

## Journal of Philosophy, Inc.

---

Prudence

Author(s): Phillip Bricker

Source: *The Journal of Philosophy*, Vol. 77, No. 7 (Jul., 1980), pp. 381-401

Published by: [Journal of Philosophy, Inc.](#)

Stable URL: <http://www.jstor.org/stable/2025536>

Accessed: 18-08-2014 20:55 UTC

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*Journal of Philosophy, Inc.* is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Philosophy*.

<http://www.jstor.org>

PRUDENCE.\*

**W**E sometimes have occasion to evaluate a person with respect to his life as a whole. We might, for example, wish to say of a person after his death that he had lived a most prudent life, that he had always acted so as to best coordinate and satisfy his past, present, and future desires. To the extent that we ourselves value prudence—desiring that our lives live up to their full potential, developing as integral wholes—we may ask, when faced with an array of alternative actions: which is the most prudent act? which act would contribute the most to my life?

In this paper, I will attempt to outline an analysis of prudence. The analysis seeks to apply the concept of prudence primarily to a person with respect to his life as a whole, and only derivatively to the individual acts that make up that life. Partly for this reason, the theory of prudence that I will present could be applied by an agent at the time of acting only if he were extraordinarily knowledgeable about the facts—past and future, actual and possible—that are relevant to his life. The chief purpose of this paper is not to argue for specific prudential maxims that would be useful to an agent at the time of acting, but rather to reveal what I take to be the basic structure underlying any adequate theory of prudence. However, as I hope some of the examples throughout this paper will suggest, the theory of prudence here presented is not without practical implications for a normative theory of decision-making.

I do not wish to claim that the concept of prudence that will emerge from my analysis will coincide with any of the uses of the term 'prudence' in ordinary language: ordinary usage is beside the point when it comes to evaluating the interest of the analysis. In

\* I would like to thank David Lewis and Derek Parfit for their helpful comments on an earlier draft of this paper.

particular, I will never use the term 'prudence', as is common in ordinary language, to refer to a psychological disposition of the agent; to be prudent, in the sense used here, is to act in a certain way, not to be disposed so to act. Moreover, the prudent act, according to the present theory, is the act that would *in fact* turn out best for the agent, whether or not the agent is or ought to be in a position to know which act this is. My discussion of the concept of prudence probably owes more to recent philosophical discussions of the concept of prudence<sup>1</sup> than it does to the practice of ordinary language. This paper differs from such discussions, however, in being wholly unconcerned with the question whether it is uniquely rational—or, indeed, rational at all—to act prudently.

## I

Let us begin with an informal exposition of some of the chief features of the concept of prudence. I take the mark of a prudent person to be roughly this: that he acts so as to get what he wants, has wanted, or will want. On the present approach, it will follow that it is also the mark of a prudent person that, whenever possible, he makes himself such as to want what he gets, has gotten, or will get. By effecting changes in himself and his environment, he attempts to make the most of his life; or, as I will develop the notion, he attempts to maximally satisfy his preferences. He creates for himself, via his actions, the best life that his circumstances will allow.

By whose standards shall we judge which life is the best life? Can we impose our own standards of the good in evaluating the prudence of another? Consider a person who has succeeded, by means of foresight and planning, in maximally satisfying his desires. However, first and foremost among those desires has been the desire to acquire great wealth. Now, it may be the case that, according to *my* concept of the good, no person who has devoted his life to the pursuit of riches can be said to have had a good life. You, on the other hand, yourself valuing great wealth above all else, note that his life has conformed in all respects to *your* concept of the good, and unhesitatingly say he has had a good life. But although we may disagree, using our own various standards of the good, as to whether or not he has had a good life, we should both be able to agree that he has *led a prudent life*; for a judgment concerning his prudence should be unaffected by how well his life has conformed to any external standards of the good. Whether or not he was prudent must be determined from the standpoint of *his own* concept

<sup>1</sup> In particular: Thomas Nagel, *The Possibility of Altruism* (New York: Oxford, 1970), pt. II; Derek Parfit, *Against Prudence* (unpublished manuscript, 1977).

of the good. Thus, in asserting that a person led a prudent life, we make not an evaluative judgment, but rather an objective claim—at least to the extent that his concept of the good is itself objectively determinate.

But the problem immediately arises: how are we to characterize a person's concept of the good? We are constantly changing throughout our lives, and, as we change, our system of values changes as well. Thus, a person's concept of the good at one time may be radically in conflict with his concept of the good at another time. It may be possible to determine, given a person's preferences at a certain time, what his concept of the good is *at that time*. But, given the fact that our preferences change over time, how might it be possible to determine a person's concept of the good, *simpliciter*? Our notion of prudence, as an attempt to evaluate a person with respect to his life as a whole, clearly requires that such a determination be possible.

Consider the following example. A man in his youth sets out various goals for himself, and, in the course of his life, succeeds in attaining them all. But as he enters old age, he looks back upon all his earlier activity with disgust and regret; he now believes that he has wasted his youth upon vain pursuits. How shall we evaluate this man with respect to prudence? In terms of the concept of the good that he had early in life, all his actions were successfully coordinated over time so as to provide the life that he then wanted for himself. Considering only this perspective, it might seem that his life has been prudent. But, from the perspective of the old man, his life has been grossly imprudent; for it has been spent in pursuit of worthless goals. Neither view, by itself, is acceptable as an overall evaluation of his prudence. A theory of prudence must attempt to take both perspectives into account in arriving at a compromise decision: his life would have been more prudent if he could have acted so as to satisfy the preferences of both his earlier and later selves (perhaps by changing himself into a later self who could accept the goals of his earlier self); but his life would have been even less prudent than it was if he had failed to act in accordance with the preferences of his earlier self, as well as failing to act in accordance with the preferences of his later self.

