**INTEGRITY, THE SELF, AND DESIRE-BASED ACCOUNTS OF THE GOOD**

***Robert Noggle***

***Central Michigan University***

***R.Noggle@cmich.edu***

**This is an author’s archived copy of a manuscript accepted for publication in Philosophical Studies.**

**The official version of record is: "Integrity, the Self, and Desire‑Based Accounts of Individual Good." *Philosophical Studies*, Volume 96 (1999): 303-331. All citations should be to the published version.**

**INTRODUCTION**

Well-being--also known variously as welfare, prudential value, or individual non-moral good--is that which makes a person’s life go well for the person living it. Many of the most popular and influential accounts of well-being link it to desire-satisfaction. Such theories claim that a person’s well-being consists of the satisfaction of her desires (or some subset of them).[[1]](#footnote-1) Part of the attraction of such accounts is that they explain why well-being is good for the person to whom it belongs. Getting what we want matters to us. According to desire-based theories, well-being matters to us (and gains in well-being affect us) because it consists of getting what we want. The fact that desire-based theories make our well-being something that matters to us seems to be an advantage over theories that simply posit a list of things that make a person’s life go well, whether they matter to the person or not. Thus desire-based theories preserve a common intuition that well-being is at least partly internal to the person herself, and that any prudential reasons that considerations of one’s own well-being generate must be internal reasons. The reason that such internalism seems appropriate in a theory of well-being is that if we are measuring the extent to which a life is valuable for the person living it, then it seems that the criteria for evaluation must be those of the agent herself. The ends and goals in terms of which we evaluate the success of a person’s life must not, it seems, be completely alien to the agent’s own ends and goals. Such, at least, is the intuition behind desire-based accounts of well-being.[[2]](#footnote-2)

However, such accounts of well-being are vulnerable to a kind of slippery slope argument. For such accounts only seem plausible if the desires in question are not mistaken or irrational. But once one begins specifying what counts as a mistaken desire, it is difficult to resist the pressure to espouse a theory which claims that a person’s well-being consists of the satisfaction of desires that she would have if she were some sort of ideal observer. However, ideal observer theories seem to sever the connection between the person and her well-being that was one of the motivating intuitions of the theory, and for many philosophers, one of its main attractions. I’ll begin by discussing the problems with the ideal observer account and why the desire account is pulled to it. I will then argue that the desire-based approach to well-being can be saved by resting it upon what I hope will be an independently plausible theory of the self. This theory will distinguish the central concerns which are constitutive of a person’s identity from more peripheral concerns that are not. Finally, I will show how a version of the desire-based account that is based on this theory of the self can avoid the slippery slope to the ideal observer account.

**FROM MISTAKEN DESIRES TO THE IDEAL OBSERVER**

Let’s begin by seeing how the desire-based account of well-being is pulled toward the ideal observer view. We’ll begin with a simple but implausible version of the desire-based account and modify it in ways that seem necessary to make it plausible. Though each modification seems necessary, the end result will be a theory that seems implausible, or at least one which severs the theory from one of its main intellectual motivations.

Consider, then, the simplest desire-based theory of well-being. It would claim that well-being consists of the satisfaction of any desire. This simple theory is highly implausible, for it seems to get the wrong result in cases in which desires are based on false beliefs. The satisfaction of such desires does not seem to increase well-being. Thus, if I want a wrench (only) because I falsely believe that it will do the job I want to do, then getting the wrench does not really improve my situation. Cases like this suggest that well-being does not consist of the satisfaction of desires that are mistakenly based on false beliefs.

This modification to the desire theory of well-being seems absolutely crucial if it is to get off the ground at all. But it puts us on a slippery slope. For there are other ways that a desire can be mistaken, and in each case it seems that we will need to modify the theory to claim that well-being consists in the satisfaction of only non-mistaken desires.

Thus while it is clearly a mistake to base a desire on a false belief, it also seems to be a mistake to fail to want something because you have false beliefs about it. Suppose Cathy does not want to bring a cat into her house (only) because she falsely believes that she is allergic to cats. Were it not for this belief, she would very much like to have a cat. Cathy seems mistaken in not wanting a cat, and having a cat would apparently make her life go better for her. Examples of this sort suggest that a person's well-being consists of the satisfaction of desires that she would have if she had no false beliefs. Thus it seems that a plausible theory should say that Cathy is mistaken in failing (because of her mistaken beliefs) to desire something that would in fact make her better off.

However, if it is a mistake to fail to want something because you have false beliefs about it, then surely it is also a mistake to fail to want something merely because you lack information about it. Consider Dave, an unhappy loner who does not realize what wonderful companions cats can be. Suppose that Dave's life would be much richer and happier if he had a feline companion. Because Dave does not know this, he has no desire to bring a cat into his home. The thought has never crossed his mind, in fact. If it had, he would desire a feline companion to ease his loneliness. Dave's lack of a desire for a feline companion seems mistaken. Cases like this suggest that we should move to a “full information” version of the desire-based account of well-being. This version claims that a person's well-being consists of the satisfaction of those desires that she would have if she had full information.

However, even a fully informed person might seem to have mistaken desires. Belief by itself is often ineffective in producing desires.[[3]](#footnote-3) Without the right experiences, I may not desire to attend an opera even though I believe my more cultured friends when they tell me (truly, let us say) that opera would enrich my life. Yet perhaps if I did experience some opera, I would find that my friends were right. In such a case my failing to desire a night at the opera seems mistaken. I make the mistake not because of any cognitive shortcoming, but because I lack certain experiences. This is a common phenomenon. Often we know quite a bit about some thing or some activity, but have no desire for it until we try it and experience it as fun, or enriching, or otherwise appealing. Then we are inclined to say that we were mistaken in our previous assessment. To avoid making mistakes of this sort, one would have to experience each of one’s options.

Sometimes, however, even experience is not enough. Sometimes we need some education or training or other kind of preparation in addition to the experience of the thing in question. Some tastes, in other words, must be acquired. Yet once we do acquire them, we often say that their objects are valuable, that they were valuable all along, and that it was only because of our previous limitations that we were unable to see them as such.

Finally, some mistaken desires result from the presence or absence of certain character traits.[[4]](#footnote-4) For example, I might now be too shy to fully enjoy being a singer or stand-up comedian. Or I might now be too impatient to fully appreciate being a parent. But perhaps I could overcome my impatience and my shyness. And if I did, I might find these occupations immensely rewarding. The problem--the source of the apparent mistake--is that I currently have traits that prevent me from desiring experiences that would enrich and improve my life once my personality has developed in such a way as to allow me to appreciate them.

In order to avoid tying well-being to mistaken desires of this sort, one might be tempted to endorse the so-called ideal observer account of well-being.[[5]](#footnote-5) This account claims that a person's well-being consists of the satisfaction of those desires that she would have if she were an “ideal observer.” An ideal observer would be fully informed and experienced, and she would have the necessary education, training, and preparation, as well as the correct personality traits, to desire all and only those things that could enrich and improve her life.

