

Preferentism and the Paradox of Desire

Brad Skow

Draft of April 25, 2008

1 The Paradox Stated

The paradox of desire is an objection to desire-satisfaction, or preferentist, theories of welfare. In a nutshell, the objection goes like this. I can certainly desire that I be badly off. But if a desire-satisfaction theory of welfare is true, then—under certain assumptions—the hypothesis that I desire that I be badly off entails a contradiction. So much the worse for desire-satisfaction theories of welfare.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.

Actualist preferentism is a theory that tells us what it is for someone's life to go well for her. On the standard atomistic version of the theory, the "atoms" of welfare are episodes of basic desire satisfaction: episodes in which the subject has a basic desire that p , and it is in fact the case that p . "Episode of basic desire frustration" is defined analogously. 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. Then the value of someone's life for her is just the net amount of

¹To my knowledge, the first appearance of this argument in print is in [Feldman 2004].

desire satisfaction in the life—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 desire satisfactions that are occurring at t , minus the sum of the intensities of the desire frustrations that are occurring at t .

The paradox of desire, then, arises in the following kind of situation. Suppose I have several 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 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: suppose that D is satisfied. Then my welfare level at t is $10 - 6 = 4$, a positive number. But then D is not satisfied. So if D is satisfied, it is not satisfied. This entails that D is not satisfied. But then, since D is frustrated, my welfare level at t is $-10 - 6 = -16$, a negative number. So D is satisfied. We have shown that D is both satisfied and not satisfied; which is absurd.

Intuitively, what generates the problem 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? I will only mention one of the responses that have been discussed.² Heathwood [Heathwood 2005] defends actualist preferentism. At the end of his paper he discusses the paradox of desire. 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. Then my desire is satisfied if and only if it is not satisfied; which is absurd. Heathwood suggests that whatever solution we find to this paradox can be adapted to respond to the argument against preferentism. So he is not worried.

This is certainly false. One could respond to the second paradox by saying that it is impossible that someone have as his one and only desire the desire that none of his desires be satisfied. But this response cannot be adapted to a solution to the original paradox. It is not impossible that I desire that I be badly off.

Still, the solution I will offer vindicates Heathwood's optimism. The idea behind my solution can be adapted to show that there need be nothing inconsistent in a scenario in which someone desires only that none of his desires be satisfied. (I will explain how in the final section.)

2 The Solution Explained

Here is my response to the argument against preferentism. The key is to recognize that, at least sometimes, desire satisfaction can come in degrees. I will say more to defend the idea that desire satisfaction comes in degrees in the next section. First, though, let us see how it helps solve the paradox.

²[Bradley 2007] discusses several other attempts to solve to the paradox of desire. He argues convincingly that none of them succeed, but he does not discuss the current proposal.

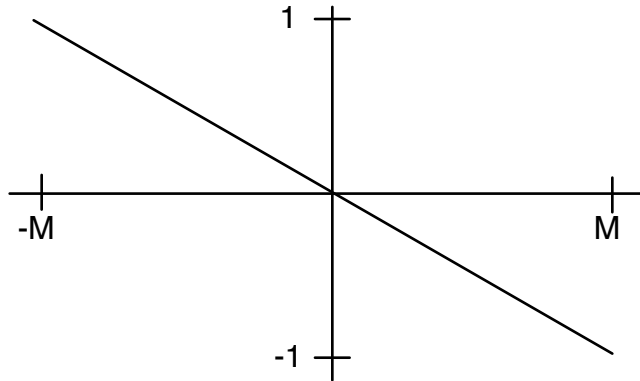
I will not try to give a general definition of “my desire that p is satisfied to degree r ” that applies no matter what the content of my desire. Instead, I will focus just on D , my desire to be badly off. In order to say how to make sense of the degree to which D is satisfied, I temporarily make two 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 can’t 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 can’t 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 . (So I say that a desire is frustrated when it is satisfied to a negative degree; and measure the degree of frustration as $(-1) \times$ the degree to which it is satisfied.) 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 satisfied 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.



Now that we have this notion of the degree to which a desire is satisfied, we need to amend the preferentist’s definition of the intrinsic value of an episode of desire satisfaction. For clearly, if preferentism is true, the degree to which I am getting what I want should play a role in determining how well off I am. (If two people have all the same desires and are having all of their desires satisfied, but the first person’s desires are satisfied to a much higher degree, then surely he is better off.) The obvious way to amend the theory is as follows: the intrinsic value of an episode of desire satisfaction is equal to the intensity of the desire times the duration of the episode *times the degree to which the desire is satisfied*. (Since I use “negative levels of satisfaction” to represent desire frustration, the intrinsic value of an episode of desire frustration is also equal to intensity \times duration \times degree of satisfaction. And as before, since I’m concerned with welfare levels at an instant, I will ignore the duration factor in my calculations.)

