The Formulation of Time Preferences in a Multidisciplinary Perspective
Their Consequences for Individual Behaviour and Collective Decision-Making

9 Time and rational choice
RUSSELL HARDIN

1. INTRODUCTION

There are two major ways in which time typically enters rational explanations of social choice. First, as in all economic analyses it enters through discounting of future values. Second, as most notably in the iterated Prisoner's Dilemma, it enters through the dynamic sequencing of moves or choices over time. I wish to consider these two issues in order to investigate the ways in which they call into question some of our standard accounts of rational choice. Through them, time affects not merely our actual choices but also what our theory of choice should be. In particular I will argue that the problem of dynamic sequencing of choices inherently brings up problems of limits to reason, especially limits to knowledge, theory, and computational ability. And certain forms of the problem of discounting call into question the nature of the very subject of rational choice: the individual chooser.

It is through the passage of time and the sequencing of choices that we gain relevant information for deciding. Many apparent anomalies in social and individual choice can be clarified by recognizing the informational role of time in our choices. Taking limits to reason into account radically alters our understanding of rational choice in dynamic contexts. Taking certain grounds for discounting into account radically affects our analysis of what it is rational to do. If one is subjectively more concerned with one's preference for present than for future consumptions, then the coherence of the individual over time may be called into question. To put this in standard philosophical language, through the passage of time personal identity can become weakened. Of course, if the identity of my present and distant future selves is weak, then my present self should not self-interestedly be as concerned with that future as with my present self.

The understanding of both the dynamic sequencing of choices and the discounting of futures can be treated as derivative from the informational role of time, although the problem of discounting of futures can also have a fundamentally different basis in the weakness of personal identity over time. As the weakness of personal identity over time increases, the long term informational role of time diminishes. Much of rational and social choice is conceived almost statically, as though time played no role in choices. In the static extreme, personal identity is not an issue. As we introduce dynamic elements into our actual choice
problems, however, we open up the problem of personal identity over time. This problem is essentially one of psychology, not of economic choice, but it implies a problem in the psychology of motivation. In the extreme of weak personal identity, all actions that do not bring immediate payoffs must be viewed as other-directed, either altruistically or otherwise.

If there were no reason to discount future payoffs, each individual would maximise or optimise total payoffs over a whole lifetime. If futures are entirely discounted, each individual should maximise this moment. If we discount some degree but not fully, we should optimise over a combination of the present and some gamble about futures. Hence, in this moment we consume more than enough merely to keep us going but we also invest some part of our resources for the uncertain future. Even if we discount not at all we should consume more than enough to keep us going if there is any tendency toward diminishing marginal enjoyment from consumption in any given moment: avoiding all pleasure until the last days of one's life, when finally there would be a small fortune for pleasure spending, would not likely maximise one's lifetime pleasure. With discounting, the extent of spending in earlier years will tend to go up.

In order to treat these issues, I will first briefly lay out the nature of the problems of the discounting of futures and of the dynamic sequencing of choices. Then I will take up the informational role of time. Finally, I will address the most difficult and troubling of the issues here in the problem of personal identity over time. Throughout, I will use the iterated Prisoner's Dilemma as the strategic interaction in which these problems show themselves. Introducing messier interactions would add confusion without adding its implications. In general, however, it should be clear that few of our actual interactions have the pristine quality of an iterated Prisoner's Dilemma with similar payoffs in each iteration. Indeed, our actual interactions are likely to be an overlay of many strategic structures at once, with some people in some of these and other people in others, some of them including many people and some including only two people, as in the simplification of the 2-person Prisoner's Dilemma. Generalisation to the more complex kinds of problems we actually face may be very messy and complicated, but it will typically have the implications that we see in the highly simplified 2-person Prisoner's Dilemma.

2. DISCOUNTING

There are two general classes of reasons that we may give for discounting futures:

1. We may discount for objective reasons. Among the most important of these is that relevant aspects of life are less sure in the future than in the present. For example, as in Tolstoy's tale, 'How Much Land Does a Man Need?' we may accumulate resources for a future which we will not be here to enjoy. The central role of discounting in economic analyses turns on the difficulty of predicting future technologies, which may make present investments less valuable - as purchasers of early, expensive pocket calculators and word processors can well appreciate.

