PREFERENTISM AND THE PARADOX OF DESIRE

BY BRADFORD SKOW

JOURNAL OF ETHICS & SOCIAL PHILOSOPHY

VOL. 3, NO. 3 | SEPTEMBER 2009
URL: WWW.JESP.ORG

Preferentism and the Paradox of Desire

Bradford Skow

1. The Paradox Stated

ACTUALIST PREFERENTISM IS A THEORY of welfare: a theory that says what it is for someone's life to go well for her. The theory's basic idea is that getting what one wants makes one's life go better. Like any theory of welfare, this one faces problem cases: cases in which someone's desires are satisfied but, intuitively, they are not made better off (or vice versa). But in addition to these problem cases, preferentism faces *the paradox of desire*. In a nutshell, this objection to preferentism goes like this: I can certainly desire to be badly off. But if a desire-satisfaction theory of welfare is true, then – under certain assumptions – the hypothesis that I desire to be badly off entails a contradiction. So desire-satisfaction theories of welfare are false.[1]

But this argument does not, in fact, establish that preferentism is false. There is a way to formulate preferentism so that the hypothesis that I desire to be badly off does not entail a contradiction. My aim is to show how this version of preferentism avoids paradox. Before I proceed, though, I need to state the version of preferentism that is the target of the argument, and spell out the argument in more detail. I will start with the first task.

On the standard atomistic version of actualist preferentism, the "atoms" of welfare are episodes of intrinsic desire satisfaction: episodes (stretches of time) during which the subject has a intrinsic desire that *P*, and it is in fact the case that *P*.[2] (An intrinsic, or non-instrumental, desire is a desire that one does not have merely because satisfying it is a means to satisfying some other desire.) "Episode of intrinsic desire frustration" is defined similarly. The intrinsic value of an episode of desire satisfaction (or frustration) is equal to the intensity of the desire times the duration of the episode. Desire satisfaction has "positive" value and desire frustration has "negative" value, so the value

[1] To my knowledge, the first appearance of this argument in print is in Feldman (2004). It is also discussed in Heathwood (2005) and Bradley (2007). (I follow Heathwood in calling it "The paradox of desire," but perhaps it should instead be called "Feldman's Paradox.") Bradley discusses several responses to the paradox, but does not discuss the one I will present.

[2] Suppose that I now want (intrinsically) to have chocolate cake on my next birthday. And suppose that, in fact, I will get chocolate cake on my next birthday, but when I get the cake I will no longer want it. Preferentism, as I have stated it, entails that my current desire is currently satisfied, and so that I am currently living through an episode of intrinsic desire satisfaction. Some preferentists may prefer a theory that does not allow this kind of episode to increase my level of welfare. Since the paradox of desire arises either way, I focus my discussion on the simpler version of the theory. (Derek Parfit (1984: section 59) discusses something like this problem.)

of someone's life for her is just the sum of the values of the episodes of satisfaction, minus the sum of the values of the episodes of frustration.

For our purposes, more important than how the theory calculates the welfare level of someone's entire life is how the theory calculates someone's welfare level at a particular time. Here is what the theory says: the value of someone's life for her at a time *t* is just the net amount of desire satisfaction that occurs in her life at *t*. Since *t* is just one instant, we can ignore the durations of the episodes of desire satisfaction and frustration. So the value of someone's life for her at *t* is just the sum of the intensities of the satisfied desires she has at *t*, minus the sum of the intensities of the frustrated desires she has at *t*.

(Actualist preferentism is not the only form of preferentism. Ideal preferentism says, roughly, that it is the satisfaction of the intrinsic desires you would have, if you were to undergo some form of "cognitive psychotherapy" (you were thinking more clearly, you knew all the relevant facts...), that contributes to your welfare. While many people accept some form of preferentism, ideal preferentism is probably more popular than actualist preferentism. (See Kagan (1998: 38) for a brief survey of the reasons.) And it looks like the paradox does not arise for ideal preferentism: it might be that if I were to undergo cognitive psychotherapy, I would not desire to be badly off. But I agree with Heathwood (2005) that none of the standard arguments for preferring ideal to actualist preferentism are any good. Since actualist preferentism is the better theory, defending it against the paradox of desire is all the more urgent.)

Now to present the paradox. The paradox of desire arises in the following kind of situation. Suppose I have several intrinsic first-order desires at *t* – like a desire for a cold beer, a desire for some salty peanuts, a desire for warm weather. (They are "first-order" because they are not desires about what desires I have, and are not desires about my level of well-being.) Suppose that not many of those desires are satisfied. So, to have numbers to work with, say that when those desires are considered alone, the net amount of desire satisfaction for me at *t* is -6.