Thus a theory of prudence must be able to arbitrate the competing claims of past, present, and future selves. Different acts would be recommended as best by different selves; if a theory of prudence is to be able to direct the agent to do that act which is best, not for

this self or that self, but for the agent timelessly considered, it must provide a method for amalgamating the various preference rankings of the earlier and later selves into a single, timeless preference ranking. The resulting system of preferences may not coincide with any actual system of preferences that the agent has had at any time; in this sense it provides an abstract and artificial perspective. But it provides the perspective from which the agent would wish to view his life if he were, at any time, motivated to perform the most prudent act, the act that would contribute the most to his life over all.

But it is not enough for a theory of prudence simply to direct the agent to act in accordance with his timeless preferences. For in deciding which act to perform, the agent is, at least in part, deciding who he will become; and by performing different acts, the agent can, in a sense, become different persons, with different systems of timeless preferences. Thus a theory of prudence cannot, on threat of circularity, direct the agent to perform that act which maximizes his timeless preferences, since it is in part by performing one act or another that those timeless preferences are created. In deciding which is the prudent act, the agent must be able to compare, for each of the acts he might perform, the preferences he would have as a result, and the extent to which those preferences would be satisfied. How such comparisons should be carried out represents one of the chief problems for an analysis of prudence.

All this talk of abstracting an agent's timeless preference pattern from his temporal preference patterns, and of comparing the degree of satisfaction of different possible timeless preference patterns, would make little sense if we did not assume that, at least in theory, an agent's preferences can be numerically represented. In this paper, I will make explicit use of a calculus of utilities. The air of precision that this lends to the analysis is somewhat bogus, since it is an idealization to suppose that a unique utility function can be assigned to an agent at a time. But stating the theory in terms of a hypothetical calculus has advantages over stating the theory in vague terms: it enables the assumptions, the structure, and the implications of the theory to be clearly displayed.

## II

Let me begin the more formal part of the presentation by setting out the basic constituents of a theory of prudence. The theory will enable one to determine, given an agent *A* inhabiting a world *w* (perhaps ours), whether or not *A* had been ideally prudent. This will be done by providing a ranking with respect to prudence of all

the persons  $A$  could have been, had  $A$  acted differently—in short, of all the *counterparts* of  $A$ .<sup>2</sup> Then,  $A$  was ideally prudent just in case he was at least as prudent as any of his counterparts.

Let us suppose that between  $A$ 's birth and death,  $A$  has had the opportunity to exercise his free will (where  $A$ 's freedom is taken to be compatible, if need be, with physical determinism). That is, there exists a series of times at each of which  $A$  could have acted differently than he in fact did. To any such time  $t$ , we associate the set of acts  $X_{wt}$  that  $A$  could have performed at time  $t$ . The set  $X_{wt}$ , as conceived of here, must satisfy the following three conditions: (1) it contains at least two members; (2) its members are mutually incompatible; that is,  $A$  can perform at most one of the acts in  $X_{wt}$ ; (3) its members collectively exhaust all the possibilities; that is,  $A$  must perform at least one of the acts in  $X_{wt}$ . The sequence of acts that  $A$  *did* perform (in  $w$ ) (it is a function that assigns to each member of the above-mentioned time series a member of its associated set of acts), we will call  $A$ 's *life-strategy* (in  $w$ ).

Furthermore, let us make the simplifying assumption that, except for the cases in which  $A$  acted freely, the world was completely determined. Thus, by adhering to the life-strategy that he did,  $A$  completely determined which possible world was to be the actual world (and similarly for all of  $A$ 's counterparts).<sup>3</sup> Given this assumption, we can say:  $A$  was *ideally prudent* if the life he had in the world he actualized (in  $w$ ) was better than the life he would have had, had he actualized any other world, by adhering to a different life-strategy. More succinctly,  $A$  was ideally prudent if he adhered to the best life-strategy that was available to him. In order to develop this notion of ideal prudence, we must characterize the set of life-strategies available to an agent, and moreover, provide a means for