However, this characterization of the ideal observer raises a crucial question: What is a "correct" or “ideal” personality trait? After all, if I developed one set of traits, I might desire things I would not desire if I developed other traits. If I became more spiritual, a life of asceticism and meditation might appeal to me. But if I became more worldly, I might desire to leave philosophy and get an M.B.A. How are we to decide which traits are ideal without assuming a prior conception of what enriches a life, of what well-being consists of? Of course we can give a neutral, non-question-begging account of correct belief by appealing to truth: a belief is correct if it is true. But how are we to decide which personality traits are correct for ensuring that the desires that result will be such that well-being consists in their satisfaction? And how are we to do this without making a circular appeal to a prior conception of what well-being consists of?

To be both ideal and neutral, an ideal observer would apparently have to consider the full information and complete experiences from the point of view of a person with each possible personality trait and each possible course of education. It is not enough, it seems, for the ideal observer merely to correctly imagine or even to experience all the possible objects of desire. She must allow for the possibility that she might become better educated, more cultivated, more sympathetic, less cynical, and so on. Similarly, ideology and world view conditions experience, and thus the ideal observer should compare all of the possible objects of desire from the point of view of “herself” but with any other world views or ideologies she could come to possess. For otherwise there will be no way to avoid the possibility that she might mistakenly overlook the possibility of undergoing a conversion, consciousness-raising, or paradigm-shifting experience that would then change the contents of her desires.

In order to know whether it would be good to have something that she cannot yet appreciate (but which she could come to appreciate), an ideal observer would have to determine whether she would desire it if her personality were different or if she were educated in some suitable way. She would have to experience the outcomes of all of her possible choices from the points of view of all the possible kinds of persons she might become, with all the possible kinds of psychologies she might have and all the ideologies and world-views she could adopt.[[6]](#footnote-6) Thus the ideal observer would consider the desires one would have under each of these possibilities and somehow select some of those desires as ideal.[[7]](#footnote-7) According to the ideal observer theory, well-being consists of the satisfaction of the desires selected by the ideal observer.

**WHO IS THE IDEAL OBSERVER?**

Clearly the ideal observer is a very special sort of critter. Perhaps too special. Many critics of the ideal-desire theory have asked: How could your well-being consist of the satisfaction of the desires of some almost godlike ideal being?[[8]](#footnote-8) The ideal observer was supposed to represent what a person would desire if she were observing from an ideal situation. But even if there is some counterpart of the actual person in some (maybe only barely imaginable) possible world who is ideal in the way the ideal observer is supposed to be, it is difficult to see how its desires could matter to the actual person whose well-being is at issue in the immediate way that her own well-being is supposed to matter to her. One of the main attractions of the desire theory of well-being was that it made it clear how well-being matters to us, and thus how it is connected to us. On the ideal observer view, we seem to have lost this connection.

Mark Johnson makes the point starkly: “Upon being told the difference between my evaluation and that of the ‘ideal’ observer, I may say, without showing any contempt for reason or value, ‘So what?’ That is, I may coherently take it that there is no reason to get into the ‘ideal’ condition and no reason to correct my valuations to accord with my beliefs about the evaluations I or another would make in those conditions.”[[9]](#footnote-9) I will attempt to formulate this worry somewhat more precisely in a bit. But first, we must consider a possible reply by the proponent of the ideal observer view to this line entire of attack.

The proponent of the ideal observer view might reply that this line of attack takes the language of the ideal observer far too literally. For, strictly speaking, the ideal observer is merely a metaphor for representing ideally corrected desires. It is merely a way to picture an aggregation of the preferences you would have if informed in various ways. Some subset of these desires, then, is the set of ideal desires.[[10]](#footnote-10) The ideal observer is just a picturesque way of representing the idea that non-mistaken desires are those that you would have if you had all of the experiences and education that you could have, and if you had them with each of the possible psychologies you could develop and ideologies you could adopt. Although this formulation is more precise, it does not avoid the original problem. Our personality traits, educational histories, ideologies, and experiences are a major part of who we are and what we care about. Even if we view the ideal observer merely as a metaphor, the theory still claims that a person's well-being is partly a function of the desires of imaginary persons quite different from her. And if the desire-based approach to well-being is forced to say this, then it has surrendered one of the main features that made it attractive.

**A DIAGNOSIS OF THE PROBLEM**

I suspect that the popularity of the ideal observer account of well-being derives in large part from the drawing (sometimes only implicitly) of a parallel between value theory and epistemology. It is tempting to think that just as there are epistemologically ideal conditions for believing, so there too there must be ideal conditions for desiring. The ideal conditions for desiring, so this line of thought goes, are those in which one has full information and complete experience, and has considered matters from all possible points of view. It is tempting to suppose that just as belief under ideal conditions might be thought to correspond to truth or knowledge or warranted assertability (or whatever other epistemic virtue belief is supposed to possess), so desire under ideal conditions might correspond to well-being.[[11]](#footnote-11) Once we begin thinking in terms of "correcting" desires that are “mistaken,” we have engaged in something that looks very much like an epistemological undertaking. Once we have succumbed to this tempting parallel, the drive toward excluding more and more kinds of mistaken desires from the set of those whose satisfaction constitutes well-being seems quite natural.

However, we should resist this parallel between value theory and epistemology. We should not correct or idealize desires as part of an epistemology in which ideally informed desires detect well-being in the way that ideal beliefs might detect truth (or other epistemic virtues). Rather, we must give some rationale for correcting desires within a value theory which is faithful to the fact that well-being is a property that is in a sense indexed to particular lives, a property in virtue of which a life is valuable to the particular person living it. For it is precisely here that the analogy between truth and well-being breaks down. Well-being is a property of individual persons and the lives they lead. It attaches to particular persons in a way that truth does not: what is true for you is true for me, but what makes you well-off does not necessarily make me well-off. Thus while epistemic conditions for truth may eradicate anything particular to a person’s point of view, a theory of well-being must not do this. For unlike truth, well-being is particular to the person whose well-being it is. In short, truth is universal, but well-being is individual.

**DO WE REALLY NEED HYPOTHETICAL DESIRES?**

Some philosophers think that desire-based accounts of well-being need not and should not go beyond actual desires: they should instead tie well-being only to desires the person actually has. The idea is that hypothetical desires are not really desires at all, and if the point of the desire theory was to connect our well-being to us via our desires, then hypothetical desires do not do this. Could moving back to an actual desire account offer a way out of the problems that the ideal observer view faces? Of course the challenge for any actual desire theory of well-being will be to tie well-being to actual desires while at the same time avoiding the apparently counterintuitive results that made the move to more and more idealized versions of the desire theory seem necessary in the first place. Can an actual desire theory of well-being get intuitively correct results in cases (like those of Cathy and Dave) in which someone fails to desire something that would clearly improve her life, simply because she either has false information about it, or entirely lacks information about it?[[12]](#footnote-12) The challenge for the actual desire theory will be to get the right answer in such cases without appealing to hypothetical desires.