2. We may discount for subjective reasons. That is, we may discount futures because we think they are not our futures, because we find it hard to identify fully with the persons who will bear our names and legal identities in the distant future.

Objective discounting is readily understood and it need not do more than complicate our usual analysis of rational choice over time. As in Taylor's account of the iterated Prisoner's Dilemma, discounting may undercut the prospects for cooperation because it reduces the future benefits of running current risks of cooperation (Taylor 1976, p.29 and passim) (1). In the usual analysis of the one-shot Prisoner's Dilemma, I have a dominant strategy of not cooperating, because I am better off not cooperating no matter what you do. To see this, consider the payoffs in the matrix of Game 1. The strategy of C is 'cooperate' while D is 'defect'. If Row and Column both cooperate, each gets a payoff of 1 unit of something valued. If they both defect, they get payoffs of -1. If Row cooperates while Column defects, Row gets -2, while Column gets 2. If I am playing you in a one-shot play of the game, I should defect because I do better by defecting than by cooperating no matter what you do.

This analysis fails for the iterated Prisoner's Dilemma because in it you can make your future choices of cooperating or not cooperating contingent on my present choice. What gives me incentive to cooperate now in the present play of the game is the prospect of higher net payoffs in the future to offset the higher payoff I would get now from not cooperating. Obviously, if those future payoffs are sufficiently discounted by the uncertainty of their ever coming to pass, then they cannot offset my present higher potential payoff from not cooperating. There is some point between total discounting and no discounting at which I am indifferent between cooperating now and not cooperating now.

An obvious reason for discounting the future is that it may never come, at least not for oneself. In Tolstoy's story, the peasant Pakhom is offered a gamble of all the land he can mark off before sunset. Naturally, he is too ambitious in his efforts and in the end he has run too big a circuit and that he may not be able to get back to the starting point before sunset, in case which he gets no land and he loses his modest savings. Already at the point of exhaustion, he now runs very hard and he makes it back just in time as the sky turns dark. Then, alas, in the moment of his success, Pakhom falls dead at the feet of his presumptive benefactor. How much land does a man need? Tolstoy's peasant needs only six feet at the end of his day.

Another objective reason for discounting is that one has relatively poor information about more distant futures and, in any case, poor theories for understanding causal relations that will determine future values. If I am in an iterated Prisoner's Dilemma with you, I typically will not know enough to be sure you will not suddenly change your discounting parameter just because you suddenly have an opportunity for great gain from a slight increase in present resources. Hence, your style of play may change and I may be the loser. What this is to say, I will not generally know how all your other interactions may affect our modest interaction. Indeed, I will not even know how my own other interactions may affect ours so that I may not even be able to predict with confidence how I will behave toward you in the future. We all have associates or colleagues with whom we have once been in relatively close, trusting relationships but with whom we now have distant, less trusting relationships. Few of us can be naively confident of the stability of such
relationships over time. But such relationships are partly model instances of the iterated Prisoner's Dilemma.

Finally, let us briefly consider subjective reasons for discounting based on my tenuous subjective concern for my future self. We all know the feeling that one can enjoy something this moment only at significant cost to ourselves in the future and most of us know well the tendency to succumb to attractions of the moment despite future costs that might seem far greater than the momentary pleasures. Philosophers treat this problem under the rubric of akrasia or weakness of will. Akrasia generally makes sense in only two ways. First, one may be a conflicting jumble of urges, instincts, reasons, and so forth, as for example in Freud's complex theory of the mind. On this account one's personal identity is a motley jumble from which we should expect conflicting results. Second, one may simply not care as much about the future as about the present quite apart from the objective grounds for discounting future values discussed above. On this account one's personal identity is weak over time.

