

Subjectivism without Desire[†]

Dale Dorsey

Department of Philosophy
University of Kansas
1445 Jayhawk Boulevard
Wescoe Hall, rm. 3090
Lawrence, KS 66045
ddorsey@ku.edu

NOTE: This is a penultimate draft of a paper that is currently forthcoming in *The Philosophical Review*.

For the purposes of this paper, subjectivism about well-being holds that ϕ is intrinsically good for x if and only if, and to the extent that, ϕ is *valued*, under the proper conditions, by x .¹ The natural contrast view, objectivism, holds that there are facts about any given person's well-being that are independent of that person's evaluative perspective. For subjectivism, unlike objectivism, a person's evaluative perspective, under the right conditions, *determines* that which is good for her, and how good it is for her.

Given this statement of the view, there is room for intramural dissent among subjectivists.² Most commonly, subjectivists have disagreed about what might be called the "theory of favored conditions": what sort of conditions must apply to x for x 's valuation of ϕ to render ϕ intrinsically good for x . For instance, some have held that ϕ is good for x if and only if x , under her *actual* conditions, values ϕ .³ Alternatively, some have held that ϕ is good for x if and only if x would value ϕ under certain idealized or counterfactual conditions, i.e., had x undergone cognitive psychotherapy, or

[†]I would like to thank Chris Heathwood, Doug Portmore, David Sobel, Connie Rosati, and anonymous reviewers for helpful comments.

¹This statement is a bit rough insofar as valuing can come in different flavors, not all of which are relevant to well-being. I hereby lay this slight complication aside until later (§3.1).

²Intramural disputes I will not discuss here include whether, and to what extent, goods should be time-relativized; whether all goods are goods at a time, or whether x 's valuing of ϕ at t entails that ϕ is good for x *simpliciter*, in an absolute fashion. For this reason, I will not time-relativise my account of subjectivism, but will leave this open for those who would prefer to do so. Other controversies in the formulation of subjectivism exist; I focus instead on the proper account of what it means "to value".

³Mark C. Murphy, "The Simple Desire-Fulfillment View" in *Noûs* 33 (1999).

were x fully informed, fully rational, fully experienced, etc.⁴

Though settling the proper theory of favored conditions is of the first importance, I elect to leave aside this discussion here. Rather, I focus on a second source of potential dispute: what might be called a “theory of valuing”. Subjectivists must offer an account of what it means for x to value ϕ under *any* conditions. Though there has been some disagreement,⁵ most hold that valuing is best characterized as—in some form or other—*desiring*. Most subjectivists hold that x values ϕ if and only if x desires ϕ . Call a subjectivist theory of welfare that also accepts a desiderative theory of valuing “desiderative subjectivism” (DS). In this paper, I to argue that subjectivists have erred in accepting a desiderative theory of valuing. Instead, I argue that subjectivism is well-advised to replace desire with *belief* or *judgment*.⁶ Instead of suggesting that x values ϕ to the extent that x desires ϕ , we should say that x values ϕ to the extent that x judges that ϕ is good for x . Call a subjectivist theory of welfare that also accepts a judgment-based theory of valuing “judgment subjectivism” (JS).

A brief roadmap. In the first and second sections, I argue that a desiderative theory of valuing is counterintuitive (§1) in a way that a suitably specified and refined judgment-based alternative can address (§2). In the third section, I argue that JS avoids a number of traditional objections that have plagued desiderative subjectivism. In the fourth section, I consider an important response on behalf of desiderative subjectivism, and in sections five through seven I address three important objections to JS in comparison to DS. Section eight concludes.

1. *Desire and Valuing*

In this paper, I do not seek to argue for subjectivism about well-being. However, to motivate an alternative to desiderative subjectivism, it is important to rehearse one of the primary advantages of subjectivism versus non-subjectivist alternatives. The spirit of subjectivism about well-being appears to be captured by the claim that there is an important constitutive connection between that which is good for a person and that person’s

⁴A number of “idealized” subjectivist views have been offered, including, most importantly, Henry Sidgwick, *The Methods of Ethics* (Indianapolis, IN: Hackett Publishing Company, 7th ed., 1981 [1907]), 111; Richard Brandt, *A Theory of the Good and the Right* (Oxford: Oxford University Press, 1978), 113, along with many others.

⁵Cf. Wayne Sumner, *Welfare, Happiness, and Ethics* (Oxford: Oxford University Press, 1996).

⁶For the purposes of this paper, I use “belief” and “judgment” interchangeably.

evaluative perspective. (For the remainder, I will treat “good for” and “intrinsically good for” as interchangeable.) Subjectivism is able to explain the seemingly plausible connection between what a person values for her own sake and what is *valuable for her* for its own sake. As noted by Richard Arneson, subjectivism is characterized, and in part motivated, by the plausible thought that a person should be *sovereign* over her good—her evaluative perspective (at least under the right conditions) should *determine* her well-being.⁷ That subjectivism links a person’s well-being with that person’s values (*modulo* a theory of favored conditions) is a source of its intuitive strength. If so, any viable subjectivist theory of welfare must adequately accommodate this connection. To do so, however, such a theory must answer the question: what does it mean for x to value ϕ ?

1.1 First-order Desires

The most natural candidate for a person’s evaluative perspective would appear to be a person’s *desires*: for me to value ϕ is for me to *desire* ϕ . This view is reflected in classic desire-satisfaction theories of well-being, from Hobbes,⁸ to Sidgwick,⁹ Rawls,¹⁰ Brandt,¹¹ and many others. However, that valuing ϕ is distinct from desiring ϕ appears to be approaching the status of philosophical dogma. Cleavage between value and desire is standardly motivated by appeal to addicts, obsessives, and various others that seem “alienated” from their desires.¹² David Lewis, for instance, writes:

[W]e’d better not say that valuing something is just the same as desiring it. That may do for some of us: those who manage, by strength of will or by good luck, to desire exactly as they desire to desire. But not all of us are so fortunate. The thoughtful addict may desire his euphoric daze, but not value it. Even apart from all the costs and risks, he may hate himself for desiring something he values not at all. It is a desire he wants very much to be rid

⁷Richard Arneson, “Human Flourishing versus Desire Satisfaction” in *Social Philosophy and Policy* 16 (1999), 116.

⁸Cf. *Leviathan*, Part I, chapter 11.

⁹Sidgwick, 111-15.

¹⁰John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971) §64.

¹¹Brandt, 132.

¹²See Harry Frankfurt, “Freedom of the Will and the Concept of a Person” in *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1981), 11-25.

of.¹³

Examples of this type abound, and thus it seems plausible to say that desire is an inappropriate foundation for a person's evaluative perspective. It would seem that we can desire any number of things without valuing them. Furthermore, it would seem that we can value any number of things without desiring them. Consider, for instance:

Stan: Stan is a successful lawyer practicing in Big City, who has an opportunity to move to Small Town, which is close to his family and friends. Stan has consistently extolled the virtues of small-town living, and has regarded Big City as a barely tolerable annoyance. But suppose that as a product of exhausting work at his law office, Stan simply comes to lack a desire for either option. When asked whether he desires to remain in Big City or move to Small Town, he replies that he has no real desires either way. When asked which option he wants, he replies, exhausted, numb: "I dunno. I guess I don't really want either."

Stan clearly lacks a desire either to move to Small Town or to remain in Big City. But a desiderative characterization of valuing would seem to entail that Stan values neither option. But this inference is too quick. We know, of course, that Stan fails to desire either option. He has, after all, a serious case of "desiderative exhaustion", or, in more common parlance, *burn-out*. But in such cases we do not ordinarily conclude that the subject in question fails to value one or the other options because he is burnt-out. Indeed, given Stan's prior pronouncements we may very well conclude that Stan values the move to Small Town, though his burn-out prevents him from desiring it.

1.2. *Second-order Desires*

So far, it appears that to value ϕ is not the same as to desire ϕ . However, this by itself does not spell doom for desiderative subjectivism. Indeed, one can reject a "first-order" desiderative approach to valuing and replace it with an appeal to *second-order* desires. On this view, x values ϕ not if x desires ϕ , but rather if x *desires to desire* ϕ . For instance, one might think that though the drug addict desires the drug's effects, he doesn't *desire to desire them*, and hence does not genuinely *value* the euphoric daze. As Lewis writes: "We conclude that he does not value what he desires, but

¹³David Lewis, "Dispositional Theories of Value" in *Papers in Ethics and Social Philosophy* (Cambridge: Cambridge University Press, 2000), 70.

rather he values what he desires to desire.”¹⁴ A similar account is offered by Railton: “Let us then say that an individual’s *intrinsic* good consists in attainment of what he would in idealized circumstances want to want for its own sake. . . were he to assume the place of his actual self.”¹⁵ For Railton and Lewis, values are best captured by what one desires to desire, or “wants to want”.

Problems arise, however. First, as Lewis mentions, just as one can be alienated from one’s first-order desires (as in the case of the drug addict), one can be alienated from one’s second-order desires.¹⁶ For instance, assume that I desire to eat an ice cream cone, and furthermore desire to desire to eat it. But assume now that I fail to desire to desire to desire to eat the ice cream cone, and in fact have a third-order *aversion* to the ice cream cone. What principled reason could there be for selecting second-order desires as the proper evaluative perspective rather than, say, third-order desires, or some further order of desire?