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.

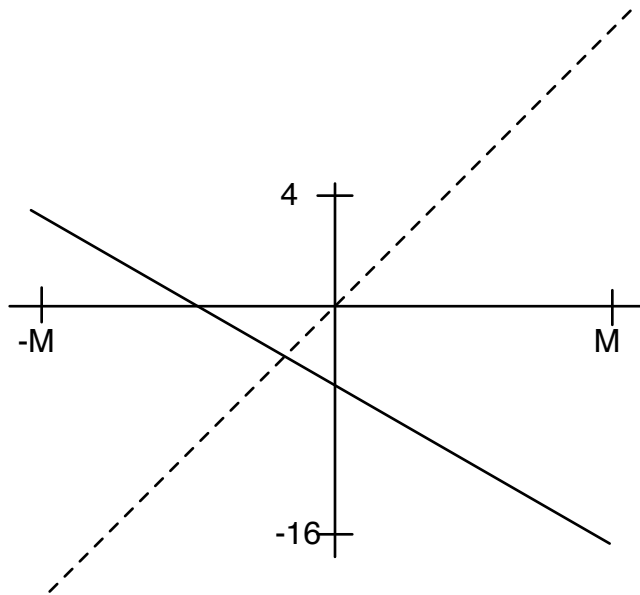
In fact, I can show directly that the amended theory is consistent. Suppose that in the sce-

nario, the frustration of my first-order desires contributes, as before, -6 to my level of welfare at t . Then, letting “ $W(t)$ ” abbreviate “my welfare level at t ,” my momentary level of welfare at t is equal to

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

Now because we know how the degree to which D is satisfied depends on $W(t)$, we can determine how $(-6 + (10 \times \text{degree to which } D \text{ is satisfied}))$ depends on $W(t)$. The graph of this dependence is in figure 2. Here the horizontal axis is my welfare level and the vertical is $(-6 + (10 \times \text{degree to which } D \text{ is satisfied}))$. 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 is special: it represents points where my net amount of desire satisfaction is equal to my welfare level. You can see that the graph of $(-6 + (10 \times \text{degree to which } D \text{ is satisfied}))$ crosses that line at a point (r, r) . That is the solution we are looking for. 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 $(-6 + (10 \times \text{degree to which } D \text{ is satisfied when my welfare level is } r)) = r$. So there is no contradiction.

So far I have just displayed graphs; but we could also calculate r directly by writing down equations for these graphs. The equation for the graph in figure 1 is

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

and so the equation for the graph in figure 2 is

$$W(t) = -6 + \left(10 \times \frac{-W(t)}{M}\right).$$

The solution to this second equation is $r = -6M/(M + 10)$. That is, in this scenario, my desire to be badly off will be 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.

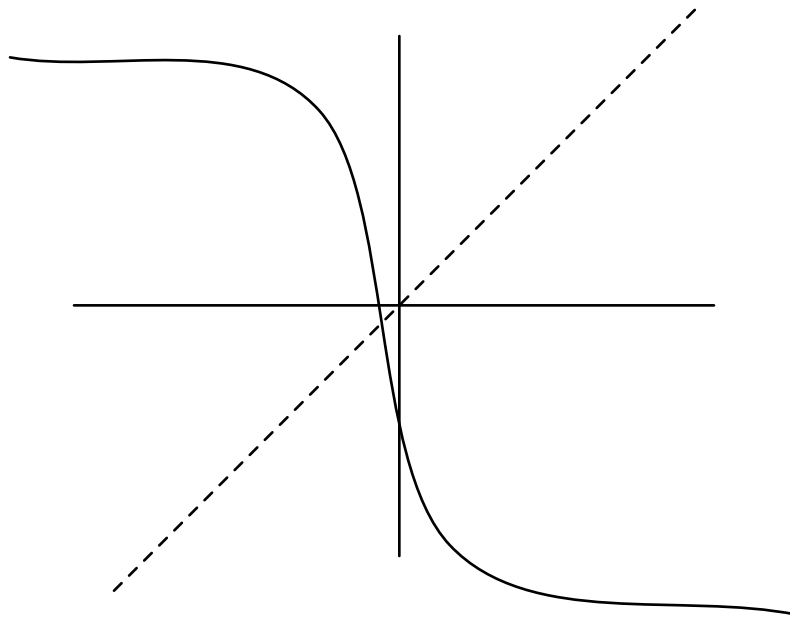
3 The Solution Generalized

My solution has two weaknesses. First, as presented, it depends on the assumptions that there is a maximum and a minimum possible level for my welfare at any time. And second, it depends on a 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 $(-6 + (10 \times \text{degree to which } D \text{ is satisfied}))$ 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. 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. For example, if my only desire is to be badly off (with whatever intensity you like), then 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 desire is that none of my desires be satisfied, then this one desire will be satisfied to degree zero.