Now suppose that in addition I have another intrinsic desire at *t*: a desire to be badly off, a desire that my welfare level at *t* be negative. Call this desire "*D*." This desire and my first-order desires are all the desires I have at *t*. Say that *D* has intensity 10. We get a contradiction. Proof: either *D* is satisfied, or not. First suppose that *D* is satisfied. Then my welfare level at *t* is $10-6=4$, a positive number. But then *D* is not satisfied, a contradiction. Now suppose that *D* is not satisfied (that is, that it is frustrated). Then my welfare level at *t* is $-10-6=-16$, a negative number. So *D* is satisfied, a contradiction.

The argument against actualist preferentism here is straightforward. It has just two premises:

(P1) The scenario I described, in which someone has a certain set of desires, including the desire to be badly off, is a possible scenario.
(P2) But if actualist preferentism is true, then that scenario is impossible.
(C) Therefore, actualist preferentism is false.

The derivation of the contradiction I just gave is the support for (P2). And (P1) seems obviously true. Surely someone could have that collection of desires. There are, of course, philosophical views that are incompatible with (P1). According to psychological egoism, for example, the only intrinsic desires anyone ever has are desires for their own welfare – desires that their welfare be positive. So if psychological egoism is true, then (P1) is false. But I do not think that psychological egoism is true.[3] (I will discuss another attempt to reject (P1) in the next section.)

Intuitively, what generates the problem for preferentism is this. Preferentism gives us an equation:

my level of welfare at *t* = the "net amount" of desire satisfaction I enjoy at *t*.

But whether or not the second-order desire *D* is satisfied at *t* depends on my welfare level at *t*. It generates a kind of negative feedback in the equation: positive numbers (and zero) on the left-hand side force negative numbers on the right-hand side, and negative numbers on the left force positive numbers on the right, so that the equation cannot be correct.

What is a preferentist to do? My solution is to revise preferentism in a way that makes (P2) false. But it is worth discussing in more detail the plausibility of rejecting (P1). For if there is a way to motivate rejecting (P1), then the revisions I propose to preferentism are not necessary.

2. Rejecting the First Premise

Heathwood (2005) discusses the paradox of desire, and he notes that a similar paradox arises that makes no mention of preferentism. Suppose that my only desire is that none of my desires be satisfied. Or suppose that I desire that this very desire not be satisfied. Then my desire is satisfied if and only if it is not satisfied – which is absurd.

This paradox is easy to solve: deny that it is possible that someone have either of those desires. It is easy to motivate this solution. No one can have either of those desires because the hypothesis that someone does have one of them leads to a contradiction.

[3] Lots of arguments against psychological egoism may be found in, for example, part II of Sober and Wilson (1998).

For those who accept a certain account of what desire is,[4] there is a similar but more illuminating way to motivate this response. Start by asking: what is it to have a certain desire? Roughly speaking, to have a certain desire is to be disposed to act in certain ways, given one's background beliefs. Desire is a propositional attitude that divides logical space into the set of desired worlds and the set of undesired worlds. So to have a certain desire is to be such that if you believed that an action open to you would make a desired world actual, then you would be to some degree moved to perform that action. But if that is what desires are, then no one could have a desire that that very desire not be satisfied, because there is no consistent way to divide the set of worlds into the desired and the undesired.

Of course, someone may say he has that desire. He may even believe he has it. (Maybe he paid a lot of money to a hypnotist who assured him that the hypnotism treatment has implanted the desire in him.) But that is not enough for him to, in fact, have the desire.

Actualist preferentists who are willing to accept this account of what desire is can tell a similar story about the desire to be badly off. They can say that it is impossible (in some circumstances) for anyone to have that desire, and for that desire to be intrinsic. (The motivation in this case is one that only preferentists can have. I will return to this point.) Preferentists will say that (in those circumstances) there is no consistent way to describe what a desired world is like, and no consistent way to describe what an undesired world is like. Of course, someone may say and believe that he has each of the desires in the statement of the paradox. But, as before, the preferentist will say that this is not enough to ensure that he does, in fact, have them. Even if this person takes every opportunity to do things that make himself worse off, the preferentist will deny that he desires to be badly off. So what desires does this person have? That will depend on the details, but there will always be some other desire that fits his behavior and background beliefs. If each of the things he does to make himself worse off involves the frustration of some first-order desire, then maybe he has a desire that none of his first-order desires be satisfied. On the other hand, maybe he holds a false theory of welfare (false by the preferentist's lights). Maybe he accepts some version of the objective-list theory of welfare, and so thinks (say) that knowledge contributes to welfare, even if he does not want or like knowing more. So he sets out to be as ignorant as possible. Since he does not want to know more, he is not frustrating any first-order desire. The best thing to say in this case is that he desires ignorance.

