to come with a greater responsiveness to information and a hypothetical desire situation of full information to which those desires can respond, it appears that all

that is left is two different sorts of desires, without any grounds for asserting the preeminence of one over the other in determining the content of an agent's good.<sup>19</sup>

Does this mean that there is a stalemate between the defender of the Simple theory and the defender of a view like Railton's? No. The Simple DF theory has the fundamental idea of DF theory on its side: that what makes an agent well-off is determined by looking at what the agent wants. Since the defenders of the second-order desires view want to rule out the relevance of some desires to the content of an agent's good, it is up to them to provide an account of why the DF theorist should follow that path. Without such an account, the Simple DF theory is the superior version of DF theory.

### **The Simple desire-fulfillment theory and the critical character of the good**

Railton writes that the Simple DF theory "has many virtues: it is uncomplicated, nonpaternalistic, and epistemically as straightforward as the idea of desire." Unfortunately, he says, "this theory is deeply unsatisfactory, since it seems incapable of capturing important elements of the critical and self-critical character of value judgments" (1986a, 11).<sup>20</sup> And James Griffin has written that the Simple DF view is doomed because

we mistake our own interests. It is depressingly common that when even some of our strongest and most central desires are fulfilled, we are no better, even worse, off. Since the notion we are after is the ordinary notion of 'well-being', what must matter for utility will have to be, not persons' actual desires, but their desires in some way improved. The objection to the actual-desire account is overwhelming (1986, 10).

These are distinct concerns. Railton's appeals primarily to a formal constraint on theories of the nature of well-being: such theories must be able to capture the fact that the notion of well-being is used to criticize one's passing desires. Griffin's, on the other hand, appeals to particular beliefs that each of us has: that on some occasions we have mistaken our own interests, and on some occasions the satisfaction of our desires has made us worse rather than better off. But in both cases the form of response that the defender of the Simple DF theory should offer is the same: that the Knowledge-Modified view does no better in fulfilling these desiderata than a Simple view does. Given that the default setting of DF theory is the Simple view, it will suffice as a defense against these criticisms of the Simple view if it can be shown that the Knowledge-Modified view succeeds in providing an account of the critical character of the good no further than the Simple view does.

Consider first Railton's criticism: that while the standard of well-being is employed to criticize agents' particular desires, the Simple view does not explain how this could be so. Since the Simple view understands the agent's well-being to be constituted by the obtaining of the states of affairs actually desired by that

agent, well-being cannot serve to criticize those actual desires. Now, there is clearly a sense in which the Knowledge-Modified view provides a standard independent of actual desire where the Simple view does not: since it is possible for the agent's desires under conditions of full information to differ from the agent's actual desires, the Knowledge-Modified view satisfies the necessary condition for critical standards that they be possibly distinct from what they are used to criticize. But it is pretty obvious that it is not sufficient for the Knowledge-Modified view to be superior to the Simple view that it satisfies this necessary condition. It is surely not a sufficient justification for finding the Simple view inferior to the *Height-Modified* DF theory—an agent's good is determined by what the agent would want if he or she were of average height—that the Height-Modified view satisfies this necessary condition, merely on the basis that one's desires might be different if one were of average height. It is not sufficient for any Modified view to be superior to a Simple view merely that it provides a standard distinct from actual desire so that actual desires can be criticized: rather, it must provide a standard that is a plausible one, one that we have a reasonable basis to believe to be a better reflection, or description, of the agent's well-being than the agent's actual desires are.

But if the earlier arguments against the rationales for invoking an information condition were sound, then we have strong reasons to doubt the plausibility of the standard invoked by the defenders of the Knowledge-Modified position. I argued earlier that the desires allegedly formed or banished in a hypothetical desire situation of full information are of one of two kinds: either they are 'non-basic,' that is, instrumental or specificatory desires, or they are merely the causal result of the agent's beliefs and the particular features of the agent's psychological constitution. In the former case, since there really are no non-basic desires, the appearance that the Knowledge-Modified DF theory offers an independent standard for one's actual desires is a false one: no changes in desires occur as the result of the adjustment of specificatory or instrumental beliefs. In the latter case, the Knowledge-Modified view really does offer a distinct standard that can be used to criticize one's actual desires, but it is far from clear that it is a plausible one. Since the beliefs in question play a merely causal, and not a justificatory, role in the generation of these desires, we lack a basis for affirming that the desires that would be caused to exist in the presence of these beliefs are a reasonable standard by which to criticize actual desires.

