

Desire, Time, and Ethical Weight

Dr. Nick Bostrom
Department of Philosophy
Yale University
New Haven, Connecticut 06520
U. S. A.

Fax: **(203) 432-7950**

Phone: **(203) 500-0021**

Email: **nick@nickbostrom.com**

Desire, Time, and Ethical Weight

1. Introduction

We want more than we can have. So we must somehow weigh our desires in order to determine which of them it is most important that we satisfy. When the objective is simply to maximize the satisfaction of one's own current desires ("egoism of the present moment"), decision theory provides a well-worked out, albeit highly idealized, framework for conducting the requisite deliberation. It takes into account the strength of the various desires one has and the subjective probabilities of the possible outcomes of the actions one is choosing between (and, on causal decision theories, also one's beliefs about the causal relations between actions and outcomes¹). All that an egoist of the present moment has to do is to introspectively ascertain these parameters, calculate (or intuitively estimate) the expected utility of the various choices, and then pick the one with the highest expected utility.

But what if one is not an egoist of the present moment? What if, for example, one's objective is to maximize desire-satisfaction simpliciter, or even just satisfaction of one's own desires, past, present, and future? Then there is a complication. For in order to get a measure of the total preference satisfaction in these cases, we have to compare preferences held at different times and for different durations. How to operationally compare the strengths of different people's preferences, or the preferences of the same person at different times, is a notorious challenge for social choice theory. We shall not address it here. Instead, we will focus on role of time: the relevance of when a preference is held, and, especially, of how long it is held.

The problem of how the duration of a preference influences how much its satisfaction contributes to overall preference satisfaction has not received much attention in the literature, which is unfortunate because it is a rather poignant one and it brings out some deep issues once one starts thinking about it systematically. Of course, one does not have to be a preference utilitarian to find the problem interesting. From a practical point of view, it is important because evaluating outcomes in terms of desire-satisfaction is the standard approach in social choice theory and welfare economics.² There are also many an ethical view that assigns weight to desire-

satisfaction in at least some contexts even if it is not the only factor of the good. Moreover, any links we can find between time, desire and ethical weight will contribute to our philosophical understanding of each of these concepts.

Let us therefore assume, for the sake of the argument, that desire-satisfaction has ethical weight. I emphasize that my purpose is not to defend this view, but rather to make some friendly suggestions for how it could be developed and to explore its consequences.

2. The problem is distinct from, but analogous to, a problem faced by hedonistic utilitarians

If a desire has been held for a longer time, is that a reason to assign it a greater ethical weight? If so, how much?

These questions are quite distinct from the problem, confronting a hedonistic utilitarian, of how to weigh pleasures that last for different periods of time. There is a widely accepted solution to this latter problem: that the ethical weight of a pleasure (for a hedonistic utilitarian) is proportional, other things equal, to its duration. This view was introduced by Jeremy Bentham.³ Holding other things – like the intensity and the purity – of a pleasure constant, it is twice as good if it lasts twice as long.

In this paper, however, desire-satisfaction is not a psychological notion. We take it, rather, in a Tarskian sense. A desire ‘that P’ is satisfied if and only if P, independently of whether the person holding the desire experiences any psychological satisfaction or is even aware that her desire is satisfied. We can therefore not straightforwardly transpose the intuition that underlies the Bentham view to the present context. For while it is true that a pleasure lasting twice as long as another otherwise similar one contributes twice as much to the total amount of pleasurable experience in the world, whence it should count as twice as good for a hedonistic utilitarian, it is not obvious that the corresponding observation about desire-satisfaction would provide a reason for thinking that the satisfaction of a desire that has been held for twice as long should count as twice as good for a desire-utilitarian. This step requires further argumentation.

Nevertheless, the analogies between the two cases are such as to suggest that this step might indeed be right. Whether we are weighing pleasure or desire-satisfaction, we should take their intensity into account. We should also take into account the probabilities of their coming about as a result of the different actions we are choosing between. And we should consider the

likely consequences of these actions for bringing about further pleasures or satisfaction of desires. Given these parallels, it seems reasonable to conjecture that the analogy extends to the factor of duration as well. A later section provides an argument for this – indeed, for the stronger claim that the ethical weight of a desire is, ceteris paribus, proportional to its duration.

All this is modulo the working assumption that we accept some kind of desire-satisfaction view of the good. It should be stressed, however, that the duration of a desire should not be assumed to be the only factor determining its ethical weight. One may also regard a desire's intensity, how well-informed it is, or how worthy the person is who has the desire, as relevant variables. And we can allow additional values beside desire-satisfaction. Yet every view that accords desire-satisfaction ethical significance ought to have something to say about how to weigh desires of different durations, and it is this challenge that is the topic of our present investigation.

3. Do latent desires count?

Active desires, as we shall use the term here, are ones that have the form of an occurrent conscious desire. Latent desires are ones that are not present in one's awareness. For example, I have had a latent desire for many years to be able to give everybody in the world a nice chocolate-mousse treat. Had I been offered this ability, I would gladly have accepted it. I was never conscious of this particular desire before I began writing this paragraph. Right now it's an active desire. Soon it will recede from my consciousness and again become a merely latent desire.