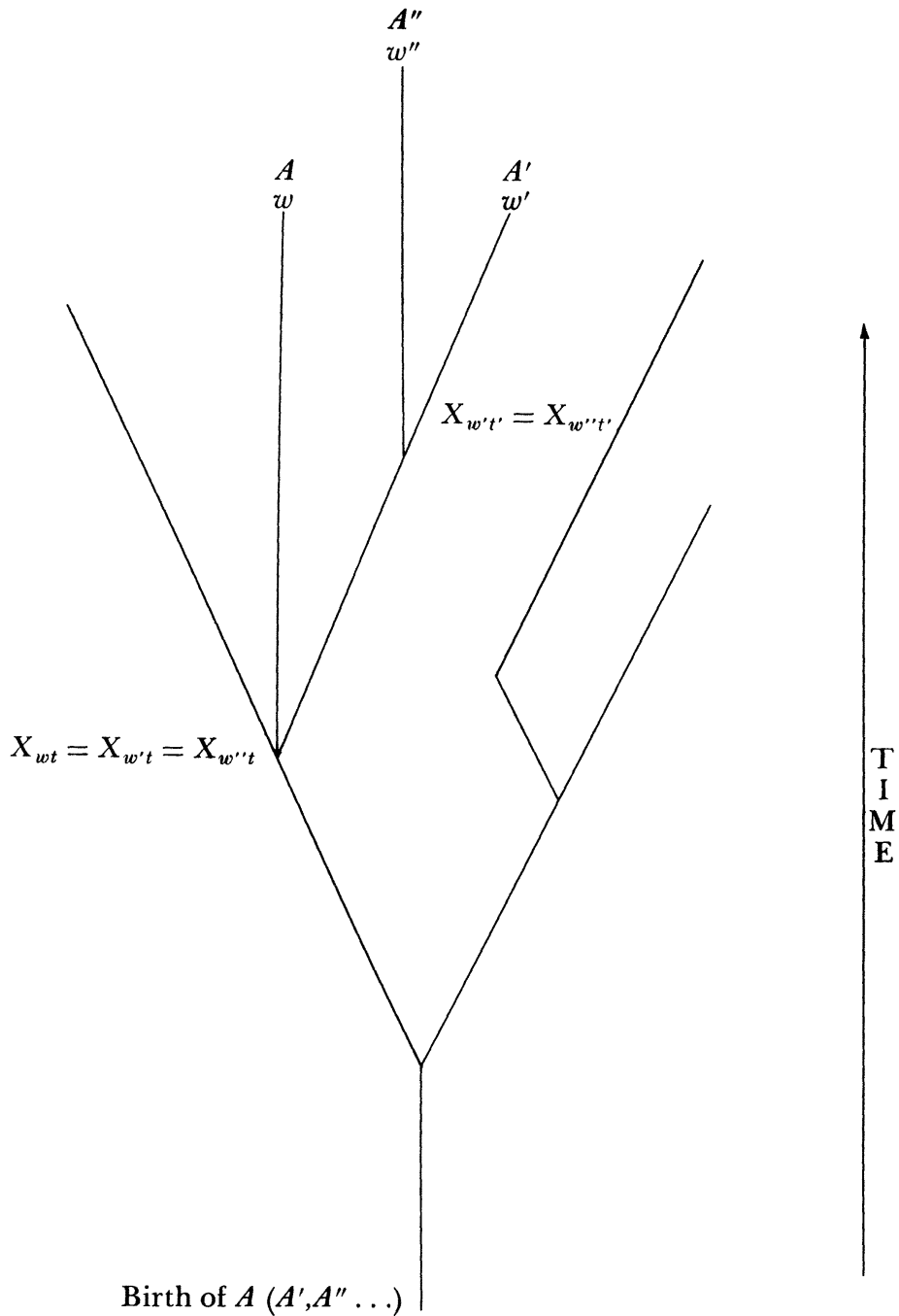
<sup>2</sup> This way of speaking, which is convenient for our present purposes, was introduced by David Lewis in "Counterpart Theory and Quantified Modal Logic," this JOURNAL, LXV, 5 (March 7, 1968): 113–126.  $A$  inhabits only the world  $w$ ; the persons  $A$  would have been had the world been different are  $A$ 's counterparts, not  $A$  himself. The use of counterpart theory carries no metaphysical weight for the present theory, and could easily be avoided; in particular, it is assumed throughout that the relation that an agent bears to his counterparts is an equivalence relation.

<sup>3</sup> This idealization allows us to keep within the province of "decision making under certainty," thereby much simplifying the presentation without essentially affecting the central features of the analysis. Generalizing the analysis to include risk is more or less routine: probability distributions over chance outcomes are introduced, and then, instead of utility, the agent is directed to maximize *expected* utility. Compare, for example, R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: Wiley, 1957), chs. 2 and 3.

evaluating which of these life-strategies is the best. I will attend to these in turn.

$A$  could have adhered to a different life-strategy than he did. For example, consider a time  $t$  which has associated with it a set of acts  $X_{wt}$ ; and consider a possible world in which a counterpart of  $A$  (who is identical with  $A$  in every respect, up until time  $t$ ) performs an act from the set  $X_{wt}$  different from the act  $A$  had performed. Call this world  $w'$  (it is identical with  $w$  in every respect, up until time  $t$ ), and the counterpart of  $A$ ,  $A'$ . The life-strategy that  $A'$  adhered to in  $w'$  is determined in exactly the same way as  $A$ 's life-strategy in  $w$  was determined; it is the sequence of acts that  $A'$  performed in  $w'$ . Continuing in this manner, we may consider a time  $t'$ , later than  $t$ , such that  $A'$  acted freely in  $w'$  at  $t'$ . Then there is another possible world  $w''$ , and another counterpart of  $A$ ,  $A''$ , who performs a different act at  $t'$  from  $A'$ 's act at  $t'$ , and thus has adhered to a different life-strategy from  $A'$ 's. And so on, as long as there are later times at which the counterpart of  $A$  in question acted freely. In this way, a branching tree structure is generated, in which each path through the tree represents a distinct possible world, to which corresponds a distinct counterpart of  $A$ , and the distinct life-strategy to which that counterpart of  $A$  adhered (see diagram). Let us call this entire structure,  $A$ 's *life-tree*, and the possible worlds represented within it,  $A$ 's *available worlds*. A life-strategy is one of  $A$ 's *available life-strategies* if  $A$  could have adhered to that life-strategy; that is, if  $A$  has a counterpart who adheres to that life-strategy.  $A$ 's life-tree,  $A$ 's available worlds, and  $A$ 's available life-strategies are equally the life-tree, available worlds, and available life-strategies of  $A'$ , and  $A''$ , and so on, for all of  $A$ 's counterparts.

Which of  $A$ 's available life-strategies is the best life-strategy? Before we can answer this question, we must first develop more precisely the notion of an agent's temporal and timeless preferences. Let us return to agent  $A$  in world  $w$  at time  $t$ . Presumably,  $A$ 's preferences pertain not only to present states of affairs, but also to states of affairs that did or will, might have or might still, hold. For example,  $A$  might prefer, at time  $t$ , that he have steak instead of fish sticks for his next meal; or he might prefer (better English: wish) that, at a certain past time, he had been in California instead of being in New Jersey, as he actually was. Moreover,  $A$  may have preferences over events not occurring within his lifetime; for example, he might prefer having his ashes rocketed to the moon, rather than having them disposed of in some more mundane fash-



*A*'s life-tree. Associated with each node of the tree is a set of alternative acts. Each point of the tree can be named by giving a world coordinate (not necessarily unique) and a time coordinate (unique) (here ignoring the possibility of time travel).

ion. *A* may even have preferences pertaining to worlds that are not represented within his life-tree—purely idle preferences; for example, disliking loud noises as he does, *A* may prefer worlds that



were not created with a big bang to those worlds which were so created.