Perhaps the defender of the Modified view will appeal only to the claim that, as a matter of fact, it appears that the results of adjusting desires by way of idealization of information matches up with our considered judgments about agents' well-being better than the Simple view does, and thus we are justified in holding this Modified view to be superior to the Simple view on account of the Modified view's critical capacities. If this were true, I suppose that it would provide at least a *prima facie* basis for moving to a Modified view—even if we would still want to know *why* exactly it turned out to be that the implications of a Knowledge-Modified view happen to cohere so well with our considered judg-

ments. But we lack evidence to believe that this is true. Once we move beyond the relatively facile cases of alleged changes in desires that occur as a result of improvement of instrumental or specificatory beliefs—cases that the Simple and Knowledge-Modified views handle equally well—I do not see that we have any evidence that the Knowledge-Modified view yields the right results with greater frequency than the Simple view does. (For a detailed discussion of the extent to which the Knowledge-Modified view generates counterintuitive implications, see Loeb, 1995.)

The defender of the Simple view can also point out that this view is not completely lacking in the capacity to serve as a kind of critical standard. We might put it this way. What seems to call most for criticism are the various passing, particular desires that grip us during the course of our everyday lives. But once we have distinguished between desires and occasions of being motivated toward particular states of affairs (e.g. between my desire for a baseball with Will Clark's signature on it and my being motivated to obtain this particular baseball), it seems likely that many events that have been loosely called 'passing desires' are, in fact, better characterized as occasions of being motivated toward some particular states of affairs, which motivation occurs in virtue of some belief-desire complex.<sup>21</sup> These occasions of being motivated toward some particular states of affairs are, however, subject to criticism on the Simple DF theory. My being motivated to get X can be criticized, on the Simple DF view, by way of the claim that my desires are such that in conjunction with true beliefs I would not be motivated to get X (e.g. my being motivated to obtain this baseball—what is misleadingly sometimes called my desire for this baseball—can be criticized in virtue of the fact that my desires are such that if I did not falsely believe it to be signed by Will Clark, I would not be motivated to obtain it.) What does lie beyond criticism in terms of well-being, on the Simple theory, is our deep, actual desires—whatever those happen to be. But once the appropriate distinctions are in place it is not as clear how powerful the objection is against the Simple view that it cannot subject the agent's desires to rational criticism in terms of well-being.<sup>22</sup>

It seems to me, then, that the distinction between desires and particular occasions of motivation blunts some of the force of Railton's objection; and even if that objection is granted force, it does not appear that the Knowledge-Modified view succeeds in providing a sufficiently plausible standard to constitute a rationale for the move from the Simple to the Knowledge-Modified version of DF theory. Let us turn, then, to Griffin's worry about the Simple view: that it seems unable to account for the fact that we can mistake our own interests. Now, the defender of the Simple view can note that his or her view allows at least one sense in which we can easily mistake our interests: even if each of us correctly believes that well-being is a matter of the satisfaction of actual desires, since what each of us desires is not transparent to us but is knowable only by a fallible process of investigation, it is possible for our false beliefs about what we want to translate into false beliefs about what our good consists in. Griffin is clearly more concerned, though, about a different sense in which one can mistake one's interests:

one can desire X, yet when X obtains one can feel that he or she is no better off. To some extent, the Simple DF theorist can respond to this objection by offering both a model of how such could occur within the constraints set by the Simple view and challenges to our intuitions about these cases. The Simple DF view surely allows that one might have a desire satisfied yet end up no better or even worse off: such could occur if the obtaining of an object of desire precludes the satisfaction of some other desire. In other cases, the Simple DF theorist will be inclined to challenge the intuitions that ground the objection. In some cases, the challenge will concern the judgment that the agent's desires have really been satisfied: a more careful analysis of the agent's motivational structure, an analysis that will include employment of the appropriate principle for individuation of desires, might yield the result that the agent's desires were not in fact satisfied.<sup>23</sup> In some cases, the challenge will concern the judgment that the agent is no better off by having the desire satisfied: even if the aim of an account of well-being is to capture the ordinary notion of an agent's good, we should not be surprised if our account of well-being calls into question some of our intuitions about the extent to which an agent is well-off. Surely those that press the view that Simple DF theories are insufficiently critical of our present particular desires would be in a strange position to hold that a theory of well-being should not be critical of our present particular intuitions about how well-off some agent is.