Supposing desire satisfaction is a good, do passive desires count? Is it good for me if my latent desires are satisfied even if they are never active, and even if it doesn't help me satisfy any of my active desires? We shall return to this issue in a later section, but some preliminary discussion is in order at this point.

One might argue that our active desires are too few and far between to form an adequate basis for defining our interests. We don't go around constantly with an active desire to be alive, yet killing somebody without warning would almost always be wrong. Since there was no active desire for not being killed, the act must have been wrong because it violated a passive desire. Passive desires therefore carry ethical weight.

This argument is much too quick. For one thing, even if only active desires count, killing somebody might be wrong for reasons that are independent of the good of desire satisfaction. But supposing that desire satisfaction is the only factor under consideration, the conclusion still doesn't follow. For although at the time of the killing the victim might not have had an active desire for not being killed (maybe he was asleep), he might have had such a desire earlier in his life, and maybe that's why it would be wrong to kill him.

Suppose a person has never had an active desire for being alive. He enjoys various activities but it never occurs to him to specifically desire that his life continue. Even here, however, we can say why, in regard to the good of active desires being satisfied, it would be wrong to kill him. Killing him would, as a "side-effect", frustrate many of his active desires. He cannot go on that ski trip that he is looking forward to if he is dead. Likewise for the ability to throw a worldwide chocolate-mousse party: I have had active desires for other things (e.g. alleviating world hunger, or being entertaining) that I would be better able to satisfy if I possessed the ability in question.

These considerations suggest that one shouldn't too quickly dismiss the view that ethical weight is confined to active desires. The issue is important because most desires are latent most of the time.

We now turn to consider the idea that the duration of a desire influences how ethically important it is that it be satisfied. Maybe that idea, when thought through, can help us decide the issue of latent desires. Let us therefore first focus on the special case where all desires are active.

4. Does duration matter?

Let's try to construct such a case, where the desires involved are of different durations but are similar in all other ethically relevant respects:

The Wall. Two adults, Alice and Bob, pop into existence and live for two hours each and are then killed. Alice and Bob are ethical equals (they are equally worthy and deserving, equally passionate, intelligent, sensitive, and similar in every other arguably relevant aspect). Your job is to paint a certain wall in a single color long after Alice's and Bob's lives are over. Alice wants throughout her two-hour life that the wall be yellow. Bob

wants, for the first hour, that the wall be blue; he then changes his mind and for the rest of his life he it to be green. Except as regards their durations, these three desires are equivalent in all relevant respects (they are equally intense, equally well-informed, etc.) We assume that there are no desires except the abovementioned, and that Alice and Bob have no way of knowing how the wall is painted. How should you paint the wall if you are perfectly ethical and your only concern is Alice and Bob's desires?

Of course, one might think that it doesn't matter how the wall is painted since the only people who care about it will never get to know its color and will not be directly affected by your action in any way. But given that we are trying to find the best way of fleshing out a desire-satisfactionist account of the good (or of some part of the good), and to explore the consequences of such a view, that is clearly not the path to choose at this juncture. So we definitely want to say that you should paint the wall in one of the three colors yellow, green, or blue, and that choosing any other color would be worse.

On the view that the duration of the desires is ethically irrelevant, it doesn't matter which of these three colors you pick. Either alternative would satisfy exactly one desire and frustrate two, and the case is one where nothing but the duration of the desires seems relevant to their ethical weight.

However, I think that this view should be rejected, on grounds that it violates the principle that we should give equal ethical weight to the welfare of persons who are equal in all ethically relevant ways (EQUAL). Alice's welfare (which, given our desire-satisfactionist premiss, must be construed as the degree to which her desires are satisfied) is no more and no less ethically important or capable of contributing to the total value of the outcome than Bob's welfare. Their concerns have equal weight. Now, painting the wall either blue or green will result in one person being half-satisfied (i.e. one out of Bob's total of two equally significant desires get satisfied) and one person (Alice) being completely unsatisfied. Painting the wall yellow achieves more – it results in one person being completely satisfied (and another unfortunately still completely unsatisfied). Claiming that one may just as well paint the wall blue or green as yellow would, in effect, be to assign Bob more ethical weight than Alice, because this view would rank

Alice's complete satisfaction as no more valuable than Bob's half-satisfaction. But since Alice and Bob are ethical equals, this would violate EQUAL.

Couldn't one argue, contrary to the above reasoning, that the situation is one where there are two different outcomes (blue and green) either of which would be just as good for Bob as yellow would be for Alice? We could then reconcile giving equal ethical weight to Alice and Bob with holding it to be indifferent which of the three desires that we satisfy. Alice's desire lasted longer, so what? We could reject the view that the duration of a desire is ethically relevant.

But is it just as good for Bob to have one of his desires satisfied as it is for Alice to have her sole desire satisfied? Presumably not. For what if Bob could have both of his desires satisfied? Surely that would be better for him than having only one of them satisfied. So we'd either have to hold that having both his desires satisfied would be better for Bob than having her only desire satisfied would be for Alice, or we have to accept that satisfying one of Bob's desires is not as good for Bob as satisfying her only desire would be for Alice. The latter yields the preferred conclusion, assuming EQUAL and assuming that Alice and Bob are ethical equals. The former alternative means that Bob has a greater capacity of contributing to the overall amount of desire-satisfaction in the world merely by virtue of having a greater number of desires in the given time interval than Alice has. The difference between Bob and Alice is that Bob changes his mind more frequently. Presumably we do not wish to reward such fickleness by ascribing Bob a greater capacity for desire-satisfaction.