Second, the second-order desiderative approach doesn’t adequately respond to *Stan*. One is tempted to point out that there is no reason to believe that Stan has any “second-order” desires, any more than he has any “first-order” desires, when it comes to the choice between Small Town and Big City. After all, he is exhausted, burnt-out. It doesn’t seem impossible, or even implausible, to describe his mental states as simply failing to want to want either option. Again, whether he does would appear to depend on the depth of his burn-out. But lack of a second-order desire doesn’t seem enough for us to declare that he doesn’t *value* moving to Small Town. We would be more likely to say, of Stan, that he fails to to desire to desire what he really, after all, values.

I find the above worries significant, but not all will agree. Hence, I propose to assume that they can be solved (or at least sidestepped). However, there is a third problem that I believe is decisive.¹⁷ The second-order desiderative approach seems at first glance both over-inclusive and under-inclusive. Two cases shed light. First, consider:

Julie: Julie is stuck in a marriage to a man for whom she feels no love. She desperately desires to leave her marriage and be-

¹⁴Lewis, 70-71.

¹⁵Peter Railton, “Facts and Values” in *Facts, Values, and Norms* (Cambridge: Cambridge University Press, 2003), 54-5.

¹⁶Lewis, 71.

¹⁷The following argument is made in more detail in Dale Dorsey, “Preferences, Welfare, and the Status-Quo Bias” in *Australasian Journal of Philosophy* 88 (2010). See also David Sobel, “Full Information Theories of Well-Being” in *Ethics* 104 (1994), 793n19.

gin anew. Unfortunately, given her financial and social circumstances, she is forced to remain married. However, because she is forced to remain married, she strongly *wants to want* to remain married. Indeed, she has made attempts, in the form of psychological therapy, to get herself to desire to remain with her husband, but to no avail.

Here it would appear that Julie desires to leave her marriage, but because of circumstances beyond her control, she cannot. Given that this is the case, Julie might, quite rationally (even from the standpoint of ideal circumstances), come to *desire to desire* to remain in her marriage. Indeed, were she able to alter her desires in this way, she would surely live a better life; less frustration and regret. But it is surely a mistake to say that Julie values remaining in her marriage. In fact, she disvalues it in the extreme.

Take a further case:

Jack: Jack maintains the goal of being a great architect. He has worked his life to become one, has spent many years in training and apprenticeship, and yet has no talent whatever for architecture. Indeed, his pursuit of a career in architecture has caused more harm than good; more-or-less constant feelings of failure, ruined relationships with friends and loved ones, etc. Jack desperately wants to be a great architect. But he does not *want to want* to be a great architect: Jack is sufficiently self-aware to realize that this desire causes him nothing but frustration and regret.

Jack does not desire to desire to be a great architect. But it would be quite wrong to say that Jack does not *value* being a great architect. In fact, he values it highly. Second-order desires thus appear to poorly capture a genuine evaluative perspective. Jack may fail to value the state of desiring to be a great architect. But he surely values *being* a great architect.

A proponent of a second-order desiderative theory of valuing might claim that I have so far interpreted this view uncharitably. I may desire to desire ϕ ; but I may do so, for instance, simply because my second-order desire is derived from a more fundamental desire for something else, ψ , to which a desire for ϕ is a means. This does not entail that I value ϕ . I value (in the sense that is relevant to determining what is good for me) only what I desire to desire *for its own sake*. Might this proposal solve some of the problems noted here?

Perhaps, but disambiguation is in order.¹⁸ The first reading holds that I value ϕ to the extent that I desire the state in which I desire, for its own sake, ϕ . In other words, I desire the state of affairs in which I maintain an intrinsic desire for ϕ . The second reading holds that I value ϕ to the extent that I desire, for its own sake, the state in which I desire ϕ . In other words, I value ϕ to the extent that I possess an intrinsic desire for the state in which I desire ϕ . A third interpretation runs as follows: I value ϕ to the extent that I desire, for the sake of ϕ , the state in which I desire ϕ . Here I desire the state of affairs in which I desire ϕ not for the sake of that very state of affairs, but rather for the sake *of* ϕ itself.¹⁹

The first reading is insufficient to solve the cases at hand. Indeed, it might be plausible to say, right now, that Julie *desires* the marriage as an instrument, i.e., as a means to assist in the raising of her children, as a means to her own financial security. But that which causes Julie the most pain is the fact that she is averse to the marriage *for its own sake*. So it would be rational for Julie to want to want the marriage *for its own sake*: only in so doing could she avoid the frustration and regret that comes along with desiring her marriage only instrumentally or derivatively. Furthermore, these considerations appear to have little to offer in response to the case of Jack. Jack does not want to want to be an architect either for its own sake or as an instrument. But this does not mean he fails to value being an architect for its own sake.

What about the second reading? Because Julie does not desire *intrinsically* to desire to remain in her marriage, this view does not imply that Julie values remaining in her marriage. However, though this proposal may solve the problem of over-inclusiveness, it worsens the contrary problem. First, Jack certainly does not desire, for its own sake, the state in which he desires to be an architect, and so this approach could not accommodate the sensible claim that Jack values being a great architect. But perhaps even more importantly, surely it is not the case that to value ϕ one has to desire *for its own sake* the state in which one desires ϕ . Indeed, to desire a desiderative state for its own sake is surely rare: when I desire to desire to, say, lose weight, I do so because maintaining the state of desiring to lose weight is instrumentally beneficial: it is instrumental to my goal of losing weight. In

¹⁸Thanks to anonymous reviewers.

¹⁹Some technical lingo might help. Read “ $D_n\phi$ ” as “I have an n -order desire for ϕ ”. Read “ $S_n\phi$ ” as “my n -order desire is for the sake of ϕ ”. On the first reading, I value ϕ to the extent that I possess the following state: $D_2(D_1\phi S_1\phi)$; the second reading: $D_2(D_1\phi)S_2(D_1\phi)$; and the third reading: $D_2(D_1\phi)S_2\phi$. These are not the only readings, but seem the most promising.

other words, and more pointedly, I can *value* losing weight without *intrinsically* desiring the state in which I desire to lose weight. To insist that, to value ϕ , one must desire *for its own sake* the state in which one desires ϕ seems wildly under-inclusive.

Finally, consider the suggestion that I value ϕ to the extent that I desire, *for the sake of* ϕ , the state in which I desire ϕ . Like the second reading, this proposal would have it that Julie does not value her marriage. Julie does not desire, for the sake of her marriage, to desire to remain in her marriage. Rather, she desires to remain in her marriage for other reasons, viz., that without remaining in her marriage she would be left shunned and destitute. And though this proposal can accommodate this verdict without the pitfalls of the previous reading, it remains unable to adequately address the cases at hand. Julie's evaluative stance toward her marriage is under-described in saying, simply, that she does not value it. Merely saying this would be compatible with pure indifference. But Julie is far from indifferent: she *disvalues* her marriage. But the view in question is unable to accommodate this: she certainly does not desire to leave her marriage, whether for the sake of leaving her marriage or for any other reason. Nor is she at all averse to the state in which she desires to remain in her marriage. Now take Jack. Jack lacks any second-order desire to be an architect. And though he is not *averse* to the state in which he desires to be a great architect for the sake of being a great architect (he is, rather, averse to that state for instrumental reasons), he possesses no second-order desire to *be* a great architect. Hence to say that Jack *values* being a great architect (which he surely does) requires us to move beyond his second-order desires. Of course, one might propose that Jack values being a great architect because he desires (first-order) to be a great architect. But this proposal fails for already-rehearsed reasons.

The proposal to capture a person's genuine values by means of her desires (whether first- or second-order) is intuitive, but problematic. For a first-order view, to desire is not necessary to value, and to value is not necessarily to desire. Even in its most successful incarnation, a second-order view is unable to correctly identify the evaluative profiles of Jack or Julie. While there are perhaps additional epicycles that may be added to a desiderative theory of valuing, the problems noted so far seem to provide sufficient reason to explore a non-desiderative alternative.

2. Judgment and Valuing

In light of the problems with a desiderative theory of valuing, the proposal I

would like to consider in more detail is that a person values that which she *believes is good for her*.

Before comparing a judgment-based view to its desiderative cousin, however, it is important to say more about the construction of a judgment-based view. First, the view I accept construes x 's genuine values not simply as beliefs of the form " ϕ is good for x ", but also as beliefs of the form " ϕ is better than ψ for x ", and various other comparative evaluative beliefs. In this way, a judgment-based theory of valuing is able to offer a profile of a person's preferences or subjective rank-ordering, as is surely necessary for any theory of valuing appropriate for a subjectivist theory of welfare. (For the sake of brevity, I treat beliefs about "good for me" as a helpful catch-all.)

Second, and most importantly, it is implausible to say that a person *values*, simply, that which she believes is good for her. Not all of my beliefs about that which is good for me express my genuine values. For instance, it could be that some of my evaluative beliefs generate absurdities: I might believe in the value of ϕ , but also believe something that straightforwardly implies the disvalue of ϕ . Imagine, for instance, that I believe that being a philosopher is good for me, for its own sake. Imagine also, however, that I believe that reading, writing, and teaching philosophy are intrinsically bad for me. In this case, it would seem that I have a belief in the value of being a philosopher, the value of which straightforwardly implies the value of the constitutive activities of being a philosopher, but that I do not believe that these constitutive activities are good for me. In this case, a simple judgment-based theory of valuing results in an absurd conclusion: I both value being a philosopher and do not value being a philosopher. Not to put too fine a point on it, any judgment-based theory of valuing that implies this conclusion fails *via reductio*.