3. LIMITS TO REASON IN DYNAMIC CHOICE

At first notice the problem of limits to information and theory may seem to work against cooperation toward the future because these limits lead to discounting. But note that this is not the full story. Limits to reason may often enhance the prospects of cooperating. Indeed, without limits to reason there is no ground for not conceiving a dynamic choice situation as essentially equivalent to a one-shot choice situation.

In game theory it is commonly supposed that we may collapse a sequence of plays of one or more games into a single play of a combined game. In playing out the sequence, I merely follow the overall strategy choice I make at the outset. It is ironic that the assumption of discounting of futures, which implies uncertainty about relevant knowledge for choice into the future, is a commonplace in rational choice but that the more radical implications of such uncertainty are seldom taken into account. Uncertain knowledge and, more generally, limits to reason play an enormous role in many other explanations of rational behaviour in economics and social life. Ignoring them in the usual collapse of a sequence of plays into a combined game is fundamentally wrong-headed for a game theoretic account of actual choice.

Introducing sensible limits to reason has far reaching implications for game theory and rational choice. The most important implication for the present discussion is that it potentially turns sequential plays into opportunities for information gathering and giving. Hence, it adds a new twist to the usual incentive to maximise one's payoff: one must use some plays to discover how the other player or players respond to various moves. Of course, this means that one can use present plays to influence how others move in future plays.

An immediate implication of this way of viewing actual plays of games is that much of the vocabulary of standard game theory is miscast. In particular, the standard notions of cooperative and uncooperative games are too ethereal or idealised for actual games. A cooperative game in the technical sense of game theory is one in which binding agreements can be made. In actual life, the bindingness can only be an uncertain notion. There is no person or institution that can be considered perfectly reliable in enforcing agreements brought to it. For example contracts at law are often not enforceable despite the best intentions of the signatories at the time of entering them. But if this is true of game contracting can be considered an 'actual' game - can be so absolutely binding as to make them 'cooperative' in the technical sense of game theory.

Nevertheless, it makes generally good sense to speak of actual games as being played under cooperative conditions when agreements on this point can be cooperative. Games played in the laboratory are motivated by forces in the control of the experimenter who can make payoffs contingent on players' assertions of how they wish to cooperate with another. The prisoner's Dilemma relies on forces outside a particular play: you present play. In actual practice, the enforcement of agreements in a as more or less cooperative. The extremes of cooperative and noncooperative games, as these terms are technically used, are not likely to occur outside pure theory. For the present discussion, we need only note that is iterated. That is to say, when time enters into our interaction, in which case the present player enters to make it possible for us to make commitments about how we will play.

What are the sources of uncertainty about the future? Apart from the standard problems of limited knowledge and the narrowness of focus of any explanation of future events we are likely to use in coming to understand the future, there is also a dramatic theoretical uncertainty many games are indeterminate. Consider one of the simplest problems for eight possible strategies for each player if we restrict choices to pure and contingent strategies. These are:

1. Choose C on the first and second moves;
2. Choose C on the first move; on the second move choose what the other player chose on the first move;
3. Choose C on the first move; on the second choose the opposite of what the other player chose on the first move;
4. Choose C on the first move and choose D on the second move;
5. Choose D on the first move and choose C on the second move;
6. Choose D on the first move; on the second move choose what the other player chose on the first move;
7. Choose D on the first move; on the second move choose the opposite of what the other player chose on the first move;

Suppose the matrix of payoffs for our Prisoner's Dilemma is as in Game 1. Now let us construct the game matrix for the twice-played Prisoner's Dilemma with each player choosing among these eight strategies. If we view the twice-played game as a single collapsed game in which each choice, we have the payoff matrix of Game 2.
In Game 1 there was a dominant strategy; choose D no matter what the other player does. In Game 2 there is no dominant strategy. For example, if we consider Row's strategy, for both players choosing D is a better choice than choosing A. So Row's best choice would be strategy 1, and Against Column's strategy, Row's best choice would be 2 or 4.

Against this, a combined game does dominate other possible strategies. Against Column's strategy, both players can play against Row's strategy, and if both players play against Column's strategy, then each player will play rationally.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1. Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.