A further problem of with the view that many brief desires count for more than a single one of the same total duration is that it leads to the difficulty of individuating desires that are only very slightly different from one another. Suppose that, instead of the clear-cut blue and green desires, Bob had a desire that was centered on some specific color but with some tolerance for small deviations in nuance. Let us suppose, to be concrete, that he started out desiring a particular shade of blue and that the center of his preferred segment of the color spectrum gradually drifted so as to reach a particular shade of green at the end of the two-hour interval. In some sense, he would then have had infinitely many (a continuum) of distinct desires, and each point of the color spectrum between blue and green would be included in the scope of infinitely many particular desires. Following the idea that duration doesn't matter, we would in this case get the conclusion that it would be definitely wrong to paint the wall yellow because that would satisfy only one

desire whereas any shade between blue and green would satisfy infinitely many desires. That seems very wrong. We can see this even more clearly if we suppose that there are a thousand people like Alice. Then there would then be a thousand desires for yellow but infinitely many for blue-green, so Bob would still win out – but that is a consequence that we certainly do not wish our theory to countenance.

Consequently, we should reject the view that it would be just as good if Bob gets one of his desires satisfied as it would if Alice gets her sole desire satisfied. Since there seems to be no other relevant difference between these desires than their unequal durations, the duration of a desire should be regarded as ethically relevant.

5. Does it matter whether a desire is in the past?

One might object to the above argument on the ground that there is an asymmetry among Bob's desires: one of them occurs later than the other. If Bob's earlier desire counted for nothing, then painting the wall green would be just as good as painting it yellow. But assuming that the earlier desires count for something then the unique best color is yellow.

However, giving only the most recent desire a person has ethical weight is an unattractive view. While the present-aim theory is a contender for the job of giving an account of individual rational decision-making, its analogues in ethics have some unpalatable consequences. I shall confine myself to commenting on one of these, which to my knowledge has not been noticed in the literature. This consequence becomes apparent when we consider the matter in light of the lessons that the special theory of relativity has taught us about the nature of time in a relativistic spacetime.

Consider the doctrine that when assigning weights to a person's desires, you should determine what the most recent desire is that she had and assign all other desires she had zero weight. According to special relativity, "most recent" is a frame-relative notion⁴. A desire that is in the past of some event relative to one reference frame may be in the future of that event relative to another reference frame. "Ethical weight" would thus also be a frame-relative notion if this doctrine were correct. This would have the consequence that how ethically important it is for you at some point in time that Alice's desire be satisfied would depend on your velocity relative

to Alice. There would be cases in which, if you had a fast spaceship, you could literally outrun your ethical obligations, which seems bizarre.

If we require ethical weight to be a frame-independent notion then we must avoid defining ethical weight in terms that explicitly or implicitly invoke the concept of simultaneity. This rules out doctrines that say, for example, that the ethical weight of peoples’ desires is greater if those desires are “in the past” or “in the recent past” or are “the most recent ones the person has yet had”.⁵

The same argument shows that doctrines that give ethical weight to only present and future desires, or to past and present desires, must be rejected if we insist on our basic ethical concepts being frame-independent. Of course, there are often instrumental reasons for giving special importance to a person’s most recent (in your reference frame) desires. But these reasons do not stem from any intrinsic ethical priority that these desires have, but rather from various practical considerations, some of which we shall describe in a later section.

6. How much does duration contribute to ethical weight?

I have argued that if desire-satisfaction is an ethical good, then desires of longer duration have greater ethical weights (other things equal).

How much does duration contribute to ethical weight? In order to examine this relation in more specific terms, we will pursue the idea that Bentham’s rule for taking the duration of a pleasure into account may have an analogous formulation pertaining to the weighting of desire-satisfactions. To this end, we shall first introduce the concept of a point-individual and use it to formulate a strengthening of EQUAL.

A point-individual is a brief subjective time-segment of a person – a “moment of consciousness”. It doesn’t matter exactly how long a point-individual lasts. For definiteness, let’s say that a point-individual is a one-minute (subjective) time-segment of a person. (The term “subjective” can be ignored for now; we’ll return to it shortly.)

We can then represent The Wall as in this table:

	<u>Alice</u>	<u>Bob</u>
1.59-2.00	yellow	green

1.58-1.59	yellow	green
...
1.00-1.01	yellow	green
0.59-1.00	yellow	blue
...
0.01-0.02	yellow	blue
0.00-0.01	yellow	blue

Each box represents one point-individual. There are 120 point-individuals desiring yellow, 60 desiring blue, and 60 desiring green. Assuming that the desires of each of these point-individuals have the same ethical weight, then the best action would be to satisfy the desires for yellow. Out of all the available actions, this will maximize the sum of weighted desire satisfaction, and by assumption the example is one where desiderata other than desire-satisfaction are not present or do not discriminate between the available alternatives. So yellow would be the best choice.