The mention of these strategies for dealing with the sorts of cases that Griffin has in mind does not show, of course, that these strategies will satisfactorily put to rest all such instances that might be urged against the Simple DF view. The Simple DF theorist might rest content with the response that until a writer sympathetic to Griffin's view brings forward a case that is not susceptible to these strategies, the status of the Simple DF theory, being the default version of DF theory, is not imperilled. But the Simple DF theorist can further point out that even if some of these cases did cause trouble for the Simple version of the theory, that would constitute a basis for preferring the Knowledge-Modified to the Simple view only if the Knowledge-Modified view were free from these troubles, or subject to them to a lesser extent. And, again, this is a claim that we have much reason to doubt. What generates Griffin's worry is not entirely a matter of the information available to the agent—if so, we would likely be able to handle the worry by employment of the same sorts of arguments used above against the other rationales offered for the Knowledge-Modified view—but rather a result of the sheer fact that the desires that Griffin has in mind are *prospective*. As Sumner has recently argued, the gap between our *ex ante* expectation and our *ex post* experience creates difficulties for DF theory even under conditions of fuller information: even if we have a clear vision of what we want and how things will go, the obtaining of the objects of our desires can feel disappointing and unrewarding (1996, 130ff). Since the worries raised by Griffin arise for both the Simple and the Knowledge-Modified view, they cannot constitute an adequate rationale for moving from a Simple to a Knowledge-Modified DF theory.

### A conditional conclusion

The Knowledge-Modified DF theory lacks grounds for affirming either of the disjuncts that distinguishes that view from the Simple DF theory. Nothing that I have said here, of course, is meant to imply that the Simple DF theory does not have its own difficulties. That theory seems to generate strange implications about well-being in cases where agents have weird or unintelligible desires (for, e.g., collecting plates of mud, avoiding touching brown boxes, counting blades of grass); that theory needs some principled rationale for excluding from the well-being of an agent the satisfaction of those desires that seem not to touch on the agent's life in any significant way (such as Parfit's Stranger on a Train case: 1984, 494);<sup>24</sup> and that theory has difficulty explaining to what extent the satisfaction of past, present, and future desires is implicated in an agent's good. But these are problems shared by the Knowledge-Modified DF theory: because Knowledge-Modified DF theories, like all genuine DF theories, do not discriminate among desires on the basis of their content, they must also imply that having a plate of mud, not touching brown boxes, or counting blades of grass could be ultimate constituents of an agent's well-being; merely as such, Knowledge-Modified theories say nothing about how desires regarding states of affairs seemingly distant from one's own good are to be dealt with; and while Knowledge-Modified views require fuller information for the desiring self, they must also provide an account of the extent to which the hypothetical desires of the past, present, and future selves are relevant to that person's well-being. My claim has not been that the Simple DF theory is without difficulties, but that those who affirm DF theory have no principled basis for the move from a Simple to a Modified version of the theory. Since the Simple and Knowledge-Modified views have these problems in common, mentioning these sorts of objections to the Simple DF theory does nothing to call that claim into question. If one is a desire-fulfillment theorist, one should be a Simple desire-fulfillment theorist.<sup>25</sup>

### Notes

<sup>1</sup>The DF theory defended in Railton, 1986b does not precisely fit this characterization; I will consider his variant of DF theory below.

<sup>2</sup>Throughout this paper I shall use the locutions 'an agent's good' and 'an agent's well-being' as interchangeable. Some authors have held that this equation is improper, that the concept of an agent's good extends more widely, and includes, the concept of an agent's well-being. See, for example, Railton, 1986a, 30. Nothing in my argument turns, though, on this implied equation.

<sup>3</sup>Of course, Dworkin offers a response to this objection on Rawls' behalf in this very article (1989, 46–53).