In general, it seems possible to form preferences over whatever sort of entity we can conceive of. For the purposes of the present theory, it will be most convenient to idealize somewhat our powers of conception and consider exclusively preferences over possible worlds. Whether these preferences over worlds are best taken as basic, or as themselves derived from preferences over entities of some other sort—say, propositions—need not concern us here.<sup>4</sup> All that is important for our present purposes is that the set of worlds over which an agent's preferences range includes at least all his available worlds, and thus all the worlds that are directly relevant to his deciding which acts to perform. Thus, let us assume that the temporal preference ranking that an agent *A* holds at time *t* is given by a weak order relation<sup>5</sup>—the relation of being at least as preferred as—over an appropriate set of possible worlds. In general, each point (*w',t'*) of *A*'s life-tree will have associated with it a weak order relation over this set of worlds, which represents the temporal preference ranking of *A*'s counterpart *A'*, at *t'*.

For an agent who is only concerned to be able to choose that act which best satisfies the preferences that he holds *at the time of acting*, only the preference ranking over worlds that he holds at that time need be taken into account: he should choose that act which is part of the life-strategy that actualizes the most preferred world still available to him at that time. But for an agent who wishes to perform the most prudent act, the preferences that he *has* held and that he *will* hold, in addition to the preferences that he holds at the time of acting, must be taken into account. If these different temporal preference rankings conflict as to which worlds are preferred to which, a method must be provided for combining these preference rankings into a single preference ranking over worlds with respect to which a choice of action can be made.

But before discussing how such conflicts should be resolved, it would be well to consider just what sorts of case, according to our theory, will involve genuine conflicts between the preferences of an agent's earlier and later selves. For many *prima facie* conflicts between the desires or interests that an agent has at different times

<sup>4</sup> The relationship between preferences over possible worlds and preferences over propositions is briefly discussed in Richard C. Jeffrey, *The Logic of Decision* (New York: McGraw-Hill, 1965), pp. 198/9.

<sup>5</sup> A weak order is transitive and connected, but need not be anti-symmetric—that is, ties are permitted.

will not be represented as genuine conflicts within a theory that is concerned exclusively with preferences over entire worlds. The distinction between apparent and genuine conflicts corresponds to the distinction between desires that are, and desires that are not, conditional upon their own persistence.<sup>6</sup> To develop this distinction, let us consider an example of a desire for a future object which is conditional in this way. I am in a restaurant about to order dinner. I now desire that I have dessert after the meal. However, although I may not be aware of this now, after eating the huge meal that I am ordering, I will no longer want the dessert. Do I now prefer a world in which I get dessert after a big meal to a world in which I don't get dessert after a big meal (all other things being equal)?

If my desire now for dessert is conditional upon my wanting it then, we are not led to the conclusion that I now prefer a world in which I get the dessert. For the preferences that I have now involve comparisons between entire worlds, in which are contained not only my getting or not getting the dessert, but also the desires that I would have at that time. Thus my preference is not simply between dessert and no dessert, but between dessert, not wanting it, and no dessert, not wanting it. If my desire for dessert now is conditional upon my wanting it then, I now prefer a world in which I don't get dessert to a world in which I do get it, given that in both worlds I will no longer want it (and that there are no other relevant differences between the two worlds). In this way, present conditional desires that clash with future desires are struck out, and no conflict between the preferences of present and future selves arises.

But not all desires are conditional in this way, and thus *bona fide* conflicts between an agent's different temporal preferences may arise. Consider the case of a man whose entire life has been ruled by the single passion to produce a masterpiece, and thereby attain posthumous fame. Suppose that he has indeed succeeded in producing such a work, but that just as it is finished he is struck by a deadly disease. As his last act he must choose between the following two alternatives: either to send the manuscript to a publisher, thus ensuring for himself posthumous fame; or not sending it, in which case it is doomed to remain undiscovered. Further suppose that just at this last moment of his life he has decided (neither whimsically nor frenetically) that he cares nothing about posthumous fame, and so now prefers not to publish his manuscript. Is there here a gen-

<sup>6</sup> This distinction is drawn, to a somewhat different end, in Parfit, *op. cit.*, p. 49.

uine conflict between the preferences of his earlier and later selves which would require a theory of prudence for its resolution?

If the preferences of his earlier selves with regard to the publication of the manuscript were all conditional upon their own persistence, then it would be prudent not to publish. For by not publishing he would satisfy his preferences at the time of acting, and would not fail to satisfy any earlier preferences, since his earlier preferences to publish the manuscript could be satisfied only in those possible worlds, if there are any, in which he continues to have those preferences. But it seems extremely unlikely that, in this case, the preferences of his earlier selves were conditional upon their own persistence. For it seems likely that, if asked in his youth whether the consummation of his life's work should be left to the decision of his altered later self, he would have answered: "definitely not!"—and this clearly contrasts with the dessert case considered above. Thus, in this case there appears to be a genuine conflict between the preferences of the man's earlier and later selves. If to be prudent is to view one's life from a temporally neutral standpoint (as I have claimed), then a theory of prudence would direct the man to take into account the unconditional preferences of his earlier selves.

(However, let me mention in passing a somewhat different conception of prudence from the one we have been considering here. On this other conception, an agent acts prudently if he acts so as to satisfy his present and future preferences; his past preferences are deemed completely irrelevant. Thus, in the case at hand, although it would have been prudent for the man to have taken into account the preferences of his dying future self when he decided, as a youth, to devote himself to writing the manuscript, it is now unequivocally prudent not to publish the manuscript, since this accords with all the preferences of his present and future selves. This conception of prudence might be called *future-directed* prudence, to distinguish it from the *bidirectional* conception of prudence being developed in this paper. The analyses given below could easily be adapted to handle future-directed prudence, although I will not take this step.)

Now, if a theory of prudence is to require the agent to take into account the conflicting preferences of his earlier and later selves in deciding which act to perform, then some means of weighing these conflicting preferences must be provided. This cannot be done simply by allowing those preferences to rule supreme which the

agent has held for the longest amount of time; it is not only how long an agent has held a preference, but also how strong that preference is, which should determine how much it is to count in the agent's decision as to which act to perform. Thus, if we are to be able meaningfully to compare the preferences of an agent's earlier and later selves, we must record not only the ordering of his preferences, but the strengths of his preferences as well.