How many point-individuals there are with a certain desire depends on two things: how many individuals have that desire and how long each of those individuals holds those that desire. The greater the accumulative duration of a desire that P, the more corresponding point-individuals there are with a desire that P; and the stronger, ceteris paribus, should be the ethical imperative that P be satisfied on theories which accord ethical significance to desire satisfaction. I propose the following as a strengthening of EQUAL:

(P-EQUAL) We should give equal ethical weight to the welfare of point-individuals that are equal in all ethically relevant ways.

Since the number of point-individuals that have a given desire is proportional to the cumulative duration for which that desire is held, we have as a consequence of P-EQUAL the following principle of proportionality:

(PROPT) The ethical weight of a desire is proportional, other things equal, to its subjective duration.

In mathematical terms, we can express the total ethical significance of \underline{P} being the case (or more guardedly, the total of that part of the ethical significance of \underline{P} that is due to \underline{P} 's contribution to overall desire-satisfaction) as:

$$DSV(P) = \sum_{i \in I} R(P_i)$$

Where $DSV(\underline{P})$ is the desire-satisfaction value of \underline{P} 's being the case, I is the set of all point-individuals, and $R(\underline{P}_i)$ is a measure of the ethical weigh of any other relevant aspect of i 's desire that \underline{P} (that remains after we've factored out the duration of \underline{P}_i). $R(\underline{P}_i)$ depends on the strength of the desire that \underline{P} and perhaps on other factors as well, such as how well informed or noble is the desire for \underline{P} . In this paper we generally attempt to try to keep $R(\underline{P}_i)$ constant so that we can focus more directly on the role of time and duration.

We can illustrate this principle on The Wall. Let's say that the satisfaction of one point-individual's desires has an ethical weight of 1 unit. Then the aggregate ethical weight of the desires for yellow is 120 units and the aggregate ethical weights of the desires for blue and green are 60 units each. As far as desire-satisfaction is concerned, it would be ethically indifferent whether we satisfy yellow-desire or both the blue-desire and the green-desire. Yet since we can only paint the wall in one color, we can only satisfy one of these, so we ought to satisfy the yellow-desires.

If we modify The Wall by supposing that there is an additional person, Clare, whose desires are identical to Bob's, then we have a situation where there is a tie between the ethical weights of the desires for yellow, blue, and green (each option has an ethical weight of 120 units). The painter could arbitrarily choose any of these colors. If we add yet another person, Doug, with the same desires as Bob and Clare, then the balance shifts in favor of painting the wall blue or green. According to P-EQUAL, either of these colors would satisfy desires whose aggregate weight is 160 units, whereas yellow would satisfy desires of weight 120 units. P-EQUAL allows us to determine the best choice of color for any combination of desire-durations and number of people having the desires.

7. Do longer-lived individuals get greater ethical weight?

One consequence of PROPT is that, other things equal, a longer-lived person's life-long desires get a greater accumulated ethical weight, as far as desire-satisfaction is concerned, than do life-long desires of a person who doesn't live as long.

We can observe a dramatic example of this if we look at a case where there is a large number of persons who live brief lives and one person who lives for a much longer, such that the combined lifetimes of the short-lived folks equals the lifetime of the long-lived person. Let's suppose that the long-lived person, Alice, has the same desire for yellow throughout her two-hour life, and that each of 120 short-lived persons has a different and separate desire for some specific color C_n for his one-minute life. All desires are equally strong etc. The situation is depicted in the following table:

1.59-2.00	Alice: yellow	Larry: C_{120}
1.58-1.59	Alice: yellow	Hal: C_{119}
...
1.00-1.01	Alice: yellow	Jim: C_{61}
0.59-1.00	Alice: yellow	Peter: C_{60}
...
0.01-0.02	Alice: yellow	Sven: C_2
0.00-0.01	Alice: yellow	Mike: C_1

If duration didn't affect ethical weight, we would have to say that painting the wall yellow would be no better than painting it in, say, color C_1 . That would be counterintuitive. Painting it C_1 would leave the world, after the first minute has passed, one in which no desire is being satisfied. In contrast, painting the wall yellow would lead to as much desire-satisfaction during the first minute as would painting it C_1 , and it would lead to strictly more desire-satisfaction the remaining time.

What PROPT tells us about this case is that the desires for C_1 - C_{120} count for as much combined as does the desire for yellow by itself. Each of C_1 - C_{120} would result in $1/120^{\text{th}}$ of the amount of desire-satisfaction that yellow would achieve. In other words, among the people in this

example, Alice's desire gets a 120 times greater ethical weight than does, for instance, Mike's. This is because Alice lives 120 times longer than Mike, who are ethical equals in all other respects (and thus it is assumed that Mike, being very young all his life, is not more passionate or more innocent, nor does Alice because of her longer life become more mature or develop desires that are better informed.)

Does this mean that longer-lived people have greater ethical weight than short-lived ones? At best, that would be a highly misleading way of putting it. Ethically equivalent desires get the same ethical weight whoever has them. True, a long-lived person's total set of desires will typically attain a greater combined ethical weight than the set of a short-lived person, but this is because they are greater in quantity, not because of whom they belong to per se. Analogously, in times of scarcity, people who live long lives are entitled to more bread over the course of their lives than are people who die young – not because they are more worthy of bread but because they are there, needing to be fed, a greater number of days.⁶

8. Subjective time

I will now introduce the notion of subjective time and give some reasons why this is the relevant measure of the duration of a preference.