I have said that preferentists who adopt this solution will say that it is impossible for someone to desire to be badly off, if that person also has certain other desires. Should preferentists who adopt this solution say that in other circumstances, when there is no inconsistency, someone can desire to be badly off? Here is a reason to think that they should not say this. Suppose

[4] This is, roughly, the account given in Stalnaker (1984).

that (the preferentist allows that) under some circumstances, I can have the desire to be badly off. For example, suppose that my only desires are: a desire for salty peanuts (with intensity 20), and a desire to be badly off (with intensity 10). And suppose that I am eating some peanuts, satisfying my first-order desire. (Why would someone who desires to be badly off eat the peanuts? Maybe I tried to resist the urge to eat the peanuts, but could not.) There is no contradiction: my desire to be badly off is frustrated, and my welfare level is positive (in fact, it is 10). Now suppose that as I eat the peanuts, the intensity of my desire for peanuts decreases. Once it dips below 10 the scenario becomes inconsistent. So at that point I will have to cease having the desire to be badly off, and begin instead to desire that none of my first-order desires be satisfied (or have some other surrogate desire). This will have to happen even though I will insist that, from my point of view, the only fact about my desires that is changing is the intensity with which I desire the peanuts.

This kind of thing may or may not bother you, depending on your views in the philosophy of mind. Preferentists who are bothered should say that no one can desire to be badly off, under any circumstances.

That, I think, is the best case that can be made for rejecting (P1). How good a case is it?

For some, the account of desire it relies on may seem implausible. Someone (with some philosophical sophistication) is in a bar, feeling down and full of self-loathing, and he thinks: "I certainly have the concept of individual welfare. I know what it is that philosophical theories of welfare are trying to analyze. Of course I have no idea which of those theories is correct, and I am not sure how to go about lowering my welfare level, but having that concept is enough to know that what I want, right now, is for my welfare level to be negative. Furthermore, this desire is intrinsic: I do not just want to be badly off because (for example) I think I have done evil, and deserve punishment."[5] I think it takes some straining to say, in the face of this speech, that this person does not in fact want to be badly off.

And, in fact, if we reject (P1) we cannot rest at saying that it is impossible to have an intrinsic desire to be badly off.[6] Suppose there are two people, *A* and *B*, and their only intrinsic desires are as follows: *A* desires that *B* be well-off, and *B* desires that *A* be badly off. Given actualist preferentism, this scenario is inconsistent. So if our solution to the first paradox is to reject (P1), we will have to say that this scenario too is impossible – that two people cannot have that pattern of desire. It takes even more straining to accept this.

Finally, maybe the fact that this solution is theory-driven is a problem. (Bradley (2007) rejects it for this reason.) Hedonists, for example, find all scenarios in which people desire to be badly off perfectly coherent. Instead of agreeing with the preferentist that such desires are impossible, they say

[5] The person's self-loathing may be a *cause* of his desire to be badly off, but that does not mean the desire is not intrinsic; certainly intrinsic desires can and do have causes.
[6] I owe this point to an anonymous referee.

that preferentists have a false theory of welfare. I myself am not sure how serious a drawback this is. The idea must be that preferentists are not playing by the rules when they use their own theory to delimit what desires are and are not possible. But is there really neutral ground on which we can agree on what desires are possible, independent of what philosophical views we accept? I am not sure.

Nevertheless, a response to the argument against preferentism that permits people to desire to be badly off (a response that accepts (P1) but rejects (P2)) avoids this controversy, and respects whatever pre-theoretic intuition we may have that it is possible to desire to be badly off. So a response like that is a better response. In the rest of this paper I will present a response that does permit people to desire to be badly off.

3. My Solution Explained

Here is my response to the argument against preferentism. The key is to modify preferentism so that facts about *how close* someone is to having their desires satisfied play a role in determining their welfare level. There are many ways to make this modification. I will present one way in detail here, and briefly mention another at the end of the paper.

The first step is to generalize our framework for thinking about desire. We have been assuming that desires are either satisfied or not satisfied. Instead let us say that, at least sometimes, desire satisfaction comes in continuously varying degrees.

We also have to revise our assumptions about desire frustration. I assumed that a desire that *P* is frustrated if and only if that desire is not satisfied (which occurs if and only if it is not the case that *P*). Since we now allow satisfaction to come in continuously varying degrees, we must also allow frustration to come in continuously varying degrees.