The solution to this problem, however, is trivial. It seems plausible to say that Stan's values are given not by whatever Stan's evaluative beliefs happen to be, but rather by what Stan's set of beliefs about what is good for Stan would be were these beliefs *rendered coherent*. This proposal requires some elucidation; I'll mention three points. First, though a thorough definition of coherence is beyond the scope of this paper, I take this constraint to be rather weak. On my view, beliefs are incoherent to the extent that they are straightforwardly inconsistent, or would be inconsistent on the assumption of relevant true *non-evaluative* sentences. The latter point is important: in judging the coherence of evaluative beliefs, one cannot treat them in isolation: "eating Julia Child's recipes is intrinsically good for me" and "eating French food is intrinsically bad for me" are perfectly coherent if we ignore the non-evaluative fact that Julia Child's recipes are thoroughly

Gallic. Though my beliefs in the value of being a philosopher and the disvalue of reading, writing, and teaching philosophy are not, on their face, incoherent, they are incoherent if we accept a relatively straightforward claim about what it means to be a philosopher, i.e., to read, write, and teach philosophy over a period of time.

Second, what does it mean for an incoherent set of beliefs to be rendered coherent? For my purposes here, revision toward coherence is performed in accordance with the standard rubric of “minimal mutilation”.²⁰ When beliefs are incoherent, priority is granted to those that are stronger, and those that are supported, or warranted, by other evaluative beliefs the person maintains. This rubric is plausible. Imagine, in the case described above, that I offer a claim about the strength of my inconsistent beliefs, viz., that I believe *more strongly* in the value of being a philosopher than in the disvalue of the constitutive activities. In that case, we would surely declare that I value being a philosopher. The contrary beliefs do not express my genuine values. Furthermore, one might imagine that (leaving aside relative strength), a belief in the disvalue of being a philosopher better comports with my other beliefs about that which is good for me: I believe that being an academic is intrinsically bad, interacting with students is intrinsically bad, being concerned with academic and intellectual matters rather than practical political matters is intrinsically bad, etc. In this case, given my *other* beliefs, we would surely be tempted to declare that I do not value being a philosopher. (A number of questions could be asked about these principles that I will not pretend to answer in any serious way here. One might account for a belief’s “strength” in a number of different ways. For instance, one might claim that the strength of a belief is identified by the “credence” of the proposition believed. Alternatively, one could judge the strength of a belief by its “centrality”, i.e., its influence on the whole of one’s belief system. One could combine these principles. Further, one could ask questions about what it means for a group of sentences to support or warrant another. Though much can and should be said on these topics, I rely on the general concept here, to await a further specified conception.)

Third, one might object that we should not eliminate the possibility of tension in a person’s valuations. For instance, an individual may value maintaining a certain sort of scientific rationality, but may also value certain religious beliefs.²¹ If so, sometimes a person may value things that are not

²⁰See W. V. Quine, “Two Dogmas of Empiricism” in *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 2nd ed., 1981).

²¹Thanks to an anonymous reviewer.

mutually supportive, or are not mutually satisfiable. But my view does not eliminate the possibility of values that are in tension, or are jointly unsatisfiable. Recall that the coherence condition is meant to be weak, viz., to rule out beliefs that are self-refuting or that result in patent absurdities, i.e., that x both values and does not value ϕ . But the valuation of scientific rationality is not incoherent with the valuation of one's religion (in the suitably weak sense of coherence I mean here), though it might be that the *achievement* of one is inconsistent with the *achievement* of the other.

Given everything so far, a judgment-based subjectivism holds that ϕ values x if and only if, and to the extent that, x would believe that ϕ is good for x were x 's beliefs about that which is good for x rendered coherent, where coherence is given a weak interpretation as specified here.²²

JS, as stated, can adequately respond to the cases that caused problems for first- and second-order desiderative theories of valuing. Stan, for instance, may be seriously burnt-out, but it seems right to say that in the midst of his "desiderative exhaustion" he can still maintain a judgment that one or more of his options is valuable or good in itself for him or that one alternative is better than the other. Indeed, given his past statements on the matter, it seems plausible to believe that though he fails to desire to move to Small Town, he still believes that moving to Small Town would be good for him. Treating a person's evaluative beliefs as her evaluative perspective is compatible with the intuition that a mere case of burn-out does not entail that Stan takes no evaluative stance toward moving to Small Town or remaining in Big City.

Furthermore, take Lewis' drug addict. It would appear that if anything is true of the drug addict, he surely believes that taking his drug, or experiencing his euphoric daze, is not good for him. Going further, though Julie may desire to remain in her marriage, she surely does not *believe* that her well-being is served by so remaining. A similar thought appears plausible in Jack's case. Jack's lack of talent leads him to fail to want to be a great architect. But surely his lack of talent would not, ordinarily, alter his *belief* that it is good for him. Jack's case is an extension of a sensible point made by Sidgwick: "a prudent man is accustomed to suppress, with more or less success, desires for what he regards as out of his power to

²²I should note that when it comes to a *theory of favored conditions*, there is reason to specify, say, a more robust coherence constraint, in addition to adding something like a "completeness" constraint, i.e., the requirement that x 's set of evaluative beliefs be rendered complete enough to issue a verdict about the value of any potential welfare goods. But insofar as a theory of favored conditions is not at issue here, I discuss only the weak coherence constraint appropriate to a judgment-based theory of valuing.

attain by voluntary action—as fine weather, perfect health, great wealth or fame, etc.; but any success he may have in diminishing the actual intensity of such desires has no effect in leading him to judge the objects desired less ‘good’.”²³ In this case, Jack wishes to diminish his own desire to be a great architect, insofar as this very desire is making his life miserable. But this instance of strategic preference engineering needn’t alter Jack’s judgments of his own good. Though he desires *not* to desire to be a great architect (and, perhaps, even desires not to *believe* that being a great architect is a welfare benefit) he surely believes that the state in which he is a great architect would improve his life.

But what if conditions are different? Imagine now that the drug addict, though he does not want to want the euphoric daze, *believes* that his euphoric daze is good for him, for its own sake. This case might be difficult to imagine in the abstract in part because we often think of second-order desires as *responsive* to judgments about value. But one might imagine that this second-order aversion to the drug might develop, for instance, if he comes to realize that a desire for the daze is somehow getting in the way of other things he desires. In this case, it seems to me plausible to say that the addict values the drug; his second-order desires are simply a product of a rational preference not to lose other things he desires. A similar conclusion holds for Julie. Though she may desire to leave the marriage, if she *judges* or *believes* that the marriage is good for her, her first-order desire (*viz.*, to leave her marriage) would appear much more like the first-order desire of the drug addict as originally presented by Lewis; a desire that is simply alienated from that which she genuinely values.

Though I discuss additional cases in §5, there seems to me good reason to believe that a judgment-based theory of valuing can adequately solve the problems that confront its desiderative cousin.

3. *Two Structural Advantages*

In this section, I argue that a judgment-based subjectivism about welfare (JS) can solve two very serious and persistent structural problems for desiderative subjectivism. First, I argue that replacing desire with evaluative belief can solve the classic problem of distinguishing prudential and non-prudential valuing attitudes. Second, I argue that combining a judgment-based theory of valuing with a subjectivist theory of welfare can artfully, and quite naturally, avoid the so-called “paradox of desire”.

²³Sidgwick, 111.

3.1. Prudential and Non-prudential Valuing

Subjectivism about welfare faces the following problem. Intuitively, there are at least two ways I can value something. This can be seen by means of a classic example. Imagine that on an otherwise unremarkable subway ride, I happen to spy a overburdened mother, call her Beatrice, with a broken leg. I form a desire that Beatrice’s leg heal. This is a genuine desire: if I could do something to further her leg’s healing, I would. Furthermore—and without begging the question in favor of DS or JS—I could certainly be said to “value” the state in which Beatrice’s leg heals. Suppose now, three months after the train ride, her leg is completely healed. But even if we are subjectivists about well-being, we would not be tempted to conclude that I’m better-off. Though I value the healing of Beatrice’s leg, I don’t value it in a—for lack of a better term—“prudential” way.²⁴ Indeed, this phenomenon seems pervasive: I value the state in which New Orleans is rebuilt.²⁵ I value the state in which a cure for SIDS is found. I value the state in which the planet is inhabitable for humans centuries down the line. But I also value, e.g., becoming a philosopher or having a good marriage; and I value the latter class in a different way than the former. Thus to properly come up an acceptable theory of valuing that could form the basis of a theory of well-being, a subjectivist must properly identify only those valuing attitudes that are specifically *prudential* (i.e., self-interested) rather than *non-prudential* (e.g., altruistic).²⁶ The problem is that desiderative views seem unable to draw such a distinction plausibly.

Perhaps the most promising option is to distinguish prudential and non-prudential desiring on the basis of the desire’s *object*. For instance, Parfit suggests that desiderative views should declare that only desires “about my own life” are examples of self-interested, or prudential, desires.²⁷ But Parfit himself is less than clear about what this is supposed to mean. Which desires are about my own life? Mark Overvold has claimed that x ’s prudential desires are those the objects of which depend, necessarily, on x ’s existence.²⁸

²⁴There are many examples of this kind; mine is slightly adapted from Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), app. I. See also David Brink, “The Significance of Desire” in *Oxford Studies in Metaethics*, v. 3, ed. Shafer-Landau (Oxford: Oxford University Press, 2008), 25.

²⁵Thanks to Luke Robinson for this example.

²⁶Cf. Brandt, 328-329.

²⁷Parfit, 494.

²⁸Mark Overvold, “Self-Interest and Getting What You Want” in *The Limits of Utilitarianism*, ed. Miller and Williams (Minneapolis: University of Minnesota Press, 1982).