Let us begin by saying what subjective time is not. By “subjective time” I do not mean how long the subject happens to think that her desire lasts. One can easily be in error about that, and misestimating how long one has had a desire should not in general affect its ethical weight. Nor do I mean by “subjective time” how long an interval “feels” for the subject, at least not in any straightforward sense. If it feels like time flies when one is happy, that is no reason to discount desires held in merry times.

To explain the concept of subjective time, it may be of help first to consider the notion an upload. In futuristic discussions, an upload is a mind that was originally implemented in a biological brain but has been transferred (“uploaded”) to a computer.⁷ The idea is that mental states supervene on brain states and that some form of computationalism is true, so that one's neuronal network (one's mind) can in principle be implemented in a computer just as well as in an organic brain. Here we are not concerned with the technical or physical possibility of this

hypothesized process⁸, or even primarily with its metaphysical plausibility; we want to use the idea merely as an aid in explaining the notion of subjective time.

The amount of subjective time experienced by an upload in an objective time-interval (say, one hour) is directly proportional to the clock speed of the computer on which the upload is implemented. For example, suppose an upload is running on a computer that is a thousand times faster than its original biological implementation. Such an upload will experience the external world as slowed down by a factor of a thousand. From the viewpoint of a biological human observer, the upload appears to think very quickly. Throw a ball up in the air and before it hits the ground the upload could have finished reading a whole paper by John Searle.

My suggestion is that we should define the concept of a point-individual in such a way the number of point-individuals associated with a mind per hour is (other things equal) proportional to the clock speed of the computational substrate on which it is running. The upload that runs a thousand times faster than biological humans has a thousand times more point-individuals in any given time interval. Upgrade the implementation of the upload to a piece of hardware that's twice as fast, and the upload will have two thousand point-individuals in the same time interval.

It is on this ground that we can claim that the ethical weight of a desire is proportional to its subjective duration. Subjective duration does not depend on anybody's estimate of how much time has lapsed. Instead it is the amount of subjective time – the amount of “experience-time” – that has actually come to pass that counts.

Why count duration in units of subjective time rather than in units of chronological time? Here is one argument.

An upload can be suspended for an indefinite period of time, meaning that the computer on which it is running is halted during some interval, although we may assume that the state of the upload's mind is stored in the computer's working memory. The period during which it is suspended should not add to the ethical weight its desires. If it did, then an upload could, implausibly, make any desire have an arbitrarily great weight simply by going into a suspended state for a sufficiently long period of time. But surely having an upload stored in a computer's memory does not increase the ethical importance of the encoded desires.

A slow upload can be conceived of, equivalently, as a fast upload that takes frequent pauses (one between every clock cycle). We should therefore regard the slow upload in the same way as we regard an upload that alternates between running quickly and pausing, supposing the total numbers of clock-cycles consumed are the same for both uploads. By PROPT, this entails assigning less weight to the desires of a slow upload than to those of a fast upload if their desires for the same objective duration (and assuming that they are equal in any other relevant respect). The faster upload's desires get weights that are greater by a factor equal to the ratio between the speed of the fast upload and the speed of the slow upload.

We can run a similar argument without using the concept of an upload. Suppose you happen to desire that it be ethically important (on a desire-satisfaction view) that you be made king. If it were objective rather than subjective duration that mattered, you could (in principle) achieve this by going into a dreamless slumber lasting for a sufficiently long time. After many, many millennia, your desire might have accumulated enough ethical weight to dominate the combined weight of others' desires that you not be their ruler. Yet no relevant developments took place during this period that could bolster your legitimacy: you had no new wishes, no conscious experiences, there was no increase in the intensity of your desires, nor did anything else happen that could sanctify your monarchical ambition. It would be absurd to maintain that your claim to a kingdom would have become automatically justified while you were sound asleep. Rather, one may regard the situation as ethically equivalent to one in which your brain was destroyed at the beginning of your hiatus and an exact replica reconstructed at the end: so far as the ethical weight of your desires are concerned, you'd simply have to pick up where you left off. – People in cryonic suspension, if they are reanimated in the future, will not be entitled to preferential treatment merely because they have been spending many years in the freezer.

The view presented here does not entail that subjective time stops whenever there is no mental development. Suppose you manage to enter a deep meditative state where for an entire hour your mind is occupied by a single unchanging idea. Does subjective time pass during this interval? It would seem that it does, for even though you remain in roughly the same mental state throughout, there is nonetheless plenty of activity in the brain on the neuronal level. To be continuously aware of something requires that the relevant cortical areas are engaged in sustained activity. This stands in contrast to the case of an upload that is paused, or that of a cryonically

suspended person, in which there is no neuronal activity. In those conditions there is not an unchanging mental state but simply unconsciousness.

Our view also does not imply that “dumb” persons have fewer point-individuals per unit of objective time. Even if a dumb person thinks fewer thoughts in a given time than a quick wit, it is not necessarily the case that subjective time runs slower for the dumb person. For the dumb person’s individual thoughts may be stretched out over a greater amount of subjective time.⁹

By basically the same reasoning we can see why somebody who takes amphetamine and as a result believes that she perceives her mind as being significantly accelerated might well be mistaken if she thought that she had a faster than normal flow of subjective time. For one thing, one of the effects of the drug might be to distort her time perception so that she is mistaken about the amount of mental development that occurs. For another thing, even if she does in fact think more thoughts per hour when she is intoxicated, this could be because she skips more frequently from one thought to the next while spending less subjective time on each thought. There is no reason, therefore, to suppose that PROPT gives crackheads a privileged moral status.