4 The Solution Motivated

Desires to be badly off create a problem for preferentism, because they generate negative feedback in the “equation of welfare.” My solution 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. Solution: allow truth to come in degrees, so that the liar sentence is true to degree $1/2$. This idea has also been applied to paradoxes of backward time travel [Maudlin 1990] and decision theory [Arntzenius (forthcoming).] But even though these paradoxes share a common form, it does not follow that they have a common solution. The key to this approach to solving them is the claim that there is continuous variation of the right kind. (For the paradox of desire, this is variation in the degree to which the desire is satisfied.) But whether it is plausible to claim that there is the right kind of continuous variation varies from case to case. Maybe truth comes in degrees but desire satisfaction does not, or vice versa, so we cannot solve both paradoxes in the same way. So does desire satisfaction come in degrees? I think it does.

Degrees of satisfaction appear in the most natural description of certain scenarios. Suppose, for example, that I want pie. And suppose I am given a tiny tiny slice of pie. Surely I would be *more* satisfied with more pie.

Or again, suppose I want exactly five apples. I am told that this is just not in the cards: there are not five apples in the store, or even the entire town. There are only four apples left. Even though I cannot get exactly what I want, surely I am better off getting four apples than getting none at all. Surely that is what I would prefer. But that must be because getting four apples at least gets me close to getting what I want. That is, at least my desire will not be left completely unsatisfied.

Or yet again, suppose I want to run a three hour marathon. In fact I run the marathon in 3:01. I would be pretty satisfied, even though I didn’t get exactly what I wanted. I would certainly be more satisfied with 3:01 than with 3:40. So, again, the satisfaction of this desire

comes in degrees.

For an example closer to the topic of this paper, suppose that I want to be badly off. Suppose my momentary welfare level is $+0.001$. Then my desire to be badly off is not satisfied to any degree. Suppose things get a little worse for me, so that my momentary welfare level is -0.001 . I am now badly off, but only as a result of an extremely tiny change in my momentary welfare level. Now I am getting what I want (my desire to be badly off is satisfied), but *just barely*. If I were worse off, surely I would be doing better as far as having that particular desire satisfied.

I do not claim, though, that the satisfaction of just any desire comes in degrees. I want to qualify for the Boston Marathon. As of this writing, that requires me to run a marathon in 3:10. If I run 3:11:01, my desire to qualify for Boston is not satisfied to any degree at all. In this case, a miss is as good as a mile (or 26.2 miles).

I suspect that the following is true: the satisfaction of a particular desire comes in degrees only if the “object” of that desire admits of degrees, in some sense I can’t state more precisely. I can only give examples. If my desire is a desire to run a marathon in a particular time, (and this desire is basic, not derived from a desire to qualify for Boston,) then the satisfaction of my desire comes in degrees, because you can measure how close or far away my time is from the desired time. Similarly, if I desire to be badly off, you can measure how close or far away my momentary level of welfare is from being negative. But it is not like that with qualifying for Boston. Even if my time is close to 3:10, the state I am in is not “very similar” to a state in which I have qualified for Boston. It is not more similar to that state than it is to a state in which I do not get to run. It is exactly like a state in which I do not get to run.

I have one final remark. As Bradley [2007] notes, an analogue of the paradox of desire arises for Truth-Adjusted Intrinsic Attitudinal Hedonism [Feldman 2004]. Briefly, this theory says that the “atoms” of welfare are episodes of attitudinal pleasure—episodes of my taking pleasure in the fact that p , for some p . But the intrinsic value of an episode is adjusted for truth: the value of an episode is the intensity of the attitude times the duration of the episode times the

“truth multiplier.” (Let us say that the truth multiplier is 1 if p is false and 2 if p is true; it is twice as good to take pleasure in true facts than in false ones.) Paradox threatens when someone takes pleasure in the fact that his life is going poorly for him. If the numbers are set up right, and his life is, in fact, going poorly, then his episode of attitudinal pleasure gets a “truth boost,” increasing his welfare above the zero mark, so that he is no longer doing poorly; as a result, his episode of pleasure loses the truth boost, and his welfare falls back below the zero mark. The scenario is inconsistent.

The cautionary remarks I made at the beginning of this section apply here. Even though this paradox for hedonism is structurally the same as the paradox of desire, I do not think it can be resolved in the same way. I can make sense of degrees of desire satisfaction. But I do not know where to insert the degrees into the truth-adjusted value of an episode of attitudinal pleasure. Even if it makes sense in some situations to introduce “degrees of truth,” it does not seem to help here, since the proposition that the person in the scenario is badly off is true either to degree 1 or 0 in the scenario.³

References

Arntzenius, Frank ((forthcoming)). “No Regrets.” *Erkenntnis* .

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.

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

³Thanks to Ben Bradley and Fred Feldman for helpful comments.

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