But there are three problems here. First, Overvold's proposal would have the effect of ruling out all desires that could only be satisfied posthumously as being prudential. To be fair, this is precisely what Overvold wishes to rule out. But it is not clear to me that we should. If I desire to be recognized posthumously as a master of the Hammond Organ, and work tirelessly to achieve this goal, it is unclear why we shouldn't regard this as an instance of prudential desiring.²⁹ Second, imagine that I desire that New Orleans be rebuilt *during my lifetime*. To my ears, anyway, this desire does not seem obviously prudential simply for the fact that its object necessarily involves my existence. Furthermore, David Sobel notes that my desires to keep my promises necessarily depend on my existence.³⁰ But this certainly doesn't guarantee that these desires are instances of prudential rather than non-prudential valuing. Third, and more generally, to distinguish prudential and non-prudential desiring on the basis of the object of desire cannot succeed, insofar as many desired objects seem to permit of both prudential and non-prudential valuing attitudes. Here's an example. Imagine that I desire to spend my life traveling to the poorest villages of the world to help vaccinate young children against common maladies. Is this a prudential or non-prudential desire? Surely it depends on the person in question: for some this might be a life-long goal upon which the success of life depends. For others this might be regarded as an instance of self-sacrifice one desires to undertake for beneficent or moral reasons. Hence the desire's object cannot seem to distinguish between prudentially relevant valuing and its contrary.

Alternatively, one might try to divide prudential and non-prudential desires abstracting from desired objects. For instance, one might say that prudential desires are those the objects of which I want *for my own sake*. But this proposal does not succeed. I can want things for my sake in a moral rather than prudential way, i.e., for the sake of my moral uprightness.³¹ A partisan of DS might respond by saying that only my *prudential* desires count in favor of my welfare. But this is obviously unsatisfactory: we have been asking which desires are prudential and which aren't. It's unhelpful to simply be told: the prudential ones. In surveying the problem of prudential

²⁹Cf. Brandt, 330.

³⁰David Sobel, "Well-Being as the Object of Moral Consideration" in *Economics and Philosophy* 14 (1998), 267.

³¹See Sobel, "Well-Being as the Object of Moral Consideration", 266-269. One could refine this by suggesting that my prudential desires are those that I maintain *for the sake of my welfare*. But, leaving aside the problem that desiring something for the sake of one's own welfare seems a curious pro-attitude to possess, this seems no more helpful than saying that prudential desires just are those that are prudential. See Arneson, 124.

desires, Arneson has declared, bluntly, that no solution can be found.³²

I do not wish to canvass all attempts to properly cleave prudential and non-prudential desires. This problem has been the object of much hand-wringing; Arneson's blunt declaration may yet prove too strong. But it is worth noting that JS can offer a particularly natural solution to this problem. For JS, the prudential valuing attitude is not a belief that the valued object is merely *good*. Rather, it is a belief that the valued object is good *for me*. Though I may believe that many things are good—such as a cure for SIDS, or the rebuilding of New Orleans—the “for me” clause is an essential element in any prudential valuing attitude: my welfare is determined by my beliefs about *my own good*. Here prudential valuing is clear as day: I take a prudential valuing attitude to ϕ only insofar as I believe ϕ is good *for me*. This seems to do the trick. Though I might believe that the healing of Beatrice's leg, or my own self-sacrifice for moral reasons, is *good*, in all but the most rare cases I do not believe these would be good *for me*.

Of course, I might for some reason come to believe (coherently, say, and under the right conditions) that Beatrice's leg's healing is *good for me*. For JS this fact would entail that I would be better-off were it to occur. One might, of course, object to this, perhaps plausibly. But this objection does not show that JS cannot properly distinguish between prudential and non-prudential valuing. It merely shows—if successful—that some objects toward which one might take a prudential valuing attitude do not count in favor of one's well-being. But this objection does not threaten JS rather than DS. It is instead a problem for subjectivism on the whole; perhaps it is reason to supplement subjectivism with a quasi-objectivist account of just what sorts of prudentially valued objects can count in favor of my welfare. If, on the other hand, we accept subjectivism as expressed in the introductory sentence of this paper, we should not bar an individual's well-being from extending beyond the boundaries of their life or experience, as mine does insofar as I prudentially value the healing of Beatrice's leg. Instead—and herein lies the real challenge—subjectivists must develop a theory of valuing that properly represents a person's prudentially relevant valuations no matter what they prudentially value. And JS completes this task in an admirably clear way. JS can distinguish the times at which x values the healing of Beatrice's leg in a prudential way (rare) versus the times at which x values the healing of Beatrice's leg in a non-prudential way (more common). Though I do not wish to argue that it is impossible for DS to mark this distinction (though offhand I can't see how it would do so), the relative

³²Arneson, 125.

ease with which JS can do so is an important advantage.

3.2. *The Paradox of Desire*

The following is a well-known problem for DS. Fortunately, JS has an easy solution. The problem runs like this. Imagine that I desire to be worse-off. In other words, perhaps as a result of a lack of self-worth or as a result of indoctrination that results in my own self-loathing, I come to desire strongly that I have a bad life. Assume that I am at a negative welfare level, and that because of the importance and intensity of this desire, having it satisfied would entail that I have a non-negative welfare level.³³ In that case, it would appear that if my welfare level is negative, it is positive (given the satisfaction of the desire that it be negative), and if it is positive, it is negative (given the strength of the desire that my welfare be negative). Hence, a paradox arises.³⁴

JS can avoid this problem. For JS, it never improves the quality of my life to be worse-off. Why? Because no matter what else I believe, *I cannot coherently believe that it is good for me to be worse-off*. Assume that the realization of some state of affairs ϕ would make my life worse. According to JS, I cannot believe that ϕ would make my life better, because for ϕ to make me worse-off, I *must* believe that it will make me worse-off, rather than better-off. Because any set of coherent evaluative beliefs will rule out the inconsistent set “ ϕ is intrinsically better for me” and “ ϕ is intrinsically worse for me”, JS does not allow that my welfare can be increased by its decrease. To believe that my life goes better as it goes worse is simply self-defeating.

This holds even if the state of affairs ϕ just is the state of affairs of my being *worse-off*. Assume that ϕ is the state of affairs in which I am worse-off than the current state of affairs ψ by the smallest possible degree. To believe that I am worse-off in ϕ , I must believe that ψ is better than ϕ . To believe that being worse-off improves my well-being, I must believe that my well-being improves by moving from ψ to ϕ , because I am worse-off in ϕ than I am in ψ . But I cannot coherently believe that I am better-off while ψ obtains than I am while ϕ obtains (which is required given the essential claim that ϕ is a state of affairs in which I am worse-off), and also believe that I am better-off (or neither better- nor worse-off) while ϕ obtains than

³³For a nice discussion of this objection, and one proposed solution to the problem, see Brad Skow, “Preferentism and the Paradox of Desire”, *Journal of Ethics and Social Philosophy* 3 (2009).

³⁴See also Chris Heathwood, “The Problem of Defective Desires” in the *Australasian Journal of Philosophy* 83 (2005).

I am while ψ obtains (which would be required were I to believe that ϕ has intrinsic prudential value simply for being worse than ψ , given that ϕ is worse than ψ in the smallest possible degree). To believe that I am better-off by being worse-off, I must believe that ϕ is ranked of *at least* equal value to ψ —otherwise my being worse-off would not make me better-off. But this is incoherent. If I am worse-off in ϕ , given JS, I must *believe* I am worse-off in ϕ .³⁵ But I cannot believe this and also believe that ϕ and ψ are of at least equivalent value. This result is just the sort of absurdity that the weak coherence condition rules out.

This result is significant. JS cannot allow that the mere fact that someone is worse-off of itself increases their well-being. Any set of beliefs that would generate such a paradox is incoherent. Though I wish to allow the possibility that a desiderative subjectivism could respond to this worry, I maintain that the response offered by JS is precisely what the doctor ordered. JS gets to the heart of the matter: one cannot have a better well-being score simply for having a worse well-being score.³⁶

One possible response suggests itself. JS appears to solve the paradox

³⁵Or, at least, I must maintain a set of beliefs that imply this, i.e., by evaluating the constituents of ϕ and ψ .

³⁶An anonymous suggests that though it is incoherent to believe, literally, that one is better-off for being worse-off, this doesn't entail that there's no way for JS to end up in paradox. For instance, I might not believe that JS is the true theory of welfare. If so, I might come to believe that it is better for me to have fewer, rather than more, of my evaluative beliefs satisfied. Surely this is coherent. But for JS, having fewer of my evaluative beliefs satisfied entails that I'm worse-off. And so I appear to maintain a set of coherent beliefs that entail, on JS, that I am better-off for being worse-off. Paradox results. But this argument goes wrong: the state of affairs (ϕ) in which fewer rather than more of my beliefs are satisfied needn't be a state in which I am worse-off. Depending on the evaluative significance I coherently grant to having fewer beliefs satisfied, I could be worse-off, better-off, or neither in ϕ . And if this is right, to believe that ϕ is better for me than ψ (a state in which more of my beliefs are satisfied) implies that I grant ϕ substantial evaluative significance in comparison to ψ , and hence there is no paradox: I *am* better-off in ϕ . (In other words, I treat the loss of the intrinsic goods as a result of the non-satisfaction of my other beliefs as worth it for the intrinsic good of having fewer beliefs satisfied. Odd, perhaps, but not paradoxical.) Of course, I might maintain other beliefs that imply ψ 's betterness than ϕ . But if so I have incoherent beliefs and, for JS, no paradox arises.