Against Column's strategy, the payoff for the better choice, A, is 36, and the payoff for the worse choice, D, is 1.

Against Row's strategy, the payoff for the better choice, A, is 6, and the payoff for the worse choice, D, is 4.
strategy, that is, one which dominates all others. But on an analysis similar to that for the twice-played Prisoner's Dilemma given above, we can reduce an x-times iterated play of the game to a simple Prisoner's Dilemma by stepwise elimination of dominated strategies. We would have to consider reduction stepwise in the sense that we would first eliminate those strategies that are dominated by any other strategy to produce a reduced game. If this game were more than a two-by-two game, we could then eliminate strategies within it that are now dominated but that were not dominated in the original full game. Eventually we would have a residual two-strategy game in which the dominant strategy would be that in which one defects on every play of the iteration.

This analysis is equivalent in its prescription to the backwards induction argument in which one defects on the last play, therefore on the next to last play, and so forth back to the first and therefore every play. The standard tit-for-tat strategy in which one chooses C on the first play and thereafter reciprocates on each play whatever the other player chose on the prior play would be eliminated on the first round of reductions because it is dominated by the slightly different strategy that combines tit-for-tat on the first x-1 plays with defection on the last play. This latter strategy is then dominated by the slightly different strategy that combines tit-for-tat on the first x-2 plays with defection on the last two plays. One can see that the backwards induction argument in which one defects on the last play, therefore on the next to last play, and so forth back to the first and therefore every play is equivalent to the backwards-induction argument in which one defects in every play of a reduced game. If this game were more than a two-by-two game, we could then eliminate strategies within it that are now dominated but that were not dominated in the original full game. Eventually we would have a residual two-strategy game in which the dominant strategy would be that in which one defects on every play of the iteration.

The tit-for-tat strategy is very good against many strategic choices of the other player - indeed, it is exceedingly good. Yet it is knocked out of contention because it is slightly inferior to another strategy that is itself similarly knocked out of contention, and so forth. Let us call the strategy that combines tit-for-tat on x-n plays with defection on the last n plays the T-n strategy. In a 100-play iterated Prisoner's Dilemma, the T-20 strategy is markedly inferior to the tit-for-tat strategy against many other strategies and yet it survives the elimination of dominated strategies that initially eliminate tit-for-tat. In that elimination we eliminate good strategies to yield a very poor finally dominating strategy. This move is similar in some respects to the so-called minimax regret rule, under which we avoid possibly bad outcomes with which the bad outcomes are riskily associated. The usual sensible objection to the minimax regret rule applies here as well: if the risk is very slight, we should run it in order to get a highly probable good result. To choose to run no risks is to shrivel and die lest one should die even sooner while trying to lead a good life. If there is a realm of aggressive strategies, the minimax regret rule, it must be quite restricted. Similarly, the argument that the all-defect strategy dominates all other strategies at least stepwise is an argument that can yield a compelling recommendation only in quite restricted circumstances - for example, it seems moderately compelling in the twice-played Prisoner's Dilemma or in a sequence of plays of a Prisoner's Dilemma whose negative payoff to one who cooperates while the other defects substantially outweighs the potential for offsetting gains from a few joint cooperations. As a recommendation in a typical large-number iterated Prisoner's Dilemma it can be no more compelling than the recommendation that one never leave one's house for fear of having an accident while walking or driving elsewhere.

In sum, the backwards induction argument for continual defection is inherently flawed, just as the general recommendation that we follow the minimax regret rule is flawed. Indeed, one can plausibly see it as logically flawed in the following sense. I know that you will play in the future is actually important for how I should play now. In this respect, the iterated Prisoner's Dilemma is clearly different from the one-shot Prisoner's Dilemma already for merely two sequential plays. In the one-shot game there is no possibility of affecting the other's choices, there is no way of making strategies contingent. Playing the game twice in succession already introduces the possibility of making strategies contingent on the other's choices, as revealed in Table 1. Hence, there is a qualitative difference between the one-shot and the iterated Prisoner's Dilemma in that the only way to manipulate learning is to have more plays. Rational players must know of the possibility of making their strategies contingent on each other's when play is iterated. And because they can gain enormously from doing so, they must do so. Of course, the more iterations there are the more likely it is that this possibility can be realised and the more heavily its possible realisation should weigh in one's choice.