Why should we think that adding degrees of satisfaction and frustration will help? Desires to be badly off create a problem for preferentism because they generate negative feedback in the "equation of welfare." The solution I am going to describe parallels solutions to other paradoxes involving a similar kind of negative feedback. (Famously, the liar sentence L says of itself that it is false. So L and $\neg L$ have the same truth-value. If the only truth-values are 0 and 1, this is impossible. One possible solution: allow truth to come in degrees, so that the liar sentence is true to degree 1/2.[7] This idea has also been applied to paradoxes of backward time travel (in Maudlin (1990)) and decision theory (in Arntzenius (2008)).)

Before we can see how degrees of satisfaction help solve our paradox, we need to develop the new version of preferentism in more detail.

[7] In fact, this approach to solving the liar paradox is no longer very popular. See Field (2008: sections 4.4-4.5) for a summary of the problems it faces.

Let us say that 0 represents the lowest degree to which a desire may be satisfied, and 1 the highest, and that a desire may be satisfied to degree *r* for any real number *r* between 0 and 1. The scale for desire frustration is the same.

Now, desire satisfaction and desire frustration are "opposites," in some sense. And the fact that they are opposites must be reflected in the scheme we use to represent degrees of satisfaction and frustration. It will not do, for example, for the scheme to permit one and the same desire to be satisfied to degree 1 and frustrated to degree 1 at the same time. So I assume that it is necessary that, for any desire, either the degree to which it is satisfied is 0, or the degree to which it is frustrated is 0 (or both). Non-zero degrees of satisfaction are incompatible with non-zero degrees of frustration.[8]

Now to say how the new theory calculates intrinsic values: Suppose I have an intrinsic desire for chocolate, with intensity 10, for one unit of time. According to the old theory, if that desire is satisfied, then that episode of satisfaction has value 10; if that desire is frustrated, then that episode of frustration has value 10 (but contributes negatively to welfare). The new theory just says that if that desire is satisfied *to degree r*, then that episode of satisfaction has value 10*r*, and similarly for frustration. The welfare value of someone's life is determined just as before, as the sum of the values of the episodes of satisfaction, minus the sum of the values of the episodes of frustration.

However, it will simplify our calculations if we use a notationally different (but equivalent) version of the new theory. So for the rest of the paper, I will adopt the convention that degrees of frustration are just negative degrees

[8] Another way to try to incorporate the idea that satisfaction and frustration are opposites (suggested by an anonymous referee) is as follows. Choose any desire and let *s* be the degree to which it is satisfied, and *f* the degree to which it is frustrated. Then the requirement is that $f = 1-s$. The difference between the two schemes is this: on the scheme I adopt, whenever a desire is partially satisfied, it is frustrated to degree 0. On the alternative scheme, whenever a desire is partially but not fully satisfied, it is also partially frustrated.

I think it is purely a matter of convention which scheme we choose. Each, when combined with some claim about how intrinsic values are determined, will lead to a different version of preferentism. But these different versions of preferentism will, I think, deliver the same verdicts about the (relative) intrinsic values of people's lives. I use the one in the text rather than the one in this footnote mainly for mathematical convenience.

It might be thought that the scheme in this footnote is superior, because it might seem like a conceptual truth that a desire that is not fully satisfied is partially frustrated. I do not agree. (In fact, I think that the claim that frustration and satisfaction are incompatible, and so that if a desire is even a little bit satisfied then it is not at all frustrated, has just as much claim to be a conceptual truth.)

What must be respected (if we are preferentists), I think, is the idea that a desire contributes less to one's level of welfare when it is less than fully satisfied than when it is fully satisfied. We (preferentists) are used to thinking that the only way for a desire to impact welfare negatively is for that desire to be frustrated; but once we have degrees of satisfaction, this is not the only way.

Still, as I said, I think it is a matter of convention which scheme we adopt. Everything I say can be re-formulated to fit with a theory that adopts the alternative scheme.

of satisfaction. That is, we permit the degree to which a desire is "satisfied" to be any real number between -1 and 1, and if my desire for chocolate is frustrated to degree *r*, then (we now say) it is "satisfied" to degree -*r*. This allows us to dispense with the distinction between episodes of desire satisfaction and episodes of desire frustration. Instead, the theory directly assigns intrinsic values to desires. That assignment works like this: the value of an intrinsic desire (at a time) is equal to

(the intensity of the desire) × (the degree to which the desire is "satisfied").