This example sheds light on a wider lesson. For JS to generate a paradox in assigning x 's welfare score, it must be the case that x 's set of evaluative beliefs assigns inconsistent evaluative valences (like “better” and “worse”) to a particular state of affairs ϕ . Otherwise one's belief set would assign a consistent evaluative valence, and there would be no paradox. But this sort of inconsistency is precisely what the coherence condition is designed to rule out: any set of beliefs that assigns inconsistent evaluative valences to a state of affairs is, strictly speaking, incoherent.

of desire in virtue of its appeal to a weak coherence constraint on valuing. But why couldn't a desiderative view appeal to the very same thing, and in so doing solve the paradox? Unfortunately, this proposal does not succeed. Say that I desire to be worse-off, but I am indifferent to everything else. This is surely a coherent desire set, if any is. But under these conditions the paradox remains. In not satisfying my desire to be worse-off, this desire is frustrated, and hence I become worse-off, and hence the desire is satisfied, and I become better-off, and hence worse-off, etc., etc. If that's correct, a coherence constraint is no savior for DS.

Again, I do not wish to make the strong claim that there is *no* way to understand DS such that it could avoid the paradox of desire. But even if such a view could be found, the solution offered by JS is far more satisfying, and gets at the heart of the issue: it is incoherent *in itself* to believe that one is better-off by being worse-off.

4. A Response: Favored Conditions

I argue that a judgment-based theory of valuing offers more satisfying verdicts than a desiderative theory of valuing, but also offers two important structural advantages: it can respond, in a natural and satisfying way, to two persistent difficulties for a desiderative interpretation of welfare subjectivism.

However, I have not yet considered a desiderative theory of valuing with the benefit of full context. In particular, most versions of DS hold that ϕ is good for x only if x values ϕ under specified idealized or counterfactual conditions. But I did not consider the possibility that a desiderative theory of welfare might avoid the problems noted here when combined with a theory of favored conditions. Perhaps, it might be argued, when placed in such context, a desiderative view succeeds.

But this response misunderstands the dialectic. I have so far argued that DS is incorrect because it offers a problematic theory of valuing: desire. But this problem remains no matter what set of favored conditions one accepts. Consider, for instance, the classic proposal that that which is good for x is that which x would desire under conditions of full information.³⁷ The problem is that that which one *desires* with full information needn't be coextensive with that which one *values* with full information. Take Stan. There is no guarantee that with full information Stan will desire either op-

³⁷This strategy is pursued by a number of desiderative theorists, in various incarnations, including Sidgwick, 111-15; Rawls, §64; Brandt, 132; Railton, op. cit.; Lewis, op. cit.

tion. The problem with Stan is not that he is ill-informed, but rather than he is *burnt-out*. But merely because, assume, Stan possesses full information and does not desire either option seems insufficient to declare that he doesn't value either option. It seems right to suggest that Stan could value moving to Small Town whether or not he suffers from a case of "desiderative exhaustion" even granting conditions of full information.³⁸ Furthermore, this possibility does little to salvage the second-order desiderative view. We may well believe that Jack and Julie are fully aware of all information. Indeed, it is the information that they possess that leads them to want to want things that they do not value, or to fail to want to want things that they do value. The problem here traces back to desire: there is no guarantee that under conditions of full-information a person will value that which they desire, any more than they will value that which they desire in their actual, benighted, state.³⁹

Before retiring this response, I should consider a further possibility at which I gestured in the previous section. JS can avoid some of the problems facing DS, or so it may seem, because JS employs a *coherence* constraint on the nature of valuing: that which I value is that which I would believe is good for me were my beliefs rendered coherent. But why couldn't a desiderative view adopt the same constraint?

However, this proposal also cannot succeed. (One could construe the coherence constraint as an element of a desiderative theory of valuing *or* as part of a theory of favored conditions. I assume here that it should be treated as an element of a theory of favored conditions, but this is neither here nor there. The problems are the same either way.) First, even if we offer an account of what it means for a desire set to be coherent (no trivial task), the general problems noted in the previous sections remain.

³⁸Further problems with the full-information account are noted in Arneson, 126-135; Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990), 18-22; David Velleman, "Brandt's Definition of 'Good'" in *The Philosophical Review* 97 (1988).

³⁹Railton suggests that the favored conditions are not just information (or what he calls "awareness"), but also *rationality*. But this does nothing to solve the problems of Julie and Jack: it would seem that Julie and Jack adopt their second-order desiderative stances in a straightforwardly rational way, given their circumstances. In addition, it is hard to see the desires of the drug addict, or Stan's lack of desires, as irrational *unless* one is willing to suggest that something other than their desiderative sets determines that which is good for them. But this is clearly a problematic suggestion for DS. Similarly, Brandt argues that the theory of favored conditions includes full-information and "cognitive psychotherapy". But though such therapy may very well cure Stan's burn-out, or the drug addict's addiction, it seems to do little in the case of Julie and Jack: that which they desire to desire after cognitive psychotherapy is unlikely to change.

Any rubric by which to render an incoherent set of desires coherent must privilege some set of desires. But there is no set of desires that might be so privileged that guarantees that a resulting coherent desire set reflects that which a person values. One obviously cannot privilege second-order desires for reasons already well-rehearsed. Furthermore, privileging the *stronger* desire (whether first- or second-order) clearly fails. The drug addict’s desire for the daze may be overwhelmingly strong, hence any coherent set of desires that privileges his strongest desires will not conform to the drug addict’s values: the drug addict clearly does not value taking the drug. Another possibility is offered by Alan Goldman.⁴⁰ Goldman claims that in rendering incoherent desires coherent, one does not keep fixed the strongest desires, but rather the “deepest” desires, where “depth” is determined not by strength, but by the power of a particular desire to explain one’s other motivational states.⁴¹ (For instance, my desire to be a professional philosopher is “deep”, because this explains a number of other desires I have, e.g., my desire to teach this class, read this book, revise this paper, etc.) Goldman’s proposal, however, is no help; alienating desires can be writ large. Take, for instance, a desire to *be thin* for its own sake on the part of a person with an eating disorder. One might explain much of what he or she desires, the actions he or she performs, etc., with reference to this deep desire, though we should not conclude, simply on this basis, that a person *values* the state of being thin.

In concluding this section, I offer a general point. I doubt very much that any theory of favored conditions could allow a general convergence of desire and value. But even if we find a set of conditions under which there is such a general convergence, does this imply that, under such conditions, desire is the *source* of value? I think not! All this would imply is that we have found a set of conditions that molds a person’s desires to fit that which she values. To determine whether desires are the source of a person’s values under this favored set of conditions we must consider the potential *counterfactual* divergence of desire and value—the extent to which desire and value converge under *other* conditions. But, as we have so far seen, under the vast array of conditions we might describe there is no proper convergence between desire and value. This is strong evidence that, even if some conditions can be described that would entail such convergence (which seems to me doubtful), any such convergence is at best accidental.

⁴⁰ Alan H. Goldman, *Reasons From Within: Desires and Values* (Oxford: Oxford University Press, 2009), 23.

⁴¹ Goldman, 38.

5. *Objection: Embarrassing Cases*

I have so far argued that judgment-based theory of valuing is superior on three grounds. First, it better accommodates our considered judgments of what it means for someone to value something, as displayed in the cases explored in sections 1 and 2. Second, it eliminates the classic problem of prudential versus non-prudential valuing. Third, it can avoid, in an extremely natural way, the paradox of desire. Hence, there are good *prima face* reasons to accept a judgment-based theory of valuing.

What remains is to consider three important objections to JS. The first objection takes aim at the first purported advantage of JS. I have so far attempted to embarrass a desiderative theory of valuing by considering a number of cases that a desiderative theory cannot appear to handle adequately. However, what's to stop a desiderative theorist from making precisely the same accusation? Are there no embarrassing cases for a judgment-based view? And if there are, surely a judgment-based view is no more accommodationist than a desiderative theory of valuing.

First, JS might be under-inclusive. Imagine, for instance, an inversion of Stan. Imagine that Stan has no belief regarding whether moving to Small Town or remaining in Big City would be good for him. However, suppose that he desires (and, perhaps, desires to desire) to remain in Big City. Surely we would be tempted to say, in this case, that Stan values remaining in Big City despite a failure to believe that so doing is good for him. Second, consider:

Janet: Janet believes that developing an interest in classic cars would be good for her. But Janet has no desires in any way connected to this belief: she *hates* classic cars, and has no second-order desire to develop her taste in classic cars.⁴²

For Janet, one may be tempted, despite her belief that it would be good for

⁴²I have purposefully stated Janet's case such that it is relatively easy to believe that none of Janet's desires would be advanced by the satisfaction of her relevant evaluative belief. An anonymous reviewer suggests that I consider an alternative case that would also guarantee this: the case in which Janet desires that the temperature on Pluto should be n degrees Kelvin. However, though I offer the same response to both cases, I consider the latter case unfairly prejudicial, given that considered judgments about whether Janet values such a state may be unfairly prejudiced by a common intuition that no one could be made better-off by such a state (which is no more a problem for JS than DS). In sorting out the relative plausibility of JS and DS, it seems sensible to focus on cases of valuing the objects of which seem relatively uncontroversial as welfare benefits.

her, to say that Janet does not value developing her tastes in this way.