9. An objection and illustration: Simbad the Sailor

Simbad had always wanted a burial at sea. In the last week of his life he changed his mind and decided that he wanted to be buried next to his wife and relatives in the local churchyard. He was fully competent when he made this decision. You are the only person that Simbad has talked to about this and you didn’t promise him anything. You are now in charge of the funeral arrangements. Should you have him buried at sea or in the churchyard?

At first blush it might seem straightforward that if one accepts PROPT one is thereby committed to thinking that it is morally best to give Simbad a sea funeral, as the desire that this option would satisfy has been of much longer duration and hence, by PROPT, is one that has much greater ethical weight than Simbad’s later desire for a land burial. However, this diagnosis oversimplifies the issue.

The reason why Simbad wanted a sea burial for most of his life might be that he thought that would best fulfill his conception of his well-being. He might have thought that being buried at sea was the perfect culmination of a sailor’s life. In the last week of his life he might have got new evidence or simply rethought the issue, and reached the conclusion that the central theme of

his life was not sailing the seas but being a father and a husband, and that the most fitting end to this existence is to be buried next to his family. If that is so, then what he really wanted all along was to have an appropriate ending, and if the views he held in the last week of his life correctly indicate what constitutes such an ending, then by burying him in the local churchyard we shall have satisfied all his sepulchral desires, not only the most recent one.

Even if he didn't earlier want an appropriate ending but had a simple desire to be buried at sea, we might still resist the conclusion that he ought to be given a sea burial, because the ceteris paribus clause in PROPT may not be satisfied. His earlier desire might have been less well informed than his later desire. If well-informedness is one of the factors that determine ethical weight, this could block the inference.

Further, we may perhaps more appropriately conceive of our role in making the funeral arrangements not as that of a benevolent despot, whose task would be to figure out how to maximize Simbad's overall desire-satisfaction, but more humbly as the executor of Simbad's last wishes, the goal being to affect a respectful conclusion of Simbad's own aborted agency.

Additionally, there may of course be all sorts of practical reasons for going with Simbad's later wishes. Most people probably prefer that persons have the prerogative to change their minds about important things up until the moment they die, provided that they remain competent. To override Simbad's last desire may be graceless gesture of paternalistic imposition, which, moreover, sets up an undesirable precedent for the rest of society.

The upshot is that we should be wary of hastily concluding, from prima facie counterintuitive consequences of PROPT in concretely described cases, that the principle is in error. For principle is an abstract one, and when we apply it to real situations, we must be recognize that there may be a multitude of confounding factors or extraneous ethical determinants that influence our intuitions. A desire-satisfactionism that adopts PROPT has many explanatory resources at its disposal to account for the fact that our actual ethical judgments are unlikely to be reducible to a simpleminded application of one short rule.

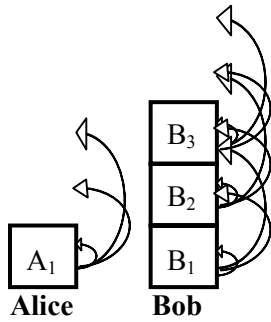
But what if we expand the example at hand by stipulating that none of the proposed possible explanations is correct? Suppose that Simbad did not change his mind about how the various burial-options would best fit his conception of what would make his life go well; and suppose that his later desire was in no way better informed than his earlier desire; and suppose

that there were no general preference for having people's later wishes take precedence over their earlier ones; and suppose any of the other explanations that one could conceive of for why, PROPT notwithstanding, Simbad should be buried in the churchyard, fails likewise. If this were the situation, then PROPT does imply that you ought to bury Simbad at sea. But I think, given these assumptions, this is actually a plausible recommendation to make to someone who intends to base an ethical choice on the exclusive goal of maximizing desire-satisfaction. If there is any lingering counterintuitiveness about this consequence, it might well be due to the fact that when we try to imagine Simbad's case we inadvertently see it as taking place in the real world where the relevant suppositions are almost never satisfied. Our intuitions are shaped by, and therefore calibrated for, the types of situations that we usually encounter, not necessarily for situations that lack an arbitrary collection of features that are characteristic of the real world.

10. Some consequences of PROPT

Long-lived persons contain more point-individuals and will therefore, other things equal, have a greater aggregate ethical weight according to PROPT. We argued earlier that it would be misleading to claim that this fact amounts to giving the desires of long-lived persons more weight than the desires of persons who don't live as long. The greater weight of the aggregate desires of the long-lived is distributed over a greater number of particular desires, so it is not the case that a desire, merely by belonging to a long-lived person, is thereby somehow ethically more significant.

Yet there is a more subtle effect which, given certain natural conditions, does lead to longer-lived persons' desires counting for somewhat more than typical desires of short-lived persons. The desires of a point-individual of a long-lived person will, on average, tend to have a greater weight than a point-individual of a short-lived person. To see this, look at this diagram:



If the arrows represent the concern that people have for their own future well-being, we can see from the picture that most point-individuals of Bob, the long-lived person, have more arrows pointing to them than does the point-individual of Alice, the short-lived person. By satisfying the desires of a late point-individual of Bob (say B_3) we satisfy not only B_3 's desires, but also the desires that B_1 and B_2 have that the desires of B_3 be satisfied.