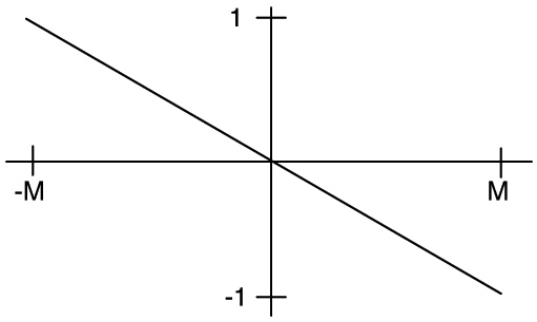
That, in outline, is the new version of preferentism. To apply it to any particular case, we need to know how to determine, for any desire, the degree to which that desire is satisfied. But I will not try to give a general theory of the conditions under which a desire that *P* is satisfied to degree *r* that applies no matter what the content of the desire. (I return to questions about what a general theory might look like in the conclusion.) For our purposes we only need to know how to determine the degree to which *D*, my desire to be badly off, is satisfied. In order to say how to make sense of the degree to which *D* is satisfied, I temporarily make two simplifying assumptions: that there is a lowest level to which my momentary level of welfare can fall, and a highest level to which my momentary level of welfare can rise. So I am assuming that, although I can be doing badly right now, I cannot be doing *arbitrarily* badly. (I will explain how to dispense with these assumptions below.) To keep things simple, let us say that there is one number *M* (for "maximum" and "minimum") such that *M* is the highest value my life could have for me at any particular time, and -*M* is the lowest.

Given these assumptions, what should we say about the way the degree to which *D* is satisfied varies with my level of welfare? Certainly *D* is maximally satisfied – satisfied to degree 1 – at some time just in case my welfare level at that time is -*M*. If I am doing as badly as possible, then things cannot be going any better for me with respect to satisfying *D*. Similarly, *D* is certainly maximally frustrated – satisfied to degree -1 – at some time just in case my welfare level at that time is *M*. But what about welfare levels between -*M* and *M*?

Figure 1 is a graph depicting one possible way to define the degree to which *D* is satisfied when my welfare level is between -*M* and *M*. The horizontal axis represents my momentary level of welfare. The vertical axis represents the degree to which *D* is satisfied. We know that any graph of the degree to which *D* is satisfied as a function of my momentary welfare level must start in the upper left and end in the lower right. This graph is the simplest one available: a straight line between those two points. According to this graph, as my welfare level increases and approaches zero, the degree to which *D* is satisfies decreases and approaches zero; then as my welfare level becomes positive and increases to *M*, the degree to which *D* is satisfied becomes negative and approaches -1. (This is the simplest way for the degree to

which *D* is satisfied to depend on my momentary welfare level; but it is hardly the most plausible way. I will show how the paradox can be avoided for less simple but more plausible patterns of dependence after I do so for the simple one.)

Figure 1: The degree to which *D* is satisfied, as a function of welfare.



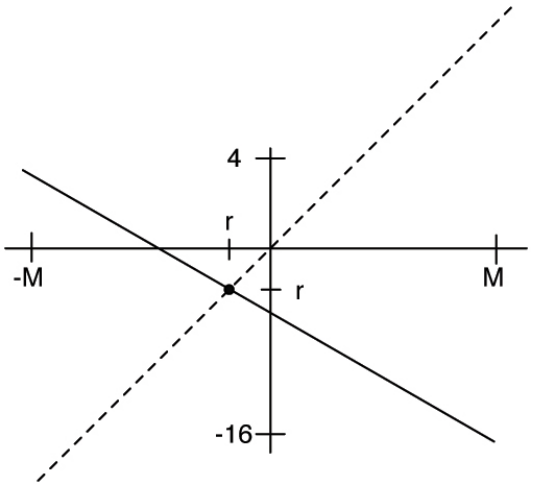
Above I showed how to derive a contradiction from preferentism and the possibility of a certain scenario in which I desire to be badly off. But that derivation assumed that my desire to be badly off was either satisfied or not satisfied; it took no account of degrees of satisfaction. So that derivation does not work against the amended theory. To derive a contradiction from the amended version of preferentism and the possibility of a scenario in which I desire to be badly off, it would need to be shown that it is inconsistent to suppose that my desire to be badly off is satisfied to degree *r*, for *every* real number *r* between -1 and 1.