I think the plausibility of these verdicts should be disputed. Take Stan. If it is really true that Stan has no belief regarding which option would be better for him, I am hesitant to say that Stan can really be said to genuinely value remaining in Big City. Let's say that you ask Stan what he thinks about the move, and he replies: "Well, I know that I want to remain in Big City, and I do have a desire to maintain this desire. But I'm just not sure that I believe that remaining in Big City is good for me." Of course, we would likely be puzzled by Stan's response, insofar as we generally believe that second-order desires are responsive to beliefs about value. But leave this aside. I'm tempted to say that in this case, we would be unsure about what Stan values: and our uncertainty derives from Stan's failure to believe that one is better than the other. We would likely say that Stan is conflicted. But this is a problem for a desiderative theory of valuing: Stan's desires are *not* conflicted.

What if Stan's failure to believe in the prudential value of either option is, say, a product of "epistemic exhaustion", which parallels Stan's "desiderative exhaustion" ("burn-out") discussed in §1? Imagine, for instance, that Stan has difficulty believing in the value of anything because of rigorous and exhaustive work in the testing of his own beliefs. He just can't bring himself to believe in the intrinsic value of either moving to Small Town or remaining in Big City. Despite this, he desires (desires to desire, etc.) to stay in Big City. Would we really believe that he doesn't value staying in Big City? To state my own intuition, I think in this case we would be reluctant to say with any confidence that he values staying put, insofar as he cannot bring himself to believe in the value of so doing. You might think this is implausible; after all he *desires* to stay in Big City. But I think Stan's desires are not evidence that Stan values staying in Big City under such conditions. Though they may be important, it seems more plausible to describe them as an indicator of Stan's *disposition* to value staying in Big City; a disposition that, given his desiderative profile, we would expect to activate once he gets over his exhaustion. But we should be reluctant to say that his desires are evidence that he values moving to Big City *now*, or, to use a bit of terminology, "occurrently". Insofar as Stan can't bring himself to believe in the value of anything, Stan can't bring himself to *value* anything.⁴³ Sometimes we are too exhausted to value—this case is a prime example. One might think that DS could make the same claim, i.e., that Stan's burn-out is just an example of failing to occurrently value. But this seems to me less plausible. In the

⁴³Thanks to an anonymous reviewer for this case and helpful suggestions.

case described in §1, it seems right to say that Stan values moving to Small Town *even if* he is burnt-out.

I think a similar thing should be said for Janet. Suppose we ask whether Janet desires to develop an interest in classic cars, and she replies: “No, not really.” Without further information, we would be generally tempted to say that Janet does not value the development of such an interest. However, if Janet goes on to say: “However, I do really believe that doing so would be good for me”, or something to that effect, we would likely alter our verdict. Perhaps she does, after all, value developing an interest in classic cars despite her desires. Much of the importance of desiderative states in determining what individuals like Janet and Stan value, it seems to me, derives from a general assumption that an individual’s desires (especially their second-order desires) are reliably correlated with their judgments about that which would be good for them. However, in cases in which one lacks a desire for ϕ , but nevertheless believes that ϕ -ing would be intrinsically good, it seems to me sensible to say that we would regard their desires as misleading when it comes to that which they genuinely value.

A further possibility should be considered. Imagine:

Betty: Betty is a 1960s housewife. She finds that she does not enjoy her life, and finds that she occasionally longs to leave it (and, perhaps, desires to desire to leave it). But, given the culture in which she lives, she has been socialized to believe that being a housewife is, in fact, good for her.

Betty’s case might present a further wrinkle for JS. If we assume that her belief is merely a result of this socialization, do we really wish to say that Betty *values* being this sort of a housewife? But wouldn’t JS—and its accompanying theory of valuing—declare that she does?

I begin by noting that *Betty* differs from *Janet* only in the element of socialization. But two points should be made here. First, it would be an error to say that someone’s values cannot be altered by socialization mechanisms to which they are subject. If so, we may be tempted to say that Betty does, in fact, value her life as a housewife. Her values have been socialized. However, and more to the point, this case does display a very serious problem for JS: if we take JS seriously as a theory of well-being, we would have to say (or so it would seem) that it is *better* for Betty to remain a housewife. And this conclusion is absurd. However, this is a problem shared by DS: desires, no less than beliefs, are subject to this sort of socialization, as noted by many.⁴⁴ In addition, and as I have discussed

⁴⁴See, for instance, Amartya Sen, *On Ethics and Economics* (Oxford: Blackwell, 1987),

elsewhere, many subjectivists avoid this troubling conclusion by offering a robust theory of favored conditions that adequately weed out values that are subject to socialization.⁴⁵ If such a theory of favored conditions can be found, there is no reason to believe it is less successful for JS than for DS. Indeed, I argue elsewhere that *only* JS can solve this problem.⁴⁶

Of course, these are difficult cases. It seems to me, however, that a judgment-based theory of valuing gives the correct answers. But it is likely that others will disagree. Furthermore, it's possible that any range of additional cases could be conjured that shed additional unflattering light on JS as opposed to DS. But even in this worst-case scenario, however, the case for JS remains strong. Point one: even if the judgment-based view is embarrassed by the cases considered here, this does not mean that there are not a host of problems for a desiderative view; note again the problems explored in §1. Hence, point two: even if the judgment-based view gets the answers intuitively wrong in some cases, when it comes to our considered judgments the results are, at worst for my argument here, inconclusive. But, point three: if we assume that consideration of cases is inconclusive between JS and DS, the balance of reasons still points to the judgment-based, rather than desiderative, theory of valuing. JS maintains two significant and important structural advantages over DS. Hence, point four: even if the results of all the variations of the cases considered here turn out to pull neither in favor of a desiderative nor judgment-based theory of valuing (which I strongly doubt), the advantages of JS should be manifest, and should provide decisive reason to accept a judgment-based subjectivism.

6. *Objection: Existence Internalism*

A classic rationale for DS is *existence internalism*. To quote Connie Rosati, “the label ‘existence internalism’ . . . refers to the general thesis that there is a necessary connection between motivation and normative status.”⁴⁷ Existence internalism claims that, necessarily, if ϕ is good for x , x must have some motivation to promote ϕ . (Note that x 's motivation to promote ϕ needn't be actual. Existence internalism, at least on Rosati's account, can be satisfied by motivation under a variety of favored conditions, such as

45-6; Martha Nussbaum *Women and Human Development* (Cambridge: Cambridge University Press, 2000), 114; Jon Elster, *Sour Grapes: Studies in the Subversion of Rationality* (Cambridge: Cambridge University Press, 1985).

⁴⁵See, for instance, Dorsey, *op. cit.*, Sumner, ch. 6.

⁴⁶See Dorsey, *op. cit.*

⁴⁷Connie Rosati, “Internalism and the Good for a Person” in *Ethics* 106 (1996), 297.

full information, full awareness, etc.) Existence internalism would cause no problems for JS as opposed to DS *were it not for* a widely accepted position concerning the nature of motivation: only desires motivate.⁴⁸ If this view is correct (which I simply accept for the purposes of this paper), accepting existence internalism appears to commit one to the authority of desire over “normative status” and hence a person’s good (on the uncontroversial assumption that a person’s good has “normative status” of some kind or other). And if this is correct, subjectivism about welfare is forced to accept a desiderative theory of valuing, warts and all.

In this section, I address arguments for existence internalism. My general claim will be that purported arguments for existence internalism fail to support existence internalism in a form that would support DS over JS. The most plausible arguments for existence internalism support subjectivism more broadly, but have no power to decide between alternative theories of valuing.

6.1. *The Principal Intuition*

Why should we believe existence internalism? While I will discuss other arguments below, a crucially important one involves a straightforward appeal to intuition. Consider, for instance, the following, from Railton:

Is it true that all normative judgments must find an internal resonance in those to whom they are applied? While I do not find this thesis convincing as a claim about all species of normative assessment, it does seem to me to capture an important feature of the concept of intrinsic value to say that what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone’s good to imagine that it might fail in any such way to engage him.⁴⁹

Rosati, in discussing arguments for existence internalism, refers to sentiments like Railton’s as *the principal intuition*:

The principal intuition supporting internalism about a person’s good, as aptly expressed by Railton, is that an individual’s good

⁴⁸See Michael Smith, “The Humean Theory of Motivation” in *Mind* 97 (1987). For a contrary view, see John McDowell, “Virtue and Reason” in *Monist* 62 (1979).

⁴⁹Peter Railton, “Facts and Values” in *Facts, Values, and Norms* (Cambridge: Cambridge University Press, 2002), 47.

must not be something alien—it must be “made for” or “suited to” her. But something can be made or suited to an individual, the thought goes, only if a concern for that thing lies within her motivational capacity: what is good for her must connect with what she would find “in some degree compelling or attractive, at least if [she] were rational and aware.” In this way, there must be a “fit” between an individual and her good.⁵⁰

I hereby stipulate that internalism of this kind *is* compelling—especially for those already inclined toward a form of subjectivism about the good. However, I argue, the “principal intuition” seems to trade not on the appeal of the connection between a person’s good and her *motivations*, but rather on the appeal of subjectivism more broadly.

Railton argues that, for instance, the good must find an “internal resonance,” that people must find their good “compelling or attractive,” that it would be problematic for an account of a person’s good to be “alienating.” Rosati notes that a person’s good must not be “alien,” that it should be “suited for” the person in question. It is worth wondering, however, whether *any* of these claims require that a person’s good be ratified by means of a motivational state. Note that Stan’s desire set, and hence motivational set, is indifferent between staying in Big City or moving to Small Town. But assume for the moment that he believes moving to Small Town would be better for him. Under such conditions it would certainly be implausible to say that Stan sees nothing “compelling or attractive” in moving to Small Town. After all, he *believes it would be good for him* to so move. There is an important sense in which Stan would not find the suggestion that moving to Small Town would be good for him “alien”. Indeed, it seems plausible to believe that this claim is ideally “suited for” Stan. As stated, the “principal intuition” appears to support *subjectivism*, rather than DS in particular. It seems straightforward that if Stan believes ϕ is good for him, even without a desire, ϕ is at the very least “suited” to him.