Perhaps the actual desire theory can appeal to some general actualdesire of the agent in cases where a more specific desire seems to be mistaken.[[13]](#footnote-13) Suppose S desires X. Suppose that getting Y is tantamount to getting X (perhaps Y is an instance of X, or perhaps Y reliably causes X). Now suppose either that S does not know that getting Y is tantamount to getting X, or that S falsely believes something that causes her to desire not-Y. I take it that the correct thing to say here is that getting Y does in fact increase S’s well-being. We might simply say that Y increases S’s well-being because it satisfies S’s desire for X. So in Dave’s case, we explain the fact that the cat increases Dave’s well-being by referring to some actual desire Dave has, presumably some sort of desire for companionship. We then claim that a cat does in fact satisfy that (actual) desire and thus avoid positing any hypothetical desire. The cat improves Dave’s life not because he would want a cat if he were informed, but because it does satisfy his actual desire for companionship. In general, instead of saying (along with the proponent of a hypothetical desire theory) that

(H) My well-being is increased if I desire X and get either X or anything I would

desire if I had full information about what things are tantamount to getting X.

the proponent of an actual desire theory would say:

(A) My well-being is increased if I desire X and get X or anything that is

 tantamount to getting X.

The proponent of the actual desire approach would presumably claim that because notions of full information and hypothetical desires do not appear in this formulation, it might avoid all the problems of the ideal observer and full information theories.

But it is, I think, a false hope. For this “actual” desire formulation is not as different from the hypothetical version as it appears. The “actual” desire theorist claims that a desire for X is satisfied either by X or by anything that amounts to, causes, or implies X. But of course strictly speaking I may not really desire this other thing at all. To say that something would satisfy a desire I have is not quite the same as saying that I desire that thing. So while I might get something that “amounts to” the satisfaction of the actual contents of my desire, it is not the same thing as getting what I actually do desire.

This just follows from the fact that desires are intensional and that one cannot substitute identicals (or “tantamounts”) into intensional contexts. Suppose someone does not know that the morning star is Venus. Then for that person, believing one has seen the morning star is not the same as believing that one has seen Venus. Nor is desiring to see the morning star the same as desiring to see Venus. (“Why are you waking me up in the morning--I told you I wanted to see Venus!”) Of course as soon as she finds out that the morning star is Venus, the substitutions will generally hold. In the same way, for Dave (who lacks the belief that cats are good companions), a desire for companionship is not the same as a desire for a cat, even if having a cat is tantamount to having companionship.

If the actual desire account is to account for the plausible pre-theoretic intuition that getting a cat would contribute to Dave’s well-being, it must move away from the real, literal (de dicto) contents of the actual desires in much the same way that the hypothetical desire theorist does. In the case of the “actual” desire theory, though, this move can get hidden in the looseness of the definition of ‘satisfy’. While the hypothetical desire account talks of things that satisfy hypothetical desires that one does not have, the actual desire account is forced (if it is to be plausible) to loosen the definition of ‘satisfy’ so as to allow things that I do not, strictly speaking, actually desire to satisfy my actual desires.

 But once the actual desire account loosens the definition of ‘satisfies’ in this way, it opens the door to the same sorts of worries that moved us toward the ideal observer theory. Suppose Jane has no desire to go skiing. But suppose that if she were to try skiing, she’d find it so fun that she’d come to be an avid skier. Thus it seems correct to say both that skiing would satisfy her desire for fun if she tried it, and that--perhaps unbeknownst to her at the moment--a ski trip would contribute to her well-being.[[14]](#footnote-14) But if we allow the actual desire theory to count skiing--which she does not presently desire--as contributing to her well-being simply because it would satisfy her desire for fun (even though she does not now desire to go skiing), then we’ve started down a slope very much like the one that led us to the ideal desire theory. For if we will allow X to satisfy a desire for Y simply because X is an instance of Y, then we leave the door open to all sorts of conditions in which the desire for Y might be satisfied. Perhaps opera would satisfy my desire for good music if I were sufficiently educated to appreciate it. So, does opera-plus-being-educated-in-music-appreciation count as satisfying my desire? Presumably so. But if the proper education plus opera satisfy my desire for music, then why not simply say thatifI were educated, then opera would satisfy my desire for music?

Thus it appears that in order to get plausible results, the actual desire account must allow substitution of equivalents into the contents of actual desires in her account of satisfaction. But because the contents of desires are intensional, this cannot be an account of the satisfaction of the actual contents of those desires. Rather, it is a kind of hypothetical satisfaction: X would satisfy S’s desire for Y if S knew that X and Y were the same. Plausible versions of the actual desire account will have to appeal to this sort of hypothetical satisfaction of actual desires, whereas the hypothetical desire account talks about the satisfaction of a hypothetical desire. This difference, though, does not help the actual desire theorist avoid the problems associated with hypotheticals, for the hypotheticals can simply be rewritten in the notation of the actual desire theory.

The problems arise because there are lots of conditions--many of them rather exotic--under which desire satisfaction would occur. As soon as either theory claims that your desires can be satisfied by any condition other than by the obtaining of the literal (de dicto) content of an actual desire, we have opened the door to considering various exotic circumstances under which our desires count as being satisfied. But unless we open this door, we will have a theory that is so grounded in our actual, uninformed, and mistaken desires (with contents read in a literal, de dicto, way) that it will be implausible to think that well-being consists in their satisfaction.

So the “actual” desire account faces questions parallel to those that drive the hypothetical desire account toward the ideal observer theory. Both accounts must ask how exotic the circumstances can be that we will count as satisfying our desires. And the intuition that more education or information makes desires line up more closely with well-being pushes both accounts to count what would go on under ideal conditions as fulfilling desires. It seems to make little difference whether we ask this question by asking what conditions count as satisfying our actual desires, or by asking what hypothetical desires we should postulate.[[15]](#footnote-15)

Though I think that the actual desire account must ultimately fail to tie our well-being to our actual desires without any idealization whatsoever, I do think it is on the right track in holding that well-being must be grounded in the motivational structure of actual agents rather than that of some hypothetical ideal observer. Where I think the actual desire account goes wrong is in its assumption that the only way to do this is to link well-being to the satisfaction of actual desires. To see why this seemingly plausible assumption might be wrong, we have to consider the relation of an agent to her desires.

**INTEGRITY**

The central intuitionbehind the desire-based approach to well-being is that well-being is individual to the person whose well-being it is because it consists of the satisfaction of her desires. But what makes her desires her own? Why is it that she can identify with and care about her own desires in a direct, first-person way, a way in which she cannot care about the desires of someone else? This is a key question, for it goes to the heart of the question of whether, to what extent, and in what sense she can identify with and care about her actual desires but not with the hypothetical desires of the ideal observer. And it goes to the heart of the question about the nature of the connection between the person and her desires, the very question, in fact, that the desire-based approach to well-being exploits in order to connect a person to her well-being. One might be tempted simply to think that the answer is that her desires are in fact her own; it is just her ownership of them that accounts for her special relationship to them. But while there is surely something to this answer, it cannot be the whole story.For not all of my desires matter equally to me. Indeed, some of them I may even repudiate. Others, however, seem to be central to shaping me as the particular person I am, making me the person who is leading the particular life I am leading. And I suggest that the distinctive relation a person has toward some of her desires but not others is due to the fact that they bear different relationships to her self. I suggest that in order for desire-satisfaction to matter to a person--in the way that it must if the claim that well-being consists of desire-satisfaction is to be at all plausible--then those desires must reflect or at least be responsive to her deepest, most central concerns and projects. If this is so, then apparently the only desires whose satisfaction could constitute her well-being must be ones that she could have without changing those deepest concerns. The rest of the paper will be concerned with spelling out and defending this suggestion.