To this end, let us unabashedly introduce, for an agent  $A$  at time  $t$  in world  $w$ , a real-valued utility function  $U_{wt}$  which assigns a numerical value to (at least) all  $A$ 's available worlds. These utility functions are to replace the ordinal preference rankings introduced earlier; they go beyond such rankings by including *how much* the agent (at the time in question) would prefer inhabiting one world as opposed to another, and not just *that* he would prefer inhabiting one to the other. Moreover, since the present theory will require that we be able to compare the utility assignments to worlds which an agent makes at different times throughout his life, we must assume that the zeros and units of the agent's various temporal utility functions can be meaningfully correlated with one another; that is, we must assume the meaningfulness of intertemporal comparisons of utility. Most likely, no amount of actual data about an actual agent could be sufficient to single out uniquely one such utility function for the agent at a time, both because worlds are generally too large for an agent to contemplate in their entirety and because there are too many real numbers to serve as possible values. But the process of deliberation for idealized agents who would have such utility functions is, I believe, similar enough to the process of deliberation in our own case to lend our theory practical significance.

Given a temporal utility function  $U_{wt}$  for each time  $t$  throughout  $A$ 's life, we can construct a timeless utility function  $U_w$  which is to represent numerically the preferences over worlds that  $A$  holds in  $w$ , not at any particular time, but considered from a timeless perspective. This is most easily done by letting the timeless utility of a world  $w'$  be the sum<sup>7</sup> of all the temporal utilities assigned to  $w'$  by  $A$ 's temporal utility functions. Thus:

$$(1) \quad U_w(w') = \sum_t U_{wt}(w')$$

<sup>7</sup> More precisely: the (Lebesgue) integral. For it has not been assumed that an agent has only finitely many distinct temporal utility functions throughout his life. But here, and in the definitions below, summation signs will be used to provide a more familiar representation of the structure of the theory.

where  $t$  ranges over all times between the birth and death of the agent.

But there is a problem. A theory of prudence which directed the agent to maximize his timeless utility as defined by (1) would be invoking the *total view* of prudence. But how does this compare with the *average view* of prudence, according to which an agent's timeless utility function is defined by letting the timeless utility of a world be the average (not the sum) of all the various temporal utilities assigned to that world throughout the course of the agent's life? Thus:

$$(2) \quad U_w(w') = \frac{\sum_t U_{wt}(w')}{t_{dw} - t_{bw}}$$

where  $t_{dw}$  and  $t_{bw}$  are respectively the times of the agent's death and birth in  $w$ . The total view and the average view need not agree in their comparative evaluations of two worlds in cases where the agent lives longer in one world than in the other.<sup>8</sup> When they disagree, the total view gives the wrong result. For example, consider the case of a man who, after leading a most successful and accomplished life, has just discovered that he has an incurable cancer. He is faced with the choice of either suffering a final year of mental and physical deterioration, thereby actualizing world  $w_1$ , or taking his own life, thereby actualizing world  $w_2$ . Let us further suppose that he has always felt, and will always feel, that there is no value whatsoever in the prolongation of life, if the extra time must be spent undergoing the sort of deterioration that he now faces; that is, according to all his temporal utility functions,  $w_2$  has a higher utility than  $w_1$ . Clearly, the prudent act, the act that would best satisfy the man's preferences, would be an act of suicide. But the total view of prudence, which defines the man's timeless utility function by (1), will often arrive at the opposite conclusion (depending upon the exact figures involved); for the total timeless utility that  $w_1$  would have to him will often be greater than the total timeless utility that  $w_2$  would have to him. Roughly, the total view goes wrong by allowing  $w_1$  to continue scoring points, even during the last year of suffering, off the successes of the agent's past.

<sup>8</sup> At least this is true for the analysis I will accept below, Analysis 2; for Analysis 1, the distinction between the total view and the average view collapses. The argument that follows tacitly assumes Analysis 2, and so may not be fully intelligible at this point.

The average view, on the other hand, gives the correct result in such cases.<sup>9</sup>

Let us then adopt (2) as our definition of an agent's timeless utility function. For an agent  $A$  in  $w$  (and similarly for any counterpart  $A'$  of  $A$  in  $w'$ ), the function  $U_w$  ( $U_{w'}$ ) assigns to each available world, and thus to the available life-strategy that corresponds to that world, the value or utility to  $A$  ( $A'$ ) of living in that world or of adhering to that life-strategy. Thus (2) provides us with the means to return to the question: which of  $A$ 's available life-strategies is the best life-strategy? and thereby: did  $A$  lead an ideally prudent life?

### III

It may seem that we are just about finished. For we have said all along that we wanted an analysis of prudence to capture the idea that an agent lived an ideally prudent life just in case he acted so as to maximally satisfy his (timeless) preferences; that is, via (2), just in case he actualized a world with maximal (timeless) utility. But, depending upon how this idea is unpacked, different analyses of prudence will result. Let us first consider the following proposal:  $A$ 's best life-strategy is that life-strategy which, if  $A$  had adhered to it, would have actualized the world that ranks highest according to  $A$ 's timeless preferences. Then,  $A$  was ideally prudent just in case the world that he inhabits is judged by him to be the best of all available worlds. More precisely, we have:

Analysis 1: Agent  $A$  in world  $w$  lived an ideally prudent life if and only if, for all worlds  $w'$ ,  $U_w(w) \geq U_w(w')$ .