This phenomenon does not represent an undue bias in favor of those with longer lives. Each segment of each individual is accorded the same ethical consideration. This results in a greater ethical weight for the desires of later point-individuals, since they are normally the beneficiaries of the well-wishes of earlier point-individuals of the same person. Since there are no successors of Alice's first point-individual A_1 , all her desires for her future well-being are unfortunately frustrated. Short of extending her life, there is no possibility of satisfying these desires.

It also follows from these considerations that giving an existing person \underline{x} extra years is ethically preferable, from a desire-satisfaction point of view, to creating a new person that will live for \underline{x} years (assuming the quality of these extra years would be the same).

It is hard to estimate how much people care about their past well-being.¹⁰ But if we assume, as seems plausible, that people generally care more about their future than about their past, then another consequence of PROPT is that it is, ceteris paribus, ethically better if a person's late desires are satisfied than if her early desires are. A life that begins low and ends high is better than one that goes in the opposite direction. This is so even if we set aside the psychological effects of knowing that things are constantly getting better or worse.

If we assume, as is standardly done in economics, that people discount their future welfare exponentially, then the weight of later desires (holding other things constant) would converge to a finite value even if a person lived forever. This is because the desires you had very long ago for your current welfare would be increasingly weak the further back in time they were, and this effect would eventually dominate the linear aggregation over time of past point-individuals. The result is that after a certain period of time (corresponding roughly to a normal human lifespan, if we assume a discount rate of a few per cent per year) the ethical weight of your desires would plateau and would not undergo any significant further increase (unless there were independent factors that increased their weight, such as their being increasingly well-informed).

If we have significant desires for our past well-being, this would entail that even the early desires of long-lived persons have more weight than the early desires of people who are destined to die young. Moreover, if we have significant past-directed desires, the weight of your desires may peak at some point before the last segment of your life. The very last of your point-individuals will not enjoy the retrograde well-wishes of any later point-individuals, so they may have less total weight than point-individuals that occur somewhat earlier in your life, who have temporally proximate (and thus not heavily time-discounted) well-wishing point-individuals both before and after them.

All of these considerations can be relevant for many practical problems, for example in setting medical research priorities or in determining socially optimal saving rates.¹¹ One should bear in mind, however, that any real-world implications are bound to be subtle and complicated. The duration of desires is at best only one of the factors that determine their ethical weight. Further, we may endorse an ethical theory that contains many principles in addition to desire-satisfactionist ones. And of course, there are all sorts of practical considerations that would have to be taken into account before making any specific policy recommendations. It would be much too rash, for instance, to jump to the conclusion that there ought to be a transfer of economic resources from young adults to middle aged adults. Nevertheless, it is progress if get a clearer understanding of the value-structures of at least clean cases in which desire-satisfaction is the sole goal and desires differ only in regard to their durations.

11. Returning to the question of whether latent desires have ethical weight

In light of the above discussion, what can we say about the problem of whether latent desires carry ethical weight? We have argued that the duration of a desire should be measured in units of subjective time for the purpose of determining its ethical weight. We argued that no subjective time lapses for an upload who is in a paused state, or for a human who is in cryonic suspension or a dreamless sleep.

Against this background, we can distinguish two cases. Suppose, first, that a latent desire exists in a chronological time interval within which no subjective time lapses. In this case it follows from what we have said that such a desire carries zero ethical weight. In other words, latent desires that exist only while the subject is unconscious lack ethical significance.

The second case is where a latent desire is accompanied by subjective experience so that subjective time flows while the desire is held. The preceding discussion is consistent with giving such a desire ethical weight. We may, however, question whether it is plausible to do so. Why should it make a difference to whether a particular desire has ethical weight whether the subject was conscious of other things while she had the desire? This seems to be a position on a hard floor between two comfy chairs. It has neither the plausibility of the view that only active desires count nor the simplicity of the view that that all desires count (which we have argued is incorrect). Instead it makes the ethical weight of a desire depend crucially on whether some mental event is taking place while the desire is held – a mental event that can be totally unrelated to the desire! Thus, if Joe has a strong latent desire for a trip to the moon while he is in a dreamless sleep, that desire counts for nothing; but if he has the same latent desire while he is dreaming about a game of chess, the desire counts for a lot. This may seem implausible. If we reject this view, we are driven to conclude that desires that aren't active have no ethical weight.

Now, rejecting the view that latent desires carry ethical weight may cast some doubt on the general view that the good consists in desire-satisfaction, since we are eliminating one way in which that general view could be correct. However, it is worth recalling some of the points we made earlier that suggested that even the limited basis consisting only of active desires may be adequate for grounding a plausible conception of the good. We pointed out that (1) you will often earlier or later have active desires for things that you currently have only passive desires for, so depriving you of these things now could be wrong because it would frustrate those earlier or later

active desires; and (2) many things for which you don't have a specific active desire have instrumental utility for you (they can help you achieve things for which you do have, or have had, or will have, an active desire). These two points go some way towards allaying the worry that a desire-satisfactionist might have: that once we restrict ethical concern to active desires then too few things get regarded as ethically valuable. And, of course, for views according to which desire-satisfaction is only one part of the good, there are additional sources of ethical value.