To begin, let us say that a hypothetical situation preserves a person’s integrity if and only if it preserves those concerns, attitudes, and other mental states that are constitutive of her identity or her self.[[16]](#footnote-16) The concept of integrity (as well as that of selfhood) to which I want to appeal is different from the concept of personal identity that is used to frame metaphysical questions about whether a person persists over time or after bizarre brain-transplant and teleportation events. Integrity is not primarily a matter of whether two spaciotemporal particulars count as temporal parts of the same person (although one might hold a theory that makes integrity necessary and sufficient for personal identity over time). Rather, integrity is a psychological notion meant to answer psychological questions about "who one is as a person," what her life is all about, what is important to her, and so on.[[17]](#footnote-17) Integrity involves the preservation of the personality, character, values and commitments that shape a human life and constitute a human self.

So while the psychological concept of integrity and the ontological concept of personal identity might be related, they are clearly not the same concept. To say that a change fails to preserve (ontological) personal identity is simply to say that two spaciotemporal particulars are not temporal parts of a single persisting entity. But to say that a change fails to preserve a person’s integrity is to say that it would involve such a complete and radical rejection of her value system, character, and convictions--her very identity, in fact--that she could no longer identify with the changed version of herself.

Thus it seems possible to imagine a course of events that would transform the numerically same physical entity as me into a Jesuit monk, or a neo-nazi. But these transformations would not preserve my integrity; they would change who I am as a person. Whether or not we decide that a new person has come into being, it is clear that, psychologically, I would be radically different from who I am. I cannot identify with the neo-nazi, for the concerns that make up his psychological identity are radically different from--and in fact radically opposed to--the concerns that make up my psychological identity. The things that he cares about are not the things that I care about. If I were to find that his change was in my future, I would find the news extremely distressing, for it would mean that I would come to repudiate those things that I now see as constitutive of the person that I am. I would feel extreme alienation from, and indeed disgust at my future self, whom I would see as morally, psychologically, and cognitively discontinuous from my current self.[[18]](#footnote-18)

Of course, some theories of ontological personal identity might claim that it reduces to psychological integrity. But my point here does not rely on this further metaphysical claim. Rather, my claim is simply that the person is psychologically, emotionally, mentally, morally, cognitively, evaluationally, and motivationally, alienated from the self that results from a radical integrity-destroying transformation. Though we do often talk about such alienation in terms that suggest literal discontinuities of personal identity--the reflective, “I’m very much a different person now,” or the dreaded “I don’t know who you are anymore”--nothing prevents us from seeing these as mere metaphors to give emphasis to claims about radical changes in our psychologies. I do not mean to rest my argument here on taking such remarks literally, and thus I distinguish between the psychological notion of integrity and the ontological notion of personal identity over time, even though one might believe that facts about the latter reduce to facts about the former.

I claim that what matters in questions about well-being is the preservation of the psychological constitution of a person’s identity (or, for short, her self), whether or not this is the only kind of persistence that matters for the metaphysical questions of personal identity. Questions about someone’s well-being must be asked from a perspective that she could occupy without loosing her self, her psychological constitution as the particular person she is.

**INTEGRITY AND THE IDEAL OBSERVER**

Persons must process information and organize their experience of the world in some way or another. But different ways of doing so will have different effects on both cognition and motivation. An ideal observer--if it is to truly represent maximal information--would need to consider every possible way that a person might come to organize information and experience the world. But since our own personalities, experiences, and moral, religious, and political commitments determine how we experience the world and organize information, an ideal observer would have to consider things from the point of view of someone with a personality, experiences, and commitments very different from our own. Thus, in order to represent maximally corrected desires, the ideal observer would have to consider how we would see things if we were very much other than we are.

Consider again my monk and neo-nazi logically possible future selves. Clearly both the monk and the neo-nazi would experience the world differently from how I do now. This means that how the monk and neo-nazi versions of me experience things, and how those experiences would affect “my” motivational structure are pieces of information relevant to knowing what a fully informed version of “me” would want. Therefore, on the ideal observer view, the desires of the monk and the neo-nazi versions of “me” are relevant to my well-being.[[19]](#footnote-19)

Clearly a radical transformation would be needed to "idealize" my preferences in such a way as to inform them with the knowledge of what I might want if I were to consider all of my options from the point of view of every kind of person I could conceivably become. I would have to take up the standpoint of every possible version of me--including those of the monk and the neo-nazi versions of “myself.” In addition to the full factual information I would have, I would also have to experience life from the point of view of a person with every possible ideology or world-view that I might have, or at least all of them that are consistent with full factual information. It is difficult to see that there would be enough left of me--of my self, of those convictions, commitments, values, and concerns that constitute my identity--for me to identify with the post-transformation “me” or to care about its preferences. Such idealized desires won’t matter to me in the direct way that my own most profound concerns matter to me.[[20]](#footnote-20)

The ideal observer theory requires that we satisfy the desires of a version of “me” that I cannot identify with since its identity is no longer constituted by the concerns, values, commitments, etc. that constitute my identity. For this reason, it will detach what is supposed to be “my” well-being from my identity as the particular person I am. It will make my well-being cease to matter to me by making it cease to be a function of the core values, concerns, and commitments make me who I am in a deep psychological (if not a metaphysical) sense.

The upshot of all this is that a desire-based theory of well-being should not abstract from the “center of psychological gravity” of the person whose well-being is at issue if it is to retain the attractive claim that a person’s well-being must have something to do with her, that it must somehow connect with the deepest concerns, values, projects, and commitments in terms of which her identity is constituted. Such a theory must, in short, respect the integrity of the person whose well-being it is. This might make things seem rather hopeless for such theories. For we seem to need to abstract away from a person’s actual desires in order to correct for mistakes; otherwise we get very counter-intuitive results. And yet, the move to correct desires seems to lead us to the ideal observer view, which severs well-being from the identity of the person whose well-being it is.

**THE WEB OF SELF**

I believe that it is possible to carve out a path between the horns of the present dilemma. The key to my suggestion is the observation that not all actual beliefs, desires, and so on seem equally important to integrity. To spell out and make good on this claim, I will need to offer at sketch of a theory of the self. To that project I now turn.