However, if we examine the consequences of Analysis 1, I think we can see that it fails to capture adequately the concept of prudence. For, according to it, an agent  $A$  is ideally prudent if his life is judged, relative to his *own* system of preferences, to have been better than the lives of any of his counterparts. But this fails to take into account how  $A$ 's counterparts evaluate their own lives. The timeless preferences of one of  $A$ 's counterparts might result in a ranking of the set of available worlds which was radically different from  $A$ 's own ranking (witness the case where  $A$  had to decide whether or not to submit himself to a brainwashing episode that

<sup>9</sup> This can be checked by running the following figures through Analysis 2 below: the man lives 51 years in  $w_1$ , 50 years in  $w_2$ ; at all times in both worlds he assigns a value of 100 to  $w_1$  and 101 to  $w_2$ . Moreover, for simplicity, suppose that he has one temporal utility function for each year of his life. Only the average view gives the correct answer.

would have changed many of his values and beliefs). The fact that *A* himself judges that he has lived the best life possible does not in itself seem sufficient to establish *A*'s prudence; indeed, perhaps all *A*'s counterparts equally judge that they themselves have lived the best life possible, in which case, according to Analysis 1, they must all be judged ideally prudent. But although they would all be judged ideally prudent, it is possible that some had their preferences satisfied by the world in which they lived (their world yielded a high positive utility according to their timeless utility function), whereas others did not have their preferences satisfied at all (their world yielded a negative utility according to their timeless utility function); for, according to Analysis 1, for each counterpart to be judged ideally prudent requires only that each, even an unsatisfied one, prefers the life that he had to the life that any of his counterparts had.

This difficulty can be brought out more sharply by imagining that *A*, at the beginning of his life, must choose between the following two alternatives. (1) *A* can take a pill that would have two effects. First, it would cause him to become the willing slave of a wicked master: *A* would never again have the opportunity to exercise his free will, and he would have a miserable life, a life with negative utility according to his timeless preferences. And, secondly, the pill would give *A* a *bias toward the actual*: it would cause him always to prefer the actual to the possible, always to prefer what does happen (in his life as a slave) to what could have happened (had he never taken the pill). (2) *A* can choose not to take the pill, thereby securing for himself, regardless of how he may choose to act in the future, a life that has a positive utility according to the preferences he has in that life.

According to Analysis 1, *A* can ensure that his life be ideally prudent by taking the pill. If he takes the pill, although he will lead a life of misery, his bias toward the actual will guarantee that, according to his own timeless preferences, the utility of any other possible world that he might have actualized will be even less than the negative utility of the world in which he lives. But this seems wrong: the mere fact that *A* prefers his own world to any other world does not show that he has led an ideally prudent life. Indeed, there might be a race of men so constituted that, as a result of some inborn psychological trait, each member of the race always preferred the world in which he lived to any other world in which he might have lived; but it should not follow from an analysis of prudence that these men could not fail to be ideally prudent.

Thus Analysis 1 goes wrong in having the evaluation of an agent's prudence depend entirely upon his actual preferences, without taking into account the preferences of his counterparts. But this leads to a further problem. Since the counterparts might disagree as to how they rank, according to their timeless preferences, the set of available worlds, the notion of prudence embodied in Analysis 1 is such that the counterparts might disagree as to the relative prudence of their fellow counterparts. Thus Analysis 1 cannot in general provide a single ranking of the counterparts with respect to their prudence upon which all the counterparts can agree. But this is exactly what an analysis of prudence must provide if it is to be able to determine, given a set of acts that an agent could perform at a given time, which of these acts is the most prudent act, that act which is part of the best life-strategy available to the agent at that time. For consider an agent who must, as his last act, choose between actualizing world  $w_1$  and world  $w_2$ . If the counterpart who inhabits  $w_1$  and the counterpart who inhabits  $w_2$  disagree as to which of  $w_1$  and  $w_2$  is the better world, then Analysis 1 cannot provide the agent with a directive as to which act to perform. For the agent cannot determine which is the more prudent act relative to his own timeless preferences—as Analysis 1 would have him do—because he doesn't yet know which counterpart he is to become, and thus which preferences he is to have. Indeed, this is the very thing that his act is to decide. Thus, Analysis 1 would be useless to the agent at the time of acting.

If an analysis of prudence is to avoid the drawbacks of Analysis 1, it must be based upon a single ranking of an agent's counterparts with respect to their prudence. Then, an agent will be ideally prudent just in case he is, according to this ranking, at least as prudent as any of his counterparts. Such a ranking is not far to seek: for any two counterparts  $A_1$  and  $A_2$  inhabiting worlds  $w_1$  and  $w_2$ , respectively, we can define  $A_1$  to be *at least as prudent as*  $A_2$  just in case  $U_{w_1}(w_1) \geq U_{w_2}(w_2)$ . Then the analysis of ideal prudence becomes:

Analysis 2: Agent  $A$  in world  $w$  lived an ideally prudent life if and only if, for all worlds  $w'$ ,  $U_w(w) \geq U_{w'}(w')$ .

Analysis 1 and Analysis 2, although they often differ in their evaluations of prudence, are easily confused with each other; for they represent two possible resolutions of an ambiguity in our informal explication of prudence. We wanted an analysis of prudence to capture the idea that an agent was ideally prudent if his pref-



erences were better satisfied in the world he actualized than his preferences would have been satisfied had he actualized some other world. But in moving from world to world in order to evaluate to what extent his preferences would have been satisfied, does 'his preferences' rigidly designate the preferences he actually had or non-rigidly designate the preferences of whichever counterpart inhabits the world under evaluation? Taken rigidly, it leads to Analysis 1: an agent  $A$  in world  $w$  was ideally prudent if the preferences he had in  $w$  were better satisfied by  $w$  than they would have been satisfied by any other world  $A$  might have actualized; taken non-rigidly, it leads to Analysis 2: an agent  $A$  in world  $w$  was ideally prudent if the preferences he had in  $w$  were better satisfied by  $w$  than the preferences of any of his counterparts were satisfied by their worlds.