And this cannot be done. In fact I can show directly that the amended theory is consistent. Suppose that in the scenario, the frustration of my first-order desires contributes, as before, -6 to the net amount of desire satisfaction I enjoy at *t*. Then my overall net amount of desire satisfaction at *t*, *NDES(t)*, is equal to

$$NDES(t) = -6 + (10 \times \text{the degree to which } D \text{ is satisfied}).$$

Now the degree to which *D* is satisfied at *t* is a function of my welfare level *W(t)* at *t*. So we can write *NDES(t)* as a function of *W(t)*. The graph of this function is in figure 2. Here the horizontal axis is my welfare level and the vertical is my net amount of desire satisfaction. The function takes its highest value of 4 when my welfare is *-M*, takes its lowest value of -16 when my welfare is *M*, and decreases linearly between those two points.

Figure 2: Graph of the net amount of satisfaction in the scenario, as a function of welfare.



Note the dashed line at a 45-degree angle in the graph. This line consists of the points where my net amount of desire satisfaction is equal to my welfare level. Let *(r,r)* be the point where the graph of *NDES(t)* crosses the dashed line. Since actualist preferentism says that *NDES(t)* = *W(t)*, this scenario is consistent with actualist preferentism provided that, in the scenario, *NDES(t)* = *r*. For suppose that, indeed, *NDES(t)* = *r* . Then what happens in the scenario that is supposed to generate paradox is this: my welfare level at that time is *r*. And *r* has the feature that

$-6+(10\times$ the degree to which *D* is satisfied when my welfare level is *r*$) = r$.

So there is no contradiction.

If preferentism is to say that *NDES(t)* = *r*, what must the degree of satisfaction of *D* be? To find the answer, we start by calculating *r*. The equation for the graph in figure 1 is

$$\text{(degree of satisfaction of } D) = -W(t)/M,$$

and so the equation for the graph in figure 2 is

$$\text{(net satisfaction at } t) = NDES(t) = -6+10(-W(t)/M).$$

We want a value for *W(t)* such that *NDES(t)* = *W(t)*. Using the second equation, we find that the solution is $W(t)=-6M/(M+10)$. Plugging this value into the first equation, we find that (preferentists should say that in this scenario) my desire to be badly off is satisfied to degree $6/(M+10)$. Since $-6M/(M+10)>-6$, the satisfaction of this desire does increase my level of

welfare at *t*. But because it is not satisfied to a very high degree, its satisfaction does not bump my welfare level above the 0 mark.

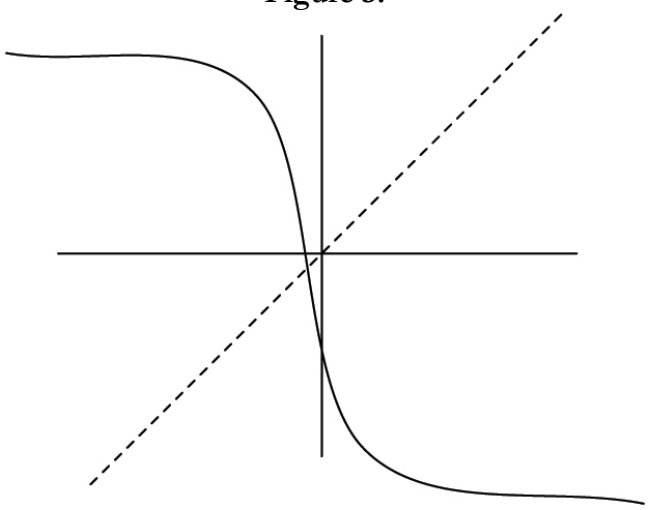
Recall how the paradox of desire was generated: introducing a second-order desire to be badly off led to negative feedback in the equation. If the left-hand side is positive, the right-hand side is negative. The problem is solved by introducing continuous variation. Then as the left-hand side increases, the right-hand side decreases. Since they are increasing and decreasing continuously, there must be a point at which they are precisely equal.

4. The Solution Generalized

My solution has two weaknesses. First, as presented, it depends on the assumption that there is a maximum and a minimum possible level for my welfare at any time. And, second, it depends on an implausible claim about the way that the degree to which *D* is satisfied depends on my welfare level.

But both of these weaknesses can be removed from the solution. The solution works as long as the degree to which *D* is satisfied varies continuously with my welfare level. Then the graph of *NDES(t)* as a function of *W(t)* must look something like the graph in figure 3: it starts out at its maximum value (4) far to the left, and decreases to its minimum value (-16) far to the right. It does not matter how close to zero my welfare level has to get before *D* stops being maximally satisfied or frustrated, and starts being partially satisfied or frustrated. It does not matter if my momentary welfare level can be arbitrarily high or low. As long as the degree to which *D* is satisfied varies continuously with my welfare level, there will always be some point where the graph will cross the diagonal.