Real players, as distinguished from the players of pure theory, will reach the same conclusion from experimenting with alternative strategies to test their adversaries. They will use multiple iterations to gain knowledge of how others play. Because most of us are real players at best, most of us find it hard to credit the supposedly logical deduction that we must defect on every play of even a one million play Prisoner's Dilemma. Our instincts are correct for the logically impoverished world in which we actually think and choose.

4. PERSONAL IDENTITY

In any economic account, time plays a positive role as an aspect of the management of any productive resource such as labour or capital. For example, it plays an obvious role in the traditional account of limited information for decision making: more or less time collecting relevant information means better or poorer decisions. In the context of limits to reason, it also plays a positive role as a source of information, as in the foregoing account of the iterated Prisoner's Dilemma. Another conspicuous role it plays in actual life, though not very much in rational choice theory, is the destructively negative role of detaching people from their supposed future selves.

Among the reasons for discounting, as noted above, is the possibility of weakness of will or akrasia. We may view the problem of akrasia as analogous to certain problems of conflicting interests over time. You do something for me today on the strength of my promise to reciprocate tomorrow. When tomorrow comes it may not be in my interest then to reciprocate. But it was generally in my interest overall to exchange your favour today for mine tomorrow. It is from this larger perspective of my overall interests that we can meaningfully say that it is in my interest to be able to bind myself against my particular interest of the moment when tomorrow I must reciprocate. Similarly, we may say of certain actions that they are akratic in the sense that they reduce one's overall welfare and are therefore against one's overall interest even though they may seem straightforwardly to enhance one's welfare in a particular moment.
Both these problems, that of exchange over time and that of akrasia, call into question the unity of the person over time. If I have to give up something today in return for something in the distant future, I may have a much harder time seeing the exchange as in 'my' interest than if the return is to come very soon, even apart from objective uncertainties. It is not so much that I am less sure about the future than the present but that I care less about it; I have less attachment to the future person who will bear my name and legal identity than to my present self.

Much of the philosophical literature on the unity of the person or on personal identity is concerned with what might be called objective identity. This is the analogue to the problem of the ship that is slowly reconstituted, plank by plank, as rot sets in, so that in the end there is no part of it that was there originally. Is it then the 'same' ship? Objective identity for a person would be a matter of constancy of personal character, preferences, and interests. For rational and moral choice, there is a more urgent concern with subjective identity, with one's subjective identification with the interests of one's future self. This is so not least because rational and moral choice are inherently concerned with subjective valuations of consumptions. Given the enormous weight of the subjective in our choices, we should not shy from considering its impact over time.

As noted above, there are two ways to make sense of akrasia: we may suppose that one is a motley set of interests, desires, instincts, and so forth in competition for control of one's actions or we may suppose that one has a weak attachment to one's self. The opposite of the former view would be to suppose that the person is very highly integrated as though there was something like Freud's superego governing the relationship between different instincts, desires, and so forth. The opposite of the latter would be to suppose that one has a strong attachment to one's future self, so strong perhaps as to make one indifferent between present and future consumptions apart from objective considerations of uncertainty about futures. These two ways of making sense of akrasia suggest two aspects of the problem of personal identity: identity may falter synchronically or diachronically, that is, in any given moment as in the motley view of our interests, or over time, as in the view that one has ever weaker attachments to further future selves. These two aspects of our problem may be overlaid, so that we may suppose that weakness of will follows both from the motley quality of the person and from the person's weak attachment to future selves. To keep these categories or dimensions straight, let us speak of the integrated and the motley identity views, and of the weak and the strong identity views.