Analysis 2 does not lead to the conclusion that it would always be prudent to discard one's freedom and instill in oneself a bias toward the actual. For, in the case considered above, letting  $w$  be the world in which the agent takes the pill, we have by assumption that, for all other worlds  $w'$ ,  $U_{w'}(w') > U_w(w)$ . Thus the agent is not ideally prudent, according to Analysis 2, if he chooses to take the pill; indeed, he is the least prudent of all his counterparts.

Analysis 2, however, rests upon an assumption that does not underly Analysis 1. In evaluating an agent's prudence according to Analysis 1, all utilities were calculated with respect to a single timeless utility function. But in evaluating an agent's prudence according to Analysis 2, utilities had to be calculated not only with respect to the agent's timeless utility function, but with respect to all the timeless utility functions that the agent might have had. Thus Analysis 2 assumes that the zeros and units of all these utility functions can be meaningfully correlated with one another. In this sense, Analysis 2 rests upon interpersonal—or to be more precise, intercounterpartal—comparisons of utility. But this assumption does not seem to be any more serious than the assumption of intertemporal comparisons of utility: variations within a single life can, in general, be just as great as variations among counterparts from world to world.

The introduction of intercounterpartal comparisons of utility enabled us to rank the counterparts in such a way as to take into account how each counterpart evaluates his own life; similarly, we can rank the available worlds: one world ranks higher than another world just in case the preferences of the counterpart inhabiting the

former world were better satisfied than the preferences of the counterpart inhabiting the latter world. Such rankings, in turn, enable us to adapt the notion of prudence as applied to an agent's entire life-strategy, to a notion of prudence that applies to an agent's individual acts. But it won't do simply to say that, of two acts between which an agent must choose, one act is more prudent than another just in case it actualizes a better world. By performing an individual act, the agent does not uniquely determine which world is to be actual (unless he will never again act freely), but determines only that one of a diminished *set* of worlds is to be actual—which one depending upon the agent's future acts. Moreover, it may be that any such set of worlds that the agent can determine by his choice of action contains both highly ranked and lowly ranked worlds. But although an individual act cannot in general by itself actualize the best world still available to the agent at the time of acting, we can at least require that the ideally prudent act be the act that leaves open the possibility that the agent will eventually succeed in actualizing that world, by continuing to perform ideally prudent acts in the future.

Thus consider an agent  $A$  in world  $w$  at time  $t$  who must choose to perform some act from the set  $X_{wt} = \{x_1, x_2, \dots\}$ . Let  $W_{wt}$  be the set of worlds still available to  $A$  at  $t$ ; that is, the set of available worlds that are identical with  $w$  up until time  $t$ . Let  $W_{wti}$  be the subset of  $W_{wt}$  which contains all and only those worlds in which  $A$  performs the act  $x_i$  (for all numbers  $i$  such that  $x_i$  exists). Then we have the following:

Analysis 2.1: An act  $x_i$  from a set of alternative acts  $X_{wt}$  for an agent  $A$  in world  $w$  at time  $t$  is an *ideally prudent act* if and only if there is a world  $w'$  in  $W_{wti}$  such that, for all worlds  $w''$  in  $W_{wt}$ ,  $U_w(w') \geq U_w(w'')$ .

Analysis 2.1, I believe, correctly captures the notion that an agent's individual acts cannot be evaluated with respect to their prudence in isolation from the effects they may have on the agent's future acts and future preferences. In this way, an evaluation of the prudence of an agent's available life-strategies is conceptually prior to an evaluation of the prudence of an agent's individual acts.

#### IV

The most interesting sorts of case upon which to test out Analysis 2 (and Analysis 2.1) are those in which an agent's choice of action will affect his future preferences. For example, consider the case of a would-be philosopher who has been offered but a single job: a

teaching position in the Black Hills of northeastern Wyoming. He must now choose between picking up and moving to Wyoming—let a typical such world be  $w_1$ ; and giving up professional philosophy to become a taxicab driver in New York City—let a typical such world be  $w_2$ . Moreover, suppose that according to his present preferences, and the preferences he would have were he to become a cab driver in New York City, any world in which a counterpart of his has freely chosen to live in Wyoming ranks near the bottom—country living is not for him! But further assume that, if he were to move to Wyoming, he would soon prefer the lowing of cattle to the honking of horns, would soon prefer a rodeo to a symphony; indeed, his attitude toward country living would change so extensively that, according to his resultant timeless preferences, he would be quite satisfied with his life in Wyoming—far more satisfied than the cab driver in New York is with his own life, according to his timeless preferences.

According to Analysis 1, the cab driver was prudent in deciding not to move to Wyoming, since  $U_{w_2}(w_2) > U_{w_2}(w_1)$ . But Analysis 2, correctly I think, judges the Wyoming philosopher to be more prudent than the New York City cab driver, since  $U_{w_1}(w_1) > U_{w_2}(w_2)$ . According to Analysis 2, how the man *would* have felt about Wyoming, had he never gone, is irrelevant to an evaluation of his prudence. What counts is how he will feel about Wyoming, once there; and we have assumed that this is better than how the cab driver feels about his own life. In general, Analysis 2 will recommend that an agent perform an act that will change his preferences, if the new preferences will be better satisfied than the old preferences would have been, had he not performed the act.

But in cases where the contemplated change in preferences is great, Analysis 2 may appear far too radical in its consequences. For example, suppose that an agent must decide whether or not to indulge in the latest discovery of medical science: a *complacency pill*. Whoever takes this pill becomes, from that time onward, completely satisfied with all the circumstances of his life, come what may; thus, as a result of taking the pill, he develops a utility function that assigns a high positive utility to all his available worlds. In analogy with the case last considered, let us suppose that, if the agent does not take the pill, he will always find the thought of such an artificially induced complacency abhorrent, but that, if he does take the pill, his resultant timeless preferences will be better satisfied by his world than the preferences that he would have had would have been satisfied, had he not taken the pill.