<sup>4</sup>It would imply the indefensibility of DF theory if, that is, their hypothetical desire situations incorporated a *full* information condition, which is the target of Sobel's and Rosati's criticisms. If a theory's information condition were more modest, perhaps it would escape those criticisms. But the impetus that leads one to give any information conditions—that is, to make a clear and intuitively plausible distinction between one's genuine good and one's apparent good—naturally pushes one toward a full information condition. As my argument will show, though, DF theorists should not be inclined to appeal to a hypothetical desire situation with *any* information condition, however modest.

<sup>5</sup>When I refer to an agent's 'actual' desires, 'actual' contrasts with 'hypothetical.' 'Actual' is not synonymous with 'present' and thus does not contrast with 'past' or 'future.' Therefore, as I have defined the theses of Simple DF theory, it is an open question whether the actual desires that the agent has had in the past or will have in the future are relevant to the agent's well-being in the same way that the agent's present actual desires are. Whatever the Simple DF theorist ultimately decides on this matter, that theorist affirms at least the strong negative claim that if an agent will never have an actual desire for some state of affairs X, then the obtaining of X will never contribute to that agent's well-being.

<sup>6</sup>For ease of exposition, the examples of information that I will use involve propositional knowledge. But the argument can also be formulated in terms of other sorts of information, such as knowledge by acquaintance.

<sup>7</sup>First baseman, Baltimore Orioles, lifetime batting average hovering around .300.

<sup>8</sup>A DF theorist might offer a different conception of desire, on which to have a desire is merely to have a pro-attitude toward some state of affairs rather than to be motivated to pursue it. Nothing essential to my argument assumes the latter conception of desire. If one inclines toward accepting the 'pro-attitude' conception of desire, one may make the following substitutions: in place of the idea that desires are ascribed to agents as explaining why the agents act as they do, say that desires are ascribed to agents as explaining why they approve of what they do; and in place of a principle of individuation for desires that appeals to motivational force, assume a principle of individuation that appeals to commendatory force. The persuasiveness of the argument, whatever it is, should be unaffected by these changes.

<sup>9</sup>It has been suggested to me that this principle for individuating desires fails because it is unable to distinguish between an agent's having a conjunctive desire that both X obtain and Y obtain and an agent's having a desire that X obtain and a desire that Y obtain. But this is not right. For in a case in which one is able to promote X's obtaining but not Y's, one will not be motivated to promote X's obtaining if one has a conjunctive desire while one will be motivated to promote X's obtaining if one has two distinct desires. If I have a desire to sing and dance, then if I am unable to sing, I will not be motivated to dance. But if I have distinct desires, one to sing, the other to dance, then my being unable to sing will not preclude my being motivated to dance. It seems to me that it is this difference in motivation that leads us to distinguish between having a conjunctive desire and having a number of distinct desires.

<sup>10</sup>Unless, of course, there is another desire that would happen to be satisfied by owning this ball—if, say, one also desired a ball autographed by Juan Gonzalez, and believed this ball to be autographed by Juan Gonzalez.

<sup>11</sup>It might be responded that this argument for the claim that I have no desire for this particular baseball might be turned around: it might be used to show that I have no desire for a baseball with 'WC' on it, but only a desire for this particular baseball. This would be correct, if the only considerations employed to ascribe desires to agents were those of individuation. But this is obviously not so. Two other considerations are also important: the capacity to ground the truth of counterfactual claims and simplicity. First, one reason to ascribe the general desire for a baseball with 'WC' on it is to explain the truth of claims like 'whenever I believe a baseball to be signed by Will Clark, I am motivated to possess that ball.' Of course, one could also explain the truth of this counterfactual by ascribing to me a potential infinity of discrete desires for particular baseballs believed by me to be signed by Will Clark. But no one ascribes desires in this way, for simplicity is also an important consideration in desire-ascription. So the fact that the argument as presented does not in itself decisively show that my desire is for a baseball with 'WC' on it does not give the defender of the Simple view any reason to be concerned about the capacity of that theory to determine which of these two putative desires is actual and therefore relevant to the constitution of my well-being.