I suggest that we think of human psychology as an interlocking network of beliefs, desires, and other attitudes. It seems clear, both from philosophical work like Quine's and from work in cognitive psychology, that beliefs can be represented as a sort of network.[[21]](#footnote-21) At each nexus is some belief, its location determined by the relations it bears to other beliefs. These relations among beliefs can be represented as strands in a web that links the various beliefs together. As Quine observes, some beliefs are closer to the center of the "web of belief" than others. Some of these core beliefs are simple logical truths and truths of basic arithmetic.[[22]](#footnote-22) Others are convictions according to which a person lives her life. These convictions are most commonly religious or political in nature, though they also commonly include scientific, sociological, and metaphysical claims as well. A person's life and personality are typically organized around her convictions; they form the foundation of her world-view. Changing a core conviction might change the person's cognitive structure almost as radically as changing her belief that 2 + 2 = 4 or that cause precedes effect.

A person’s core beliefs and commitments are the foundation for the rest of her cognitive structure, the lens through which she sees the rest of the world. They are the basic organizing principles that the person uses to interpret and organize all other incoming information. A change to them constitutes a profound paradigm shift--a conversion of sorts--that produces a radical cognitive discontinuity between the pre- and post-change selves. Such a change is more than a mere change in beliefs, but a change in the foundational cognitive structure with which the person organizes her mental life. Ideological beliefs sometimes form cognitive cores of this sort as well. A core commitment to Marxism or feminism will lead a person to organize and interpret other beliefs and information far differently from someone whose cognitive economy rests on a foundational commitment to a patriarchal or capitalist ideology.

A similar picture can be drawn for human motivation. Evidence from motivational psychology strongly suggests that a few core desires are the foundation for most if not all of our motivation.[[23]](#footnote-23) It appears that the structure of motivation is something like a web of desire, with core desires, along with beliefs, supporting layers of more peripheral desires. The contents of the core desires will usually be very general. Typically their objects are not particular actions, but states of affairs that can be attained in many possible ways. Means-end beliefs contribute to the production of more particular desires, ultimately desires to perform particular actions. In this way, the core motivations are the foundations for other motives.

Some core desires--for such things as food, love, self-esteem, achievement, and pleasure--are probably hardwired into all of us. Others are particular to individuals. These constitute our central projects and core values. Core values of honesty, fairness, loyalty, nurturing, and so forth are the bases for whole sets of other, more particular preferences. Similar remarks apply to what Bernard Williams calls "projects”: such things as political goals, raising families, pursuing truth, fighting oppression, and self-development of various kinds.[[24]](#footnote-24) Such projects are often core organizing principles in a person's life. Thus central projects, core values, and hard-wired basic motives are centers around which webs of other motivation are woven. They define who a person is; they make up her psychological identity and thus determine who she is and what her life is about.

Indeed, we can understand the distinction that some philosophers make between a “mere” desire and a value as a distinction between core motives that are constitutive of the self and motives which are more peripheral and thus not constitutive of the self.[[25]](#footnote-25) A core motive is one that I cannot contemplate changing without simultaneously contemplating changing my very identity. It is a motive to which I have committed myself by building much of the rest of my motivational structure around it. It is a motive that I have taken as an ultimate end, as a foundation for other ends rather than as a means to some other, more foundational end. It is an end that is thus at the core of my motivational and evaluative structure. When I engage in critical reflection on my current motives, these values supply much of the material for making decisions about which ones to endorse and which to repudiate.

In fact, I think that the higher order desires that many writers on autonomy mention are best seen as getting their content from core desires.[[26]](#footnote-26) That is, the endorsement of a desire by a higher order desire will typically reflect one’s core desires. That would make sense of the common claim that higher order desires are somehow more authentically part of the self than lower order desires. Nevertheless, there is no guarantee that higher-order desires (in the logical sense of desires that have other desires as their objects) and core desires will always coincide. On the view I am suggesting, when such divergences between higher order desires and core desires occur, it is the latter rather than the former that are constitutive of who the person is.[[27]](#footnote-27)

**INTEGRITY AND MODERATE IDEALIZATION**

Attitudes at the center of these webs constitute the self. They affect, are connected to, and support the rest of the belief-desire system. They give meaning to a person's life and constitute the self or the psychological identity of the person. They matter to us in the most central way possible, for they make us who we are. Indeed, part of what makes them central in the web of the self to whom they belong is that they matter so crucially to that person; the rest of the self has been spun around it, so to speak. The attitudes at the core of the self constitute the foundations of our attempts to make sense of the world, and to define what we stand for, what our purpose is, and what is important to us. They form the cognitive and evaluative perspectives that determine what matters to us. They are literally at the psychological core of our very being.[[28]](#footnote-28)

Attitudes at the periphery, on the other hand, do not constitute the self. They include things like one’s perceptual beliefs and one’s current motivational state, whims, and so on. They are certainly affected by the core, but the influence is mostly one-way. When they change--as they frequently do--those attitudes at the core of the self, in terms of which the self is constituted, do not also change. This means that quite a lot of change can occur within a person’s psychological economy without her self being affected. Consequently, many beliefs and desires can be idealized without any damage to a person’s integrity.

This means that we need not give up all idealization, all appeals to hypothetical desires. Some changes are compatible with the preservation of integrity, and so some correction can be allowed by moderate, integrity-respecting idealization. Beliefs and desires that are closer to the periphery of the web of self are not necessary for the preservation of integrity, for they do not define who the person is. They can and should be idealized to determine what affects a person’s well-being.

Idealizations of this moderate sort also leave enough of our personality and our motivational and cognitive structure intact for there to be determinate answers to questions about what we would desire if our desires were corrected. If we idealize too far--asking what I would want had I become a Jesuit monk or a neo-nazi, for example--there may be no determinate answer as to what I would want. The changes necessary to effect such a transformation would be so radical that it's not clear how much of my present psychological make-up would be left to make decisions about what the post-transformation "me" would want.[[29]](#footnote-29) Trying to figure out what the post-transformation "I" would want would be rather like trying to guess what I would want if I had a lobotomy: how can we be sure which parts of me would be left intact? But if we limit our corrections to the periphery of the web of self, we leave intact enough of the personality to meaningfully answer questions about what the person would want if her desires were corrected. If we leave the core motivational structure largely intact, we will be able to decide what peripheral desires a person would have if her peripheral beliefs were correct. In this way, her well-being will be a function of the actual desires that lie at the core of her identity.

And this observation brings us to the issue of the proper rationale for correction in a desire-based theory of well-being. On the view I am suggesting, the purpose of invoking idealized desires is not to offer an epistemological theory about accurate detection of value. Rather, it is a way of bringing well-being into line with the core desires of the particular person whose well-being it is. If the person’s (psychological if not metaphysical) identity is partly constituted by core commitments, convictions, projects, and values, then limiting correction to the periphery of the web of self makes well-being more responsive to the actual person, and to the core of the psychological constitution of her self. Idealization of peripheral desires (and the peripheral beliefs on which they are often based) brings well-being closer to the objects of our most central and important actual desires than do our actual peripheral desires.