Figure 3:



In light of what I have said, then, it should be clear what this modified version of preferentism says about other seemingly paradoxical situations. I have discussed a scenario in which I desire to be badly off – in which I desire that my welfare level be less than zero. Scenarios involving an intrinsic desire that my welfare level be less than 10, or less than -40, will receive exactly the same treatment. For another example, suppose that an intrinsic desire to be badly off is my *only* desire. Let this desire have whatever intensity you like. This scenario is consistent (with my version of preferentism). My desire will be satisfied to some positive degree if my welfare is negative, and satisfied to some negative degree if my welfare is positive. The result: my desire will be satisfied to degree zero, and my welfare level will be zero. Similarly, if my only intrinsic desire is that none of my desires be satisfied, then this one desire will be satisfied to degree zero.[9]

[9] I earlier said something about the connection between the paradox of desire and the liar paradox. There is a dual problem: the problem of the truth-teller. The liar sentence cannot consistently be supposed to be true, or to be false. The truth-teller, on the other hand, says of itself that it is true, and *can* consistently be supposed to be true. It can also consistently be supposed to be false. Which is it? Its truth-value is ungrounded – not determined by anything else.

So (an anonymous referee asked) what about the analogous scenario for (the old version of) preferentism: a person whose only intrinsic desire is a desire to be well-off? His welfare level is ungrounded: it is consistent to suppose he has positive welfare (for then his desire is satisfied), and also consistent to suppose he has negative welfare (for then it is not). Since I have not presented a detailed theory about the way the degree to which a desire is satisfied depends on the circumstances, the version of preferentism I present does not answer the question whether there is exactly one consistent assumption about this person's level of welfare. (Symmetry considerations suggest that the new version entails that there is *at least* one consistent assumption about his welfare level: the assumption that both it and the degree to which his desire is satisfied are zero. But nothing follows about whether there are other consistent assumptions.) I also do not have a fixed view about what preferentists *should* say about this scenario. I do not think they are required to have a theory that assigns a de-

5. Conclusion

In order to solve the paradox of desire, I have proposed an extension of our theory of desire. Instead of letting desire satisfaction be on/off, I let desire satisfaction come in continuously varying degrees. Unlike the first response I discussed, mine permits people to desire to be badly off.

There are lots of interesting questions about degrees of desire satisfaction that I have not tried to answer in this paper. Let me say something here about a couple of them.

First, we will want to know: for any *P* and any real number *r*, under what conditions is a desire that *P* satisfied to degree *r*? For example, suppose I desire to run a marathon in three hours, and in fact I finish in 3:01. To what degree is my desire satisfied? Let us use "the satisfaction conditions" of a desire to name the way in which the degree to which that desire is satisfied varies with the way things are. Then we are asking what determines the satisfaction conditions of a desire. On the old theory, without degrees of satisfaction, the satisfaction conditions of a desire were determined by its content: a desire that *P* is satisfied iff *P*. But on the new theory, satisfaction conditions are not determined by content. So what does determine them?

One kind of question about satisfaction conditions is especially interesting: questions about the "zero point" for the satisfaction of some desire. Consider again my desire to run a marathon in three hours. How close do I have to get to three hours for this desire's level of satisfaction to stop contributing negatively to my welfare, and start contributing positively?

I have said nothing that allows us to answer these questions. I have not tried to present any general theory of satisfaction conditions. But my solution does not depend on any particular theory of this kind. In the paradoxical situations I have been talking about, we know that however the degree of satisfaction of *D* varies with welfare, there is a consistent solution.

Another question is this: for which *P* is it the case that a desire that *P* can be satisfied to an intermediate degree? Is this true for all desires, or only for some of them? I myself suspect that it is true only for some of them. But I do not know exactly what criterion distinguishes desires that can be satisfied to intermediate degrees from those that cannot. Still, if degrees of desire satisfaction make sense at all, then *D* (and other desires like it that threaten to generate paradox) can be satisfied to intermediate degrees, and that is what is needed for my solution to work.

One theory of satisfaction conditions that answers both of these questions says that truth comes in degrees, and that the degree to which a desire

termine level of welfare to the person in this scenario; perhaps there is no fact of the matter about how well off he is. (If someone wanted to combine preferentism with a standard version of consequentialism, though, he would need to assign the person in this scenario a definite level of welfare, in order to assign a determine value to the possible world in which he resides.)

that *P* is satisfied is determined by the degree to which *P* is true. I do not like this theory, though, because I do not believe in degrees of truth. I also do not think the theory would deliver the right verdicts. (I suspect that my desire to finish a marathon in three hours is satisfied to some positive degree when I run 3:01, even though it is perfectly false that I finished in three hours.) I suspect that a desire that *P* admits of degrees of satisfaction when the "object" of the desire itself admits of degrees, in some sense that I cannot state more precisely. I only have examples: my marathon time, for example, admits of degrees. I do not know how to develop this suspicion into a more concrete proposal.