Perhaps Railton and Rosati wish to define “non-alienation” or “resonance” or “to be suited to” in such a way that for a particular good ϕ to be “suited to” x , ϕ must engage her motivational set. But even leaving aside the fact that neither argue for this interpretation, there are three problems with simply stipulating this. First, in so doing, the principal intuition appears to cut little ice for existence internalism. If to care about something simply is to be motivated by it, the principal intuition is little more than table-pounding. Second, Rosati herself *explicitly* rejects a strictly motivational

⁵⁰Rosati, 298-9.

interpretation of existence internalism. Rosati suggests that her internalist constraint requires a “double motivational link” between a person and her good. But Rosati claims that internalism should be linked to motivation “in the broad sense.” She understands the notion of motivation in a “broad sense” as linked to the development of “a proattitude—such as desiring, liking, being glad of, caring about, and so on—an attitude which may *or may not* be a motive to action. To say that something must motivate, in the broad sense, to be a part of a person’s good, is to say that it must be something that can, in a positive way, matter to her or be an object of her concern.”⁵¹ Beliefs, however, are attitudes, and x ’s belief that something ϕ is good for x is arguably just as “pro” an attitude as a desire or motivation. It is strongly intuitive to believe that, for Stan, moving to Small Town is something that is a matter of his concern. After all, it is something he believes is good for him, to which he is committed even in the absence of a corresponding desire.

Third, it seems wrong to say that x ’s good should be “suited to” an attitude possessed by x that does not reflect what x values. But if this is correct, whether a judgment-based subjectivism can satisfy the principal intuition is entirely dependent on whether a judgment-based theory of valuing is successful. If so, the principal intuition can offer no *independent* objection to JS: if JS is otherwise plausible as a theory of valuing, this *implies* that it satisfies the principal intuition.⁵²

⁵¹Rosati, 301; my emphasis.

⁵²An anonymous reviewer suggests the following response. Imagine a person, Jane, who believes ϕ is good for Jane. But also imagine that Jane does not *value her own good*; she is, perhaps, a serious self-loather. If Jane is a self-loather, we might be tempted to say that though she believes ϕ is good for her, she does not take a pro-attitude toward ϕ . But I think this is the wrong way to understand the case. That Jane takes an anti-attitude toward ϕ , given her self-loathing, does not mean that she does not *also* take a *pro*-attitude toward ϕ in the form of her belief that ϕ is good for her. However, this prompts a deeper question about the cogency of JS as a version of subjectivism. JS holds that x can plausibly be said to value ϕ if x believes that ϕ is good for x . If Jane is a self-loather, can she really be said to *value* ϕ ? If not, it would seem as though a belief that ϕ is good for her is insufficient to say that Jane *values* ϕ .

Before I respond, it is worth noting that this is no less a problem for DS than for JS. Assuming that DS characterizes “valuing” as “desiring” (in some form or other), DS must admit that Jane desires ϕ , but because Jane is a self-loather, she fails to value ϕ . Hence “valuing” cannot be identified as desiring, either. But JS, and not DS, can avoid this critique. JS can make a distinction in ways of valuing that is unavailable to a desiderative theory. (See §3.1.) For JS, a reasonable interpretation of this case may be to characterize the attitude Jane takes to her own good as follows: though Jane prudentially values ϕ , she perhaps fails to believe it is good *on the whole*, despite its being good *for her*. But a desiderative theory of valuing—as already explored—has a very difficult time making the

6.2. The Metaphysics of “Good”

Rosati notes a further argument for existence internalism. Rosati writes, following R. B. Perry, that “if value can exist only if there are creatures who can be affected by and react to their world, then value, and more specifically, goodness for a person, must be a motivational property. What else, after all, could it be? The only alternative might seem to be that the property of being good for a person is a Moorean, nonnatural property, but this alternative introduces special metaphysical and epistemological problems.”⁵³ The argument from the metaphysics of “good” has an important place in arguments for DS. DS is often defended on grounds of naturalism, i.e., that it is most consonant with our naturalist, scientific worldview.⁵⁴ Though this argument is important, it again fails to distinguish desiderative and non-desiderative theories of valuing. The point of the argument is that value has to be centered in persons—*not* found in Moorean nonnatural properties, or properties that cannot be explained given the assumption of naturalism. But this is perfectly consonant with JS. JS does not posit Moorean nonnatural properties: value is centered in persons, just not in a person’s *desires*. Rosati does not believe that this version of Perry’s argument is particularly strong (for starters, it fails to consider a number of possible externalist views). But Rosati’s *reconstruction* of the argument even more clearly cannot support a desiderative subjectivism: “if goodness for a person is not a complex motivational property, then either it is a peculiar nonnatural property, or it is some other kind of property which itself presupposes the truth of internalism. The conclusion would then be not ‘either a strange non-natural property or a motivational property,’ but ‘either a strange nonnatural property or internalism.’”⁵⁵ But stated in this way, JS is clearly compatible with the argument from the metaphysics of “good”. Rosati insists that any view that is not committed to Moorean non-natural properties must be committed to goodness *qua* motivational property, or goodness *qua* some other property that properly falls under the internalist heading. But because JS is in no way committed to Moorean non-natural properties it *must* be compatible with internalism. It would thus appear

distinction between prudential and non-prudential valuing.

⁵³Rosati, 313-14.

⁵⁴This defense is evident in Railton, “Facts and Values” and “Moral Realism” in *Facts, Values, and Norms*, op. cit. Thanks to Connie Rosati for reminding me of this important point.

⁵⁵Rosati, 314-15.

that the argument from the metaphysics of “good” supports subjectivism on the wholesale and hence cannot rule out JS.

6.3. *Judgment Internalism*

A final argument for existence internalism—which, if successful, really *would* be an argument for DS, appeals to the truth of *judgment internalism*, viz., the view that “[a] person cannot sincerely judge that something is good for herself unless she has some tendency to approve of or pursue that thing.”⁵⁶ The connection between motivation and judgment needn’t guarantee motivation sufficient for action; rather, “sincere judgment” requires that the person whose good it is “normally have some inclination, not necessarily overriding, to promote or to care about that thing.”⁵⁷ The move from judgment internalism to existence internalism is common.⁵⁸ Rosati states the argument like this:

The truth of judgment internalism might seem to support the claim that a plausible account of the good for a person must satisfy existence internalism. . . An account of the good for a person must permit judgments about a person’s good to serve their characteristic action-guiding functions. It must be able to explain how it is that, at least normally, judgments about a person’s good motivate, and it must also preserve their characteristic recommending and expressive functions or normative force. An account can succeed in this, without embracing noncognitivism and its antirealist implications, only if it satisfies simple internalism. By limiting a person’s good to some subset of those things that can matter to her, an account insures that it will be at least possible for judgments about a person’s good to perform their characteristic functions.⁵⁹

Assume for the moment that we should interpret “promote or care about” or “approve of or pursue” as requiring a desire sufficient to motivate. Under such an interpretation, judgment internalism claims that to have a sincere judgment about one’s own good, it must normally be the case that the person

⁵⁶Rosati, 310.

⁵⁷Rosati, 310.

⁵⁸See also Simon Blackburn, *Ruling Passions* (Oxford: Oxford University Press, 1998); Thomas Nagel, *The Possibility of Altruism* (Princeton, NJ: Princeton University Press, 1970), 7.

⁵⁹Connie Rosati, “Internalism and the Good for a Person” in *Ethics* 106 (1996), 310-11.

in question has some desire the content of which is the purported good in question. In this way, DS is well-placed to explain judgment internalism in a way that non-motivational subjectivisms (like JS) are not.

Whether judgment internalism is true is a vexed philosophical question. I will assume it for the sake of argument. At first glance, however, it is a little difficult to see why judgment internalism would support DS. After all, if judgment internalism is true, this just provides a link between evaluative judgment and desire; necessarily, they “normally” co-exist. But this appears compatible with a view that says that ϕ can only be a genuine prudential good for x if x judges that ϕ is a genuine prudential good. This view would simply have it that, as a matter of psychological fact, whenever persons judge that ϕ is good for them, they are normally motivated to promote ϕ . Hence nothing about judgment internalism is incompatible with JS. Furthermore, the only tool that would allow DS to explain judgment internalism is its peculiar constraint on prudential value: for some claim “ ϕ is good for x ” to be true, ϕ must motivate x . This will imply that whenever x *truly* judges that ϕ is good, x will be motivated to promote ϕ . But it says nothing about cases in which x might *falsely* judge that ϕ is good. It is fully compatible with DS that x will falsely judge that ϕ is good for x , and fail to be motivated by it under any circumstances (even “normal” ones). Thus DS *of itself* cannot explain judgment internalism. To explain judgment internalism, DS must rely on a further fact: a reliable psychological connection between x ’s state of judging that “ ϕ is good for x ” and x ’s desire to promote ϕ .⁶⁰ But any theory of prudential value—whether compatible with existence internalism or not—can accept judgment internalism if it is explained, not by one’s theory of prudential value, but rather psychologically, by a reliable connection between the dual mental states of judgment and desire.

We should reject the claim that existence internalism provides a rationale for DS as opposed to JS. The most plausible arguments for existence internalism support subjectivism but have no power to rule out JS.