This position also combines the compelling Kantian insight that persons are ends in themselves with the plausible claim that persons are at least partly constituted by their convictions, commitments, and characters. It does this by claiming that because persons are at least partly constituted by their ends, to give persons pride of place in moral theory requires that we also give pride of place to the ends, values, and convictions by which they are at least partly constituted. But because desires and beliefs are not equally constitutive of the identity of the persons to whom we owe respect, we can draw moral distinctions among them and assign greater moral weight to those that are more constitutive of the identities of the persons to whom they belong.

**TWO NOTIONS OF THE GOOD LIFE**

On the view I am advocating, questions about a person’s well-being must be asked from within the evaluative perspective that helps to constitute who she is. In this way well-being can in fact be a property of the psychological entity who is actually living the particular life that she is in fact living. For some purposes in moral and political theory, this (rather narrow) notion of well-being seems to be just what we want. Thus for instance, one might want to argue that the operative notion of happiness in utilitarianism is the narrow notion of well-being that is tied to the evaluative perspective that is partly constitutive of particular persons. Or one might want to argue that the duty of benevolence should aim at increasing others' well-being, construed in this way. Or perhaps liberal political theorists will need to appeal only to well-being in order to measure and manipulate levels of welfare without adopting a full-blown state-sanctioned evaluative perspective.

But of course we do not confine even our first-person prudential deliberations to questions about what courses of action would produce the most well-being. We sometimes ask questions about what kind of person to become, what sorts of evaluative perspectives to take up.[[30]](#footnote-30) The difference between the two sorts of questions is rather like the difference between asking what move to make in a game and asking what game to play. If one is playing chess, knowing what move would be good if one were playing checkers is irrelevant to deciding what is a good chess move. One's goals qua chess-player are defined relative to the particular game of chess. This is parallel to the question of what makes a particular life go well. It must be asked from within the life. But there are other questions, about games and about life. One might quite sensibly ask what game to play. One does not ask that question from within the standpoint of any particular game, for the question is about what game to take up. Similarly, some questions about the good life are not tied to the standpoint of a particular life; they are questions about what kind of life is best.

Often we make important decisions about how to live on grounds other than our potential well-being. Should I become an underpaid social worker or a wealthy drug dealer? A stockbroker thriving on Wall Street, or a frustrated educator fighting a losing battle against intellectual mediocrity? A complacent bigot in a racist society or a progressive reformer living a life of conflict and turmoil? The amount of well-being in each particular life is only one consideration in such a choice. Although well-being matters, it is not the only thing that matters. Often one choice will lead to more well-being, but another will be more worthwhile or virtuous or more or morally good. One life may be happier in terms of how the life goes to the person who lives it, but the other is a better--from a more objective point of view--life to lead, even if it is a less (subjectively) happy life. Thus we do seem to ask questions about what kind of life is best that are not purely questions about subjective well-being.

So although we should separate the notion of a valuable self from that of a valuable life for a given self, we seem to need both notions. If we want to decide whether a particular action or policy will make a particular person better off, then we need the narrow, more subjective notion of well-being. To decide which among two wildly different lives to lead, or two very different directions that one might develop one’s character, we will typically appeal to moral ideals, and perhaps to what we might call “aesthetic” ideals like Aristotle's notion of a virtuous life worthy of being lived.[[31]](#footnote-31)

**WELL-BEING, THE GOOD LIFE, AND VIRTUE**

Given that we are interested in what makes the particular person's life go well for her, my strategy has been to decide how different a person can be while retaining the psychological integrity necessary for her to be able to identify with the hypothetical person in such a way that the desires of the hypothetical person can matter to the actual person. The trick, of course, is to figure out exactly how much of a change is compatible with the preservation of a person’s integrity. If this is indeed the correct strategy, then any plausible desire-based theory of well-being will have to appeal to a theory of the integrity and the self.

I have just offered such a view, but it has an odd consequence when plugged into the desire-based theory of value, and I want to say a bit about that consequence now. On the view I am suggesting, there will likely be certain changes in which it will be unclear whether integrity has been preserved. For on my view, integrity-preservation is a matter of retaining convictions, traits, and commitments that are at the nexus of the causal network of psychological states that forms the core of the self. But there is likely to be no strict division--no bright line--between the core and the periphery of the web of self, and so hypotheticals will vary in the *degrees to which* they preserve integrity.[[32]](#footnote-32) This lack of a clear division between hypotheticals that preserve integrity and those that do not might have interesting implications.

In particular, it seems to imply that questions about well-being will gradually shade into questions about the good that appeal to more objective moral or aesthetic criteria. This shading will occur as we contemplate greater and greater changes in our attitudes, and as we consider changing attitudes that are closer and closer to the core of the web of psychological states that constitutes our identity. Such questions blur the boundaries between questions about (subjective) well-being and questions about moral or aesthetic value of a life, as judged from a point of view at least partially external to it. They involve the possibility of partially changing who one is, by partially changing one's evaluative perspective. They are questions about what kind of person to be, about how to plan not only one’s life, but the evolution of one’s identity and character. They require some transcendence of our current situation and perspective. In particular, they require transcendence of part but not all of our evaluative framework.

To see how these different questions can blend together, consider an example suggested by Connie Rosati.[[33]](#footnote-33) It involves a man, whom I’ll call Randy, who has a core commitment to the principles of Ayn Rand: “He . . . has memorized The Virtue of Selfishness. He fervently believes in the tenets of objectivism and has organized his life around living in accordance with his creed.” But, continues Rosati, “he is not entirely happy. For one thing, he has difficulty making and keeping friends. Others, including members of his own family, find him--to his puzzlement--rude and dogmatic. He would very much like to live a life that is less isolated.” Randy finds that his commitment to his core principles prevents him from realizing certain values--friendship, for example. So it appears that friendship, which would presumably contribute to Randy's well-being, is unavailable to him because of his core commitments. Indeed, he seems to be made worse off by the satisfaction of desires emanates from his core commitment to “the virtue of selfishness.” What should we say about Randy? As Rosati tells the story, it is Randy who craves friendship and sees his life as impoverished without it. It appears, then, that Randy holds two incompatible core values. There are two things that would improve Randy's life, but Randy cannot do both. Tragic, perhaps, but then no one ever guaranteed that all goods could be realized at once.

If Randy values friendship as a core commitment--perhaps one that he is only now becoming aware of--then his very self is deeply conflicted.[[34]](#footnote-34) If the conflict becomes too great, Randy may have to consider giving up one of the two core values. To find out whether to make the change, he must ask which of his incompatible core values to keep: Ayn Rand or friendship. Questions like this are quite unlike ordinary questions about what will make us better off, for they can only be asked by at the same time asking whether to at least partly change who we are. Notice, however, that even a choice to make such a transformation is not completely detached from Randy’s evaluative framework. A question of this sort does partly transcend one’s evaluative standpoint, for it is a question about that standpoint. A person commonly asks such a question when she notices that her evaluative standpoint is inconsistent and wonders what to do about it. If it remains a question about how best to fit the bulk of one’s core commitments and values into a coherent evaluative perspective, then it is still a question that requires only a partial transcendence of one’s perspective. Thus it lies in between questions about what improves a particular self and questions about what self to be. It is a question that asks how best to be the kind of person that can coherently maximize most of the values that make up one’s self.