On the strong identity view, my interests are always the same: to maximise my consumptions overall, subject to objective grounds for discounting future consumptions. To say that I suffer from akrasia at any given moment is to say that somehow my reason nods and lets me act in a way that does not maximise consumptions overall. It is merely an error on my part. To speak of weakness of will, as we commonly do, instead of akrasia suggests the motley identity view, as for example in the Freudian conception of the compartmentalised mind with a superego concerned with consumptions or well being overall and the ego and the id competing for pleasures or other urges of the moment in disregard of the future. It is not simply that I make an error but that the 'wrong' part of my motley identity gains momentary control. To make such a claim stick, to say that there is something wrong or weak about my action, we must have a strong theory of value for the person that morally differentiates consumptions that feed different parts of the motley self. The notion of strong identity cannot easily withstand parcelling into a motley of urges or identities without such a dubious theory.

To say, on the grounds that it will reduce other opportunities later, that one ought not now do what one enjoys doing now will force us inexorably into valuing not the person but the life. If we do parcel our identity while insisting that it is still all one, then any supervenient claims will be about what actions benefit and what actions harm us will be inescapably paternalistic. One cannot occupy the position of the life in order to overlook actions within it to say which contribute more or less to the life. Suppose, as must happen if akrasia is a serious issue, that one's momentary, partial self disagrees with any supervenient claims about what one ought to do in this moment. What person or part of a person is to arbitrate the difference? And if we have an answer to this question, How is the difference to be arbitrated? A coalescence of all our partial interests and valued consumptions into one overarching utility function suggests the quality of the utility theory in Arrow's impossibility theorem (Arrow 1963). In that utility theory, it is whole states of affairs, not marginal contributions to whole states of affairs, that must be compared (2). On a strong identity theory it is variant whole lives, not variant consumptions per se, that must be compared.

A genuinely strong, integrated identity view seems finally incoherent. At the very least, on this view the notion of akrasia reduces to the trivial notion of carelessness or nodding reason. The interesting questions therefore arise for a view of identity that is at most moderately strong. Such a view automatically brings in some degree of discounting of future consumptions for reasons of weakened personal identity over time. And it turns some judgements of akrasia into judgements of distributive justice or of fairness to some future self. It may also make it less incoherent to view the person as a motley whose overall well being, considered in the moderately short run, makes sense. In the sufficiently short run, we can imagine that there is a central tendency to one's desires and values and also that there is a, perhaps still somewhat crude, reasonable grasp of relevant causal relations and expectations for fulfilling one's desires.

For rational choice theory, the most destructive of the implications of a high degree of consistency of personal identity over time is that it radically infects our value theory. We are forced to consider whole life values rather than marginal choice values. But the implication of supposing a weak identity of personal identity is even worse. It is that choices about the present versus various futures are inherently moral choices that involve trading off well being for one person against that for others. Hence, a narrowly rational, self-interest theory no longer makes sense for choices over time. Rational choice theory is therefore reduced in scope to the very near term interests of choosers. The very subjectivity of rational choice value theory radically reduces the apparent scope of the theory as a theory of purely individual behaviour.
5. WEAK IDENTITY AND STRONG SOCIETY

It is a common criticism of rational choice explanations of behaviour that society cannot be based on self interest. Seemingly, this criticism must be greatly strengthened if personal identity over time is extremely weak or if interactions are extremely anomic. Iteration produces incentives for cooperation when relationships are thick or intense enough to bring interactions close together in time, but not otherwise. The intrusion of stiff discounting clearly inhibits cooperation in iterated Prisoner's Dilemmas, as argued by Taylor (1976) and Axelrod (1984). This means that spontaneously cooperative social relations to resolve Prisoner's Dilemmas can essentially be based on rational incentives only in very small communities with relatively stable ongoing relationships. To secure more than this requires the inculation of relatively strong norms or the creation of institutions that generalise across the interests of people. Such institutions may overcome even relatively weak identity and relatively anomic interactions to produce substantial cooperation.