7. *Objection: Circularity*

The argument I consider in this section might be thought of as the desiderative subjectivist’s trump card. It is all well and good, it might be claimed, to defend the intuitive authority of one’s judgments over that which one values. Nevertheless, JS has special problems that DS does not have. JS is

⁶⁰Cf. David Lewis, “Desire as Belief II” in *Papers in Ethics and Social Philosophy* (Cambridge: Cambridge University Press, 2000), 57-9.

stuck with a circular analysis of the concept of welfare. David Brink states this objection concisely:

[S]omeone might analyze goodness as a property of objects that tends to elicit in ideal conditions and appraisers the judgment that it is good or valuable. Here we invoke the very value we are analyzing in our analyses. It is true that, on this view, we analyze X, not in terms of X, but in terms of beliefs about X. But if we accept the not unreasonable assumption that any story about what makes a belief a belief about X must eventually advert to X, then it appears that this sort of analysis is ultimately circular.⁶¹

Here it would appear that in analyzing the nature of “good for x ”, we must make use of the term “goodness for x ”. After all, the analysis contains “good for x ”: that which is good for x is that which x (coherently) believes is good for x . But this seems to provide a straightforwardly uninformative analysis of what it means for something to be good for x . In providing an analysis of welfare in terms of judgments about welfare, we must then proceed to analyze judgments about welfare. But if judgments about welfare are analyzed in terms of the concept of prudential value, we have offered a straightforwardly circular analysis of the nature of this concept.

This problem can be sidestepped. There may be many solutions, I limit my exploration to one. (I am officially neutral on whether there are additional, perhaps better, solutions.) Note that JS as stated bears important similarities to views that have come to be known as “constructivist” in the normative domain.⁶² For instance, Sharon Street states metaethical constructivism in the following way: “the fact that X is a reason to Y for agent A is constituted by the fact that the judgment that X is a reason to Y (for A) withstands scrutiny from the standpoint of A ’s other judgments about reasons.”⁶³ This view is similar in structure to JS, and seems to present similar worries about circularity.

But this is significant for JS. Constructivists have developed sophisticated responses to the circularity worry, from which JS could borrow liberally.⁶⁴ For instance, Street responds to the charge of circularity by marshal-

⁶¹David Brink, “The Significance of Desire”, op. cit., 20-21. See also Lewis, 69-70.

⁶²See, for instance, Sharon Street, “Constructivism about Reasons” in *Oxford Studies in Metaethics*, v. 3, ed. Shafer-Landau (Oxford: Oxford University Press, 2008).

⁶³Street, 223.

⁶⁴I have explored these possibilities in detail in Dale Dorsey, “Truth and Error in Morality” in *New Waves in Truth*, ed. Wright and Pedersen (Palgrave, 2010) and “A Puzzle for Constructivism and How to Solve It”, MS.

ing a thought familiar from response-dependence theorists of color. Very roughly, such a view holds that “red” should be analyzed in terms of “looks red”. But “looks red” should be understood as a primitive *redness* experience that, presumably, we all understand.⁶⁵ Street’s view is similar. Street writes: “The idea of one thing’s being a reason for another cannot successfully be reduced to thoroughly non-normative terms. Instead, I would argue, our understanding of this idea is given by our knowledge of *what it is like* to have a certain unreflective experience—in particular, the experience of various things in the world as ‘counting in favor of’ or ‘calling for’ or ‘demanding’ certain responses on our part.”⁶⁶ For Street, the proper response to the charge of circularity is to hold that “the notion of a reason” in terms of which we form judgments, is “primitive”.⁶⁷ A similar maneuver can be exploited by JS. Judgments about the prudential good, it might be claimed, cannot be fully reduced to non-evaluative judgments: they are judgments that necessarily refer to a certain unreflective understanding of what it means for something to improve the quality of a person’s life. Though I refrain from making the claim (which may be plausible or implausible) that these beliefs refer to a *sui generis* experience (à la redness), certainly the notion of ϕ ’s being intrinsically good for me is something I have a basic understanding of and on the basis of which I can form beliefs, which in turn constitute my evaluative perspective, and which in turn form the basis of my welfare.⁶⁸ For JS, the concept of welfare—about which we form beliefs that constitute our good—is primitive. In this way, circularity (or, at any rate, vicious circularity) is avoided.

One might think that if one cannot analyze the concept of welfare in non-evaluative terms that this is a problem for JS. However, as Street and David Wiggins note,⁶⁹ whether such an analysis is a failure depends on the sort of theory we intend to be offering. As Street notes, constructivism about reasons is designed simply to offer a *truth condition*: constructivism “reduces facts about reasons to facts about what we *judge* or *take* to be

⁶⁵See, for instance, Philip Pettit, “Realism and Response-Dependence” in *Mind* 100 (1991), 602-5.

⁶⁶Street, 239-40.

⁶⁷Street, 242. Street’s response to the charge of circularity is similar to Jonathan Cohen’s rejection of the charge of circularity for his “functionalist” account of the nature of color. See Cohen, “Color: A Functionalist Proposal” in *Philosophical Studies* 113 (2003), 10-11.

⁶⁸Street, 241.

⁶⁹Cf. David Wiggins, “A Sensible Subjectivism?” in *Needs, Values, Truth* (Oxford: Oxford University Press, 1987), 188-9.

reasons”.⁷⁰ Same here. Subjectivism about well-being, when combined with a judgment-based theory of valuing, offers a theory of when a particular ϕ is good for a particular person x , and the extent to which ϕ is good for x . In this way it “reduces facts about” welfare to “facts about what we *judge* or *take*” to be in our interest. But it does not seek to offer any robust analysis of the latter notion; this idea it takes as simply primitive. As Wiggins writes, “Circularity is no objection...so long as the offending formulation is also *true*.”⁷¹

Of course, DS *can* offer an analysis of “goodness for x ” that is fully reducible to non-evaluative terms, i.e., in terms of desire. Insofar as DS offers such a reductive analysis and JS—given its primitivism—does not, is this a reason to prefer DS to JS? Perhaps; I grant for the sake of argument that reductive accounts, other things being equal, are to be preferred to non-reductive accounts.⁷² However, as I have so far been at pains to argue, other things are not equal. Whether we should accept a reductive or non-reductive analysis of “goodness for x ” seems to me to depend most significantly on whether such an analysis can allow us to say things about the *facts* of prudential value we wish to say. If, therefore, our considered judgments concerning when ϕ is prudentially valuable for x , or what conditions must hold such that ϕ is prudentially valuable for x , align more closely with JS than DS, or if JS is a more appealing account of what a person genuinely values, there should be no objection to its primitivist analysis. To accept a view that allows a reductive analysis at the price of an inferior theory of welfare seems to me bad philosophical business.

Let me put this point in a slightly different way. Some hold that an important motivation for subjectivism about welfare is that, at least in its desiderative guise, it can offer a reductive analysis of welfare; prudential value can be fully analyzed in non-evaluative terms.⁷³ JS—if we accept the

⁷⁰Street, 242.

⁷¹Wiggins, 189. I should note a further circularity worry I discuss and attempt to solve elsewhere. In particular, constructivism, when combined with a standard truth predicate, seems to require a circular semantics of normative judgments, leaving aside the analysis of “good for x ”. However, insofar as I propose solutions elsewhere, I refer readers interested in this technical concern to “Truth and Error in Morality”, op. cit., and “A Puzzle for Constructivism and How to Solve It”, MS.

⁷²Note, of course, that JS is reductive in the sense already indicated: it reduces facts about welfare to facts about what we *believe* about welfare. But it is not reductionist in this broader sense. (Compare Street, 242.) I assume for the sake of argument that this is an important cost.

⁷³Cf. Railton, “Facts and Values”, 62-3; “Moral Realism”, op. cit. Thanks to an anonymous reviewer.

“constructivist” interpretation I’ve been sketching here—cannot accommodate this motivation. I admit that for some this is a cost. However, surely another important motivation is the central intuition noted at the opening of this paper: that an individual’s welfare should be determined, in full, by that which she values. But if JS offers a better theory of valuing *and* requires a primitivist analysis (I stress again that I leave open the possibility of that the right-hand conjunct is false), it would appear that these motivations are in tension: a primitivist analysis is the price of capturing the important connection between x ’s welfare and x ’s true values. And though I don’t speak for all subjectivists, it seems to me that this is a price worth paying.

8. Conclusion

I hope to have done the following things in this paper. First, I hope to have shown that a judgment-based theory of valuing is plausible not just as a result of a number of intuitive verdicts, but also as a result of consideration of two structural defects in DS that JS artfully avoids. In addition, I hope to have shown that existence internalism, though a powerful argument in favor of subjectivism, has no power to decide between versions of subjectivism. Furthermore, the argument from circularity can be plausibly dispatched either by treating the concept of prudential goodness as primitive, in a manner analogous to constructivism about reasons.

I also wish to mention briefly what I have not done. First, I have not fully articulated a judgment-based subjectivism. In particular, I have not offered any theory of favored conditions.⁷⁴ I have argued only that in constructing a subjectivist theory of welfare, we are well-advised to replace a desiderative theory of valuing with a judgment-based theory of valuing. Moreover, I have not argued for subjectivism of any kind against *objectivism*. The argument of this paper is addressed to those—like me—who believe that the personal good must be responsive to a person’s own evaluative perspective. But in arguing for subjectivism, the first step is to find the most plausible articulation of its main themes. I am optimistic that JS fits this bill.

⁷⁴Though I do so elsewhere, see “Preferences, Welfare, and the Status-Quo Bias”, *op. cit.*