<sup>12</sup>One might hold that there is a tension here in my asserting both (a) that the principle of individuation for desires that I offer is justified by the feature of our practice of ascribing desires in order to account for motivation to act and (b) that there are no instrumental desires: after all, it is also part of our practice of ascribing desires that we ascribe instrumental desires to agents! Two responses are to the point here. First, not all features of the practice of ascribing desires are on the same level, as it



were: the aim of the practice itself, to explain agents' actions, is the highest-order feature of the practice, which governs the others. I take it, then, that if it turns out that the ascription to agents of instrumental desires is unhelpful in explaining agents' motivations to act, then one has reason to revise that practice by refraining from ascribing those desires to agents. Secondly, it seems to me that our practice of ascribing desires is not entirely of one mind about the full propriety of ascribing specificatory or instrumental desires. For consider the following commonplace exchange. *Student A*: "What are you doing tonight?" *Student B*: "Studying." *Student A*: "Studying? But it's Friday night! Why do you want to study tonight?" *Student B*: "I *don't* want to study; I *want* to pass my philosophy exam." (I owe thanks to an anonymous reader for raising this objection.)

<sup>13</sup>Thomas Nagel discusses and argues against the views of those that do not believe in "motivational action at a distance" in his 1970.

<sup>14</sup>How then do we explain why an agent acts to promote one means to a desired end rather than another? Suppose that an agent desires X, believes Y to be a means to X, and believes Z to be a means to X; the agent promotes state of affairs Y, and not Z. This case can be dealt with in a number of ways. In some cases, there will be some other desire that Y is believed to satisfy while Z does not (perhaps a desire to employ the most efficient, or most aesthetically pleasing means). In some cases, one might form a distinct, basic desire to promote Y. In some cases, there will be no explanation for why the agent promoted Y rather than Z—at least no explanation that can be framed in terms of desires and beliefs.

<sup>15</sup>For a more detailed critical discussion of Brandt's views, see Velleman, 1988.

<sup>16</sup>For further discussion of these points see Hubin, 1996, 36–43, which raises them in the context of criticism of Knowledge-Modified DF theories of rationality rather than well-being.

<sup>17</sup>I owe thanks to a referee at *Noûs* who pointed out that I had ignored this way in which Railton's view differs from other DF theories.

<sup>18</sup>Railton mentions an alternative reason for moving to a full information view, that is, that the Simple view would be unable to supply an account of well-being that can serve as a critical standard for the agent's desires. I will consider this rationale below.

<sup>19</sup>One could claim, I suppose, that it is just obvious that second-order desires have this sort of authority; but to me this is not at all obvious. One might also appeal to intuition here, claiming that the view that second-order desires have authority over first-order ones enables us to explain certain intuitions that we have about the content of agent's well-being. I find this sort of appeal, if ungrounded in at least a minimal account of why second-order desires are authoritative over first-order ones, *ad hoc*. But one might also respond by noting that there are cases in which the second-order view seems clearly to produce sharply counterintuitive results. Consider the example offered by Parfit against the view that desires that one desires not to have are not relevant to one's well-being: "Suppose that you are in great pain. You have a very strong desire not to be in the state that you are in. On our revised theory, a desire does not count if you would prefer not to have the desire. This must apply to your intense desire not to be in the state you are in. You would prefer not to have this desire. If you did not dislike the state you are in, it would not be painful. Since our revised theory does not count desires that you would prefer not to have, it implies, absurdly, that it cannot be bad for you to be in great pain" (1984, 498).

<sup>20</sup>The version of the Simple view that he refers to is Hobbes'; but it is unclear whether Hobbes' view is so simple. For discussions of this issue, see Hampton, 1986, 38–40, and Murphy, forthcoming.

<sup>21</sup>Not all 'passing desires' are occasions of being motivated that are resultant upon a desire proper and some specificatory or instrumental belief. Some are genuine basic desires that are just quite transitory, and these may not be open to criticism on the Simple view. One might say that since a Simple view must allow these ephemeral desires to determine in some way the content of the agent's well-being, this constitutes an objection to that view. I doubt, though, both that it is objectionable to allow such desires to shape the content of an agent's good and that a Knowledge-Modified view would provide any way to eliminate such desires. First: why shouldn't the satisfaction of whims play a role in making an agent well off? And second: do we have any reason to think that improvements in information would have any effect on an agent's tendency to form ephemeral desires?