If it is institutions that glue things together in the face of problems of weak personal identity, then these institutions are likely to determine much of the content of the self, giving it much of its objective continuity. Hence they may enhance stability not only because they can stabilise certain incentives but also because they can stabilise our expectations about each other's and our own preferences over time. I may not subjectively identify very strongly with my future self but I can fairly reliably predict what I will want even in the distant future instantiation of myself with whom I hardly identify. Indeed, on the weak objective identity view, I am who I am now largely because I have lived my life in the context of certain institutions and not others. If we insist on very strong objective identity that is somehow built-in and largely independent of the vagaries of what institutions we live with and under, it is hard to imagine that identity could be. This is similar to the conceptual problem we have if we insist on a very strong notion of personal autonomy: if the person is sufficiently autonomous, then the person must be hollow, as hollow perhaps as the so-called wild boy of Aveyron who was discovered living in the wild in 1799 (Lane 1976).

If identity is not very strong and if the incentives for cooperation are often contingent on future rewards, even distant future rewards, from repeated interactions, then rational choice theorists must address the ways in which societies mould individual values and the ways in which institutions create short term incentives to behave in ways that lead to long term stability. If they do not, they are left with explanations that turn very heavily on malleable preferences that they cannot explain. Yet it may be the very malleability of preferences over time that can yield explanations of the institutional structure we have.

REFERENCES


Tuesday evening workshop on contemporary moral and political theory at the University of Chicago.

1. Taylor introduces discounting in part for technical reasons. Without it, payoffs in the infinite future would swamp payoffs in the realistic short term. One must either use discounting or termination of play after a finite series to avoid the lack of realism and the intractability of unbounded future payoffs.

2. This utility theory leads to perverse complications in our understanding of human choice (Hardin forthcoming).

NOTES

For their comments on an earlier draft of this paper, I wish to thank Robert Chirinko and William H. Kruskal; the organisers of and the participants in the conference, 'Time Preference', at the Wissenschaftszentrum in Berlin, December 1985; and the participants in the informal
### Game 1: Two-person prisoner's dilemma

<table>
<thead>
<tr>
<th>Column</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1,1</td>
<td>-2,2</td>
</tr>
<tr>
<td>Row</td>
<td>2,2</td>
<td>-1,1</td>
</tr>
</tbody>
</table>

### Game 2: Twice-played 2-person prisoner's dilemma

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,2</td>
<td>2,2</td>
<td>-1,3</td>
<td>-1,3</td>
<td>-1,3</td>
<td>-4,4</td>
<td>-4,4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2,2</td>
<td>2,2</td>
<td>-1,3</td>
<td>-1,3</td>
<td>0,0</td>
<td>0,0</td>
<td>-3,1</td>
<td>-3,1</td>
</tr>
<tr>
<td>3</td>
<td>3,-1</td>
<td>3,-1</td>
<td>0,0</td>
<td>0,0</td>
<td>-1,3</td>
<td>-1,3</td>
<td>-4,4</td>
<td>-4,4</td>
</tr>
<tr>
<td>4</td>
<td>3,-1</td>
<td>3,-1</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>-3,1</td>
<td>-3,1</td>
</tr>
<tr>
<td>5</td>
<td>3,-1</td>
<td>0,0</td>
<td>3,-1</td>
<td>0,0</td>
<td>0,0</td>
<td>-3,1</td>
<td>0,0</td>
<td>-3,1</td>
</tr>
<tr>
<td>6</td>
<td>3,-1</td>
<td>0,0</td>
<td>3,-1</td>
<td>0,0</td>
<td>1,3</td>
<td>-2,2</td>
<td>1,3</td>
<td>-2,2</td>
</tr>
<tr>
<td>7</td>
<td>4,4</td>
<td>1,3</td>
<td>4,4</td>
<td>1,3</td>
<td>0,0</td>
<td>-3,1</td>
<td>0,0</td>
<td>-3,1</td>
</tr>
<tr>
<td>8</td>
<td>4,4</td>
<td>1,3</td>
<td>4,4</td>
<td>1,3</td>
<td>1,3</td>
<td>-2,2</td>
<td>1,3</td>
<td>-2,2</td>
</tr>
</tbody>
</table>