Assuming, then, that desires that are merely latent don't have ethical weight, there is a further step that might suggest itself. We may wonder: if a desire gets ethical weight precisely when the subject is conscious of it, may it not be that the real source of ethical weight is the quality of the subject's experiences rather than the satisfaction of her active desires? This conclusion does not follow logically, but the preceding discussion may make it a more attractive move. If we take this route, we would have been led from the view that all desires count to the view that only active desires count to the view that only the quality of experiences counts (though not necessarily all the way to hedonistic utilitarianism). In this paper we won't go so far as to argue for this view. We will just note that the line of argument that we have pursued may tempt some to proceed in that direction.

Acknowledgments

I'm grateful to many for helpful discussions, and especially to Shelly Kagan.

References

- Bentham, J. (1970). An Introduction to the Principles of Morals and Legislation. London, The Athlone Press. (Originally published in 1780.)
- Caplin, A. and Leahy, J. (2000). "The Social Discount Rate." National Bureau of Economic Research, Working paper #7983.
- Drexler, K. E. (1985). Engines of Creation: The Coming Era of Nanotechnology. London, Forth Estate.
- Feldman, A. (1980). Welfare Economics and Social Choice Theory. Boston, Martinus Nijhoff Publishing.
- Lewis, D. (1986). Philosophical Papers, Volume II. New York, Oxford University Press.

- Merkle, R. (1994). "The Molecular Repair of the Brain." Cryonics 15(1 and 2).
- Moravec, H. (1989). Mind Children. Harvard, Harvard University Press.
- Ng, Y.-K. (1990). Social Welfare and Economic Policy. New York, Harvester Wheatsheaf.
- Parfit, D. (1984). Reasons and Persons. Oxford, Clarendon Press.
- Taylor, E. F. and Wheeler, J. A. (1992). Spacetime Physics: Introduction to Special Relativity, 2nd ed., W. H. Freeman & Company.

¹ See e.g. (Lewis 1986) for an introduction.

² See e.g. (Feldman 1980; Ng 1990).

³ (Bentham 1970 [1780]).

⁴ See e.g. (Taylor and Wheeler 1992).

⁵ Note, however, that this kind of argument does not (at least not in any straightforward way) rule out theories of individual decision-making that gives a person's own desires different weight depending on whether the desires are in the past or the future. For in this case there is a natural preferred reference-frame relative to which one should interpret these temporal concepts: namely, the person herself. We cannot settle the controversy among these theories by appealing to the theory of relativity. It is when you have to aggregate changing attributes of people or objects that are traveling along a different trajectory from your own that the problem referred to in the text arises.

One view that is unaffected by this problem is that only a person's last desires (the wishes one has at one's death bed) carry ethical weight. However, I don't think that anybody espouses this position.

⁶ One could develop a version of the theory we are exploring that would factor out this effect by dividing the ethical weight of a person's desire by the length of that person's life. Similarly, one could add a component of egalitarianism to desire-satisfactionism by adding the rule that the overall amount of good in the world is a function of not only the total amount of desire-satisfaction (evaluated according to the prescriptions set forth in this paper) but also of how evenly distributed this desire-satisfaction is between different persons (or even between different desires!) We will not develop either of these theoretical options here. (They seem to address different issues than the role of time in desire-satisfactionism.)

⁷ See e.g. (Drexler 1985; Moravec 1989).

⁸ One method of uploading a human mind that would seem to be possible given sufficiently advanced technology is as follows: (1) Through continued progress in computational neuroscience, create a catalogue of the functional properties of the various types of neurons and other computational elements in the human brain. (2) Use e.g. advanced nanotechnology to disassemble a particular human brain and create a three-dimensional map of its neuronal network at a sufficient level of detail (presumably at least on the neuronal

level but if necessary down to the molecular level). (3) Use a powerful computer to run an emulation of this neuronal network (such a computer could be built with molecular nanotechnology). This means that the computations that took place in the original biological brain are now performed by the computer. (4) Connect the emulated intellect to suitable input/output organs if you want it to be able to interact with the external world. Assuming computationalism is true, this will result in the uploaded mind continuing to exist (with the same memories, desires etc.) on its new computational substrate. (The intuitive plausibility of the scenario may be increased if you imagine a more gradual transformation, with one neuron at a time being replaced by a silicon microprocessor which performs the same computations. At no point would there be a discontinuity in behavior, and the subject would not be able to tell a difference; and at the end of the transformation we have a silicon implementation of the mind. For more detailed explanations, see e.g. (Merkle 1994).

⁹ One can compare this to the case of a smart human being uploaded and run on a slow computer. Such a mind would have the same experiences (assuming no interaction with the external world) as if it were running faster; only, its experiences would be more “spread out”. That is to say, the physical processes on which the experiences supervene would cover a longer stretch of objective time.

¹⁰ For some relevant thought experiments, see (Parfit 1984), especially chapter 8.

¹¹ See (Caplin and Leahy 2000) for an argument that uses our conclusion as a premiss and attempts to show that governments ought to adopt a lower social discount rate and ought to subsidize savings and generally promote future-oriented policies