Such inquiries can be fairly radical: they might well tell us to make significant changes to who we are. But just as the Neurath's ship must be reconstructed while it is sailing, our lives must often be made and remade while we live them. In such cases, there will be a large amount of the core left in tact, and indeed, the part of the core that is not being questioned will typically supply the evaluative criteria for deciding whether to make the change. Such questions require a partial transcendence of the evaluative perspective that helps to constitute the self and make us who we are. The fact that we ask--and often answer--them shows that some questions about the good transcend the limits of our own integrity.[[35]](#footnote-35)

1. . See Shelly Kagan, "The Limits of Well-Being," Social Philosophy and Policy (1992): 169-89, esp. p. 185; David Lewis, "Dispositional Theories of Value" (Proceedings of the Aristotelian Society Supplementary Volume 63 [1989]); Connie Rosati, "Persons, Perspectives, and Full Information Accounts of the Good" (Ethics 105 [1995]: 296-325); and David Sobel, "Full Information Accounts of Well-Being (Ethics 104 [1994]: 784-810). See also James Griffin, Well-Being: Its Meaning, Measurement, and Moral Importance (Oxford: Clarendon, 1986); Derek Parfit, Reasons and Persons (Oxford: Clarendon, 1984), appendix I; Henry Sidgwick, The Methods of Ethics, Indianapolis: Hackett, 1981 [orig. pub. by Macmillan & Co., 1907), pp. 111-12) and Richard Brandt, A Theory of the Good and the Right (Oxford: Oxford University Press, 1979), pp. 110-129. [↑](#footnote-ref-1)
2. . My purpose here is not to defend this intuition, but rather to examine certain problems that arise for theories that embrace it. It is worth pointing out that holding a subjective conception of well-being does not in any obvious way commit one to a more thoroughgoing subjectivism about rationality or morality. Nor does it commit one to begging the question against those who, like Aristotle, claim that there is more to the good life than subjective satisfaction. Well-being should be seen as a quasi-technical term referring to the subjective component of the individual’s good: how her life goes from her own point of view. There may or may not be an additional, objective component of an individual’s good as well. [↑](#footnote-ref-2)
3. . See Peter Railton, "Facts and Values" (Philosophical Topics 14 [1986]: 5-29); and "Moral Realism" (Philosophical Review 95 [1986]: 163-207); and David Lewis "Dispositional Theories of Value" (op. cit.), pp. 122-6. [↑](#footnote-ref-3)
4. . See Rosati, pp. 314-15; Norman Daniels, "Can Cognitive Psychotherapy Reconcile Reason and Desire?" (esp. p. 779-782); David Lewis, op. cit., 126; Susan Babbit, op. cit.; Sobel, p. 796. [↑](#footnote-ref-4)
5. . Perhaps somewhat ironically, the full-blown ideal observer account of well-being, as distinct from the full information account, is to some degree a creation of critics of the desire-based approach to well-being. The most notable of such critics is probably Connie Rosati. In “Persons, Perspectives, and Full Information Accounts of the Good,” she argues very persuasively that the pressures which push philosophers toward the full information version of the desire-based account in fact push all the way to an ideal observer account (she calls it an “Ideal Adviser account”). She also provides an excellent description of the kind of ideal observer account that would be necessary to respond to the pressures that motivate the move to the full information version of the desire-based account. My discussion of the ideal observer view is heavily indebted to both of these aspects of Rosati’s excellent article. [↑](#footnote-ref-5)
6. . See Rosati, pp. 315-17. The idea here is presumably an elaboration of Mill’s insight that only someone who has known both pig-pleasure and Socrates-pleasure would be in a position to know which is better. That is, only someone who had experience various options from all possible points of view and perspectives would be in an ideal position to know which options are most desirable. [↑](#footnote-ref-6)
7. . Rosati (op. cit.) wonders how the ideal observer is supposed to select some set of desires as "ideal" from among the desires of all the various possible persons. I think that the most sensible answer is in terms of the strength various desires would have in an ideal observer. If the ideal desire theory really does embody Mill’s insight about Socrates and the pig, then surely the idea is that desire for A is more ideal than a desire for B if an ideal observer would prefer A to B. Thus it might be better to talk about ideal preference orderings rather than ideal desires. [↑](#footnote-ref-7)
8. . See the works by Rosati, Sobel, and Johnson cited above, as well as a recent paper by Don Loeb (“Full-Information Theories of Individual Good,” -------- ). [↑](#footnote-ref-8)
9. . Mark Johnson, “Dispostional Theories of Value,” p. 157. [↑](#footnote-ref-9)
10. . More precisely, the ideal observer is a metaphor for the strongest preferences in a set of all the fully informed preferences from each possible psychology a person could have. [↑](#footnote-ref-10)
11. . Mark Johnson, "Dispositional Theories of Value," (A symposium in Proceedings of the Aristotelian Society Supplementary Volume 63 [1989]), Arthur Ripstein, "Preference" (in R. G. Frey and Christopher Morris, ed., Value, Welfare, and Morality [Oxford: Cambridge University Press, 1993]); and D. Goldstick, "Interests" (paper presented at the 1996 meeting of the CPA) explicitly draw parallels of this general sort; David Lewis (p. 21) and Peter Railton ("Moral Realism," p. 175) succumb to this temptation as well. [↑](#footnote-ref-11)
12. . Cases in which we have a mistaken actual desire (e.g., one that is based on a false belief) pose no problem for the actual desire theory. For claiming that well-being is a function of the satisfaction of actual desires does not imply that the satisfaction of every actual desire contributes to well-being. [↑](#footnote-ref-12)
13. . I thank an anonymous reader for pressing this point and for suggesting how such an actual desire theory might attempt to deal with cases like Dave and Cathy’s. An audience member at a presentation of an earlier version pressed a similar point. Mark Murphy pushes this line of thought in “The Simple Desire-Fulfillment Theory” (unpublished manuscript). [↑](#footnote-ref-13)
14. . Such a fact might make a ski trip an excellent gift, for example. Indeed, much of the institution of gift-giving--especially when the recipient is relatively affluent--assumes that a person might not know that certain things will in fact contribute to her well-being. Such things, of course, make the best gifts. [↑](#footnote-ref-14)
15. . Suppose I desire a minor political career. If I were to convert to an ideology of extreme racism, become a KKK member, and rise to the upper echelons, then I would in fact have a minor political career. Now, the principle that a desire is satisfied if you get anything that is equivalent or tantamount to the thing you actually wanted will imply that the KKK career does in fact satisfy the desire for a minor political career. So it appears that the actual desire theory is committed to saying that converting to a racist ideology and then becoming Grand Dragon of the KKK increases the well-being of anyone with minor political aspirations. This intolerable result follows directly from the principle A above, which the actual desire account needs if it is to be at all plausible. Of course one way to resist this conclusion would be to say that only certain forms of hypothetical satisfaction of actual desires count--namely those in which the actual desires would be satisfied under conditions that we would regard as improvements. This very tempting strategy would put the actual desire theory squarely on the road to a version of the ideal observer theory. [↑](#footnote-ref-15)