<sup>22</sup>Though we could subject them to criticism in terms of other kinds of value (moral, aesthetic, perfectionist, etc.), as Railton points out (1986b, 177).

<sup>23</sup>Is, then, the Simple view not so simple after all? If our desires are not apparent to us, and are knowable only by an eminently fallible process of investigation, do we not lose the attractiveness resulting from the seeming straightforwardness of the Simple position? Perhaps some of the initial attractiveness is lost, but we do need to keep in mind that this version of DF theory remains relatively simple. For, after all, every DF view will have to put forward, or assume, a view on the ascription and individuation of desires. So, even if the Simple DF theory appeals to an account of the individuation of desires that makes the ascription of desires a more complex matter, it remains simple in comparison to the Knowledge-Modified account. Given the thesis of this paper—that there is no rationale for moving from a Simple to a Knowledge-Modified view—this comparative merit is all the simplicity that the Simple view needs to claim for itself.

<sup>24</sup>Parfit imagines a situation in which he meets a stranger who has a disease that is believed to be fatal. Parfit forms a desire that the stranger be cured of the disease. Suppose that he never hears of the stranger again, yet the stranger in fact recovers from the disease. Parfit finds the idea that the satisfaction of this desire makes him better off highly implausible (1984, 494).

<sup>25</sup>I received a lot of help in writing this paper. I owe thanks to Richard Arneson, Terence Cuneo, Christoph Fehige, Trenton Merricks, Elijah Millgram, Madison Powers, Peter Railton, Henry Richardson, Connie Rosati, and Heath White, as well as to the members of the Georgetown / Maryland Moral Psychology Discussion Group and audiences at the Georgetown philosophy department colloquium and at the ISUS meeting in New Orleans. I also benefitted from the critical comments of two anonymous referees at *Noûs*. I am grateful to Georgetown University and to the Erasmus Institute for the fellowship support that I received from them during the drafting and revising of this paper.

## References

- Brandt, Richard. (1979) *A Theory of the Good and the Right* (Oxford: Clarendon Press).
- Dworkin, Ronald. (1989) "The Original Position," in *Reading Rawls*, ed. Norman Daniels (Stanford: Stanford University Press), 16–53.
- Griffin, James. (1986) *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford: Clarendon Press).
- Hampton, Jean. (1986) *Hobbes and the Social Contract Tradition* (New York: Cambridge University Press).
- Hubin, Donald C. (1996) "Hypothetical Motivation," *Noûs* 30 (1), 31–54.
- Loeb, Don. (1995) "Full-Information Theories of Individual Good," *Social Theory and Practice* 21 (1), 1–30.
- Murphy, Mark C. (forthcoming). "Hobbes on the Evil of Death," *Archiv für Geschichte der Philosophie*.
- Nagel, Thomas. (1970) *The Possibility of Altruism* (Princeton: Princeton University Press).
- Parfit, Derek. (1984) *Reasons and Persons* (New York: Oxford University Press).
- Quinn, Warren. (1993) "Putting Rationality in its Place," *Morality and Action* (New York: Cambridge University Press), 228–255.
- Railton, Peter. (1986a) "Facts and Values," *Philosophical Topics* 14 (2), 5–31.
- Railton, Peter. (1986b) "Moral Realism," *Philosophical Review* 95 (2), 163–207.
- Rawls, John. (1971) *A Theory of Justice* (Cambridge: Belknap Press of Harvard University Press).
- Rosati, Connie. (1995) "Persons, Perspectives, and Full Information Accounts of the Good," *Ethics* 105 (1), 296–325.
- Sandel, Michael. (1982) *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press).
- Sidgwick, Henry. (1981). *The Methods of Ethics*, 7th ed. (Indianapolis: Hackett).
- Sobel, David. (1994) "Full Information Accounts of Well-Being," *Ethics* 104 (4), 784–810.
- Sumner, L. W. (1996) *Welfare, Happiness, and Ethics* (Oxford: Clarendon Press).
- Velleman, J. David. (1988) "Brandt's Definition of 'Good'," *Philosophical Review* 97 (3), 353–371.