I will finish by addressing an objection, and then saying a few words about another way to develop the idea behind my solution. Although my solution permits people to desire to be badly off, there are still desires that my solution says no one can have. No one can have a "sharp cut-off" desire to be badly off: one that is satisfied to degree 1 if the subject's welfare level is negative, and is satisfied to degree 0 otherwise. Now, there may be plenty of pre-theoretic reasons to think that desires to be badly off are possible. But is there any reason to think that sharp cut-off desires to be badly off are possible?

Some may think that just as they can come to know the contents of their desires by reflection alone, they can come to know the satisfaction conditions of their desires by reflection alone. And they may say that when they reflect, they see a desire to be badly off with sharp cut-off satisfaction conditions. But I am skeptical that the satisfaction conditions of our desires are available to reflection.

One might think that the satisfaction conditions of a desire are available to reflection, and so think that one has a sharp cut-off desire to be badly off, by having the following thoughts: "I've had a sharp cut-off desire to be badly off many times. When I have been in a bad mood and wanted my life to be going poorly, I would not have been satisfied at all to learn that my welfare level was non-negative; the only thing that would have satisfied me was learning that my welfare level was negative – and it would have satisfied me completely." But this kind of reflection does not tell us about the satisfaction conditions of our desires. There is a confusion at work here about what desire satisfaction is. It is wrong to think of the degree to which a desire is satisfied as the intensity of some *feeling* of satisfaction that the agent experiences. This is a lesson we can learn from preferentism without degrees of satisfaction. According to that theory, someone's desire that *P* is satisfied iff it is the case that *P*. Even if it is the case that *P*, the person may not know that *P*, and may even believe that *P* is false. So even though her desire is satisfied, she may experience no feelings of satisfaction at all.

Still, there may be some who firmly believe that we can have sharp cut-off desires, and some who object to allowing desire satisfaction to come in degrees in the first place. The basic idea behind my response is compatible with these convictions. That idea is to allow facts about how close someone

is to having his desires satisfied play a role in determining his welfare level. I have incorporated these facts by modifying my theory of desire. For people who insist that desire just does not work that way, there is a way of developing the response that is consistent with the view that desire satisfaction is on/off. This way of developing the response goes, in outline, like this: Suppose that desire satisfaction is on/off, and that my welfare level is positive, so that my desire to be badly off is frustrated. There are still facts about how close my desire is to being satisfied. If my welfare level had still been positive but slightly lower, then my desire would still have been frustrated, but it would have been closer to being satisfied. Modify preferentism so that, roughly speaking, the closer my desires are to being satisfied, the better off I am. Formally speaking, this response works in the same way as the response I presented in some detail above, but it does not require the claim that desire satisfaction itself comes in degrees. I like the version of the theory that incorporates degrees of satisfaction, because I find the claim that desire satisfaction comes in degrees plausible, and because I suspect that degrees of satisfaction may have other theoretical applications. But I offer this alternative solution to the paradox to those who will have no truck with degrees of satisfaction.[10]

Bradford Skow
Massachusetts Institute of Technology
Department of Linguistics and Philosophy
bskow@mit.edu

[10] Thanks to Ben Bradley, Fred Feldman, and audiences at the Rocky Mountain Ethics Congress and the MIT philosophy retreat for helpful comments. Thanks to Agustin Rayo and Caspar Hare for helping me re-think how to present my solution. Finally, I am very grateful to my two anonymous referees. Their feedback led to substantial improvements to many parts of the paper.

References

Arntzenius, Frank (2008). "No Regrets." *Erkenntnis* 68: 277-297.

Bradley, Ben (2007). "A Paradox for Some Theories of Welfare." *Philosophical Studies* 133: 45-53.

Feldman, Fred (2004). *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. New York: Oxford University Press.

Field, Hartry (2008). *Saving Truth from Paradox*. New York: Oxford University Press.

Heathwood, Chris (2005). "The Problem of Defective Desires." *The Australasian Journal of Philosophy* 83: 487-504.

Kagan, Shelley (1998). *Normative Ethics*. Boulder: Westview Press.

Maudlin, Tim (1990). "Time Travel and Topology." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1: 303-315.

Parfit, Derek (1984). *Reasons and Persons*. New York: Oxford University Press.

Sober, Elliott and David Sloan Wilson (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Stalnaker, Robert (1984). *Inquiry*. Cambridge, MA: MIT Press.