16. . The term `integrity' also (intentionally) evokes some ideas and themes in the work of Bernard Williams. See "Persons, Character, and Morality" (in Moral Luck [Cambridge: Cambridge University Press, 1981]). [↑](#footnote-ref-16)
17. . I borrow this distinction--and much of my thinking about it--from the work of Marya Schechtman. See The Constitution of Selves (Ithaca: Cornell University Press, 1996). [↑](#footnote-ref-17)
18. . Such cases raise interesting questions: Should we provide for a future self whom we now repudiate? Presumably a future self has at least some claim on our concern merely because it is in fact our future self. Yet we surely also have an interest in continuing to be the kind of person who is committed to the values and projects to which we are now committed. An important round of this debate is found in Derek Parfit, “Later Selves and Moral Principles” (in Montefiore, ed., *Philosophy and Personal Relations*) and Bernard Williams, “Persons, Character, and Morality.” Distinguishing between the psychological notion of integrity and ontological notion of personal identity might help us to at least clarify the conflicting considerations. [↑](#footnote-ref-18)
19. . Remember that we apparently have to let the ideal observer take into account every possibility that I could actualize, for there is no non-circular way to select some possibilities that count in determining well-being and others (presumably that of the neo-nazi, at least), that do not count. [↑](#footnote-ref-19)
20. . This is not to say that I might not change our profoundest concerns on finding out what concerns and ideal version of “me” would want. But such a decision is in a sense (to be discussed below) more a decision about what person to be than about what makes my life go well for me given the person I am. [↑](#footnote-ref-20)
21. . See Word and Object (Cambridge: MIT Press, 1960), pp. 9-13; "Two Dogmas of Empiricism" in From a Logical Point of View (New York: Harper 1953), pp. 42-46; Neil Stillings, et al., Cognitive Science: An Introduction (Cambridge: MIT Press, 19--), esp. pp. 26-30, 73-86, and 142-67; J. R. Anderson, The Architecture of Cognition (Cambridge: Harvard University Press, 1983) and Human Associative Memory (Hillsdale, N.J.: Erlbaum, 1972). [↑](#footnote-ref-21)
22. . Though I follow Quine in speaking of a web of belief, I do not mean to commit myself to anything more Quinean than the claim that beliefs are parts of complex networks. [↑](#footnote-ref-22)
23. . See John Atkinson and David Birch, David, An Introduction to Motivation (New York: Van Norstrand, 1978); Robert Audi, "The Structure of Motivation" (Pacific Philosophical Quarterly 61 [1980]: 258-75); Heinz Heckhausen, Motivation and Action (New York: Springer-Verlag, 1991); Frederick Toates, Motivational Systems (Cambridge: Cambridge University Press, 1986); and Aaron Sloman, "Motives, Mechanisms, and Emotions" (Cognition and Emotion 1 (1987): 217‑33). An early version of this idea is Maslow’s Motivation and Personality (New York: Harper and Row, 1970). [↑](#footnote-ref-23)
24. . See Williams, "Persons, Character and Morality" and "A Critique of Utilitarianism" (in J.J.C Smart and Bernard Williams, Utilitarianism: For and Against, Cambridge: Cambridge University Press, 1963), pp. 108-118; and Norvin Richards, "A Conception of Personality" (Behaviorism 14 [1986]: 147-57). [↑](#footnote-ref-24)
25. . See, for example, Gary Watson (“Free Agency” [Journal of Philosophy 72 (1975)]). [↑](#footnote-ref-25)
26. . Classic examples of such theories of autonomy can be found in Harry Frankfurt, “Freedom of the Will and the Concept of a Person” (The Journal of Philosophy 68 [1971]:5-20); and Gerald Dworkin, The Theory and Practice of Autonomy, Cambridge: Cambridge University Press, 1988. [↑](#footnote-ref-26)
27. . Proponents of higher order desire theories of autonomy will, no doubt, disagree with me here, but this is not the place to settle this particular issue. My point here is twofold. First, there is a connection between the topic of this paper and certain issues in the literature on autonomy, namely, the idea of some desires being more “truly one’s own” than others. Second, that to the extent that higher order desires do give us (what I think are) the right answers about what is truly one’s own, it is simply because they will probably tend to embody values that are more central than the motivations that they endorse or repudiate. I develop and defend an alternative to the higher order desire view of autonomy in “Autonomy, Value, and Conditioned Desire” (American Philosophical Quarterly 32 (1995): 57-69). There I suggest that heteronomous desires arise from informational states that are causally and informationally isolated from normal beliefs. One result of this isolation is that they are maximally peripheral in the sense defined here. [↑](#footnote-ref-27)
28. . Indeed, this fact means that even if the ideal observer theory were correct, it would likely be motivationally inert. For if one has a core commitment to some value, then one will typically be unable to be convinced that an ideal observer would repudiate it. This is simply because holding a value is parallel to believing something. To believe that P is to believe that P is true. Similarly, to hold V as a value is to hold that V is indeed valuable. A person who believes that P will also believe that an ideal observer will believe that P, and similarly someone who holds V will believe that an ideal observer would hold V as well. [↑](#footnote-ref-28)
29. . David Velleman makes this point in "Brandt's Definition of `Good' (Philosophical Review 97 [1988]: 353-71, p. 368-9, n. 15). [↑](#footnote-ref-29)
30. . Marilyn Friedman offers many useful insights about the role that friends play in helping us to ask such questions. See her What Are Friends For (Ithaca: Cornell University Press, 1993). [↑](#footnote-ref-30)
31. . See Richard Kraut, "Two Conceptions of Happiness" (Philosophical Review 88 [1979]: 167-97) and Griffin, op. cit., pp. 56-72 and 127-162. The term ‘aesthetic’ is meant to capture the admiration that we feel toward various excellences, beyond any consideration of their usefulness to society. [↑](#footnote-ref-31)
32. . There might be other sources of indeterminacy as well. For there will almost certainly be more than one distinct core, and calculations of centrality may turn out not to be perfectly transitive. [↑](#footnote-ref-32)
33. . In her reply to the version of this paper presented at the Central Division meetings of the American Philosophical Association, April 1996. [↑](#footnote-ref-33)
34. . Perhaps most evil people are actually conflicted, and thus have some good core desires will be frustrated when their evil desires are satisfied. But our theories of well-being should allow that evil people might genuinely flourish by satisfying evil desires. Indeed, that is part of what makes them evil. [↑](#footnote-ref-34)
35. . Early versions of this paper were presented at meetings of the Florida Philosophical Association (October, 1995) and the American Philosophical Association, Central Division (April, 1996). My thanks to audiences at both presentations. Special thanks are due to Connie Rosati, who commented on the APA presentation, for her insightful remarks. Comments and criticisms by anonymous readers have also been helpful, and I thank them. Finally, I thank Sam Black, Don Brown, and Colin McCleod for fruitful comments and discussion of the penultimate draft. [↑](#footnote-ref-35)