Analysis 2, it may seem, must judge that the agent is prudent if he takes the complacency pill, just as it judged that the would-be philosopher is prudent if he moves to Wyoming. For, according to Analysis 2, it is prudent for an agent to do whatever will bring his preferences and his world into closer accord, even if this involves, as in the case before us, radically altering his preferences through the use of drugs. But most of us, I think, would limit the extent to which an agent can tamper with his own psychological make-up without thereby exceeding the bounds of prudence. Can Analysis 2 be modified to accommodate this intuition?

One might be tempted to diagnose the problem as follows. According to Analysis 2, the ranking of an agent's counterparts with respect to their prudence depends only upon how each counterpart evaluates his own life. For this reason, an agent who has chosen to alter his preferences through, say, the use of drugs, hypnosis, or brain surgery, might still rank high on the scale of prudence, even though the great majority of his counterparts find his life contemptible. This suggests that a more moderate account of prudence can be obtained by allowing the ranking of an agent's counterparts to depend also upon how each counterpart evaluates the lives of his fellow counterparts. Thus, a worldless utility function can be defined—thought of as assigning utilities to worlds from the standpoint of the counterpart set as a whole—by letting:

$$(3) \quad U(w) = \sum_{w'} U_{w'}(w)$$

for all available worlds  $w'$ .<sup>10</sup> Then, an agent was ideally prudent if he actualized a world with maximal utility according to this utility function; and Analysis 2 is replaced by:

*Analysis 0: Agent A in world w lived an ideally prudent life if and only if, for all worlds w',  $U(w) \geq U(w')$ .*

If, as seems plausible, the great majority of an agent's counterparts would disapprove of radically altering one's preferences by unnatural means, then Analysis 0 will not recommend that the agent take a complacency pill.

But Analysis 0 won't do as an analysis of prudence. For not only would it allow me, for example, a say in evaluating the prudence of my more immoderate counterparts; it would also allow them a say in evaluating my prudence. But the evaluation of my prudence

<sup>10</sup> The use of summation instead of integration, with its implicit assumption that the number of available worlds is finite, covers up a host of problems (which cannot be entered into here) about introducing measures over the space of possible worlds.

should not depend upon whether or not my life is in accord with the preferences of my counterparts. My counterparts, taken all together, form a pretty shabby lot. For every mistake I might have made in my life, I have a counterpart who made it; indeed, I have counterparts who have never chosen to do the right act (on whatever standard of rightness one has in mind). Why should I care how such counterparts would evaluate my life?

The rejection of Analysis 0, if correct, shows that there is a fundamental difference for a theory of prudence between how the systems of preferences of an agent's temporal stages relate to one another, and how the systems of preferences of an agent's counterparts relate to one another. For the move that was just rejected, the establishment of a worldless standpoint from which evaluations of prudence could be made, is exactly parallel to the move that was earlier endorsed, the establishment of a timeless standpoint. Personal identity over time is held by a theory of prudence to be fundamental, not personal identity across worlds. (This, incidentally, motivates the decision to couch a theory of prudence in the language of counterparts, rather than the language of transworld identity.)

How then, if only Analysis 2 remains, are we to avoid the conclusion that it would be prudent for an agent to take a complacency pill? In my own case, I think that the reason I would have doubts about taking such a pill, even knowing that it would result in the satisfaction of preferences, is that I would have doubts as to whether the ever-complacent person that resulted from my taking the pill would be *me*, and as to whether it would be *my* preferences that were being satisfied. A sharp enough break in the continuity of one's preferences is tantamount to death; taking the pill, then, can be viewed as an act of suicide. On this assumption, Analysis 2 gives the right answer after all: it would not recommend that the agent take the complacency pill, unless all possible futures were so bleak as to make nonexistence the most preferred option.

But perhaps the situation is not, in general, quite so simple. Personal identity, I suspect, can admit of degrees; and cases may arise which cannot be counted straightforwardly either as cases of death or as cases of survival, but must be counted as something in between. Thus, a full treatment of prudence would have to find a way of incorporating degrees of personal identity into the calculations of Analysis 2—a task that will not be carried out in these pages.<sup>11</sup>

<sup>11</sup> For suggestions as to how to handle degrees of personal identity numerically, cf. David Lewis, "Survival and Identity," in A. O. Rorty, ed., *The Identities of Persons* (Berkeley: Univ. of California Press, 1976), pp. 32–36.

## V

Analyses 2 and 2.1, because they involve a number of simplifications and idealizations (as I have always tried to point out), are probably best thought of as providing a skeletal outline of a theory of prudence, and not the full theory itself. Nevertheless, that outline is ample enough to embody successfully the basic features of prudence set out in the introductory section of this paper. Prudence directs the agent: Act so as to be maximally satisfied by your world! or, more precisely: Actualize a world  $w$  for which  $U_w(w)$  is a maximum! This same edict can be written in two different ways, which serve to highlight different aspects of the theory of prudence. On the one hand, prudence directs: Make the world conform to your preferences! On the other hand, prudence directs: Make your preferences conform to the world! These two principles of prudence are not independent of one another, but represent two facets of a single phenomenon; they must be jointly coordinated by the agent so as best to achieve the prudential goal, the maximal satisfaction of preferences. Taken together, the two principles epitomize the nature of prudence: to be prudent is to effect a reconciliation between oneself and one's world.

PHILLIP BRICKER

Princeton University

## THE EVIL OF DEATH \*

So death, the most terrifying of ills, is nothing to us, since so long as we exist, death is not with us; but when death comes, then we do not exist. It does not concern either the living or the dead, since for the former it is not, and the latter are no more.

Epicurus, *Letter to Menoeceus*

**T**HE common-sense view is that a person's death is one of the greatest evils that can befall him. Most of us, to be sure, would concede that in extreme circumstances—e.g., when one is suffering from a terminal, excruciatingly painful, illness—

\* I am grateful to several participants at a meeting of the Washington State University/University of Idaho Philosophy Colloquium, and particularly to Marv Hebner and Michael Ferejohn, for helpful criticisms of an earlier draft.

0022-362X/80/7707/0401\$02.30

© 1980 The Journal of Philosophy, Inc.