Weakness of Will and Time Preference

Naoyuki Shiono*

Abstract

This paper presents a theory of akrasia that draws essentially on the notion of time preference. When we make a decision over which course of action to pursue, it often happens that what can be obtained in the near future motivates us more strongly than what lies temporally further afield, although we are aware that the latter is to be evaluated more highly than the former. This is the effect of time preference, and I will argue that it plays a crucial role if we are to understand how our behaviour can exhibit weakness of will.

I

Since Donald Davidson published his influential paper "How is Weakness of the Will Possible?"\(^1\), the problem of the weakness of will, namely akrasia, has been discussed extensively in the philosophical literature. In this paper, I will briefly introduce and summarize the issue and the current discussions surrounding it, and then present one way in which it may be resolved.

The problem of akrasia can be recognized as a paradox which is inherent in the folk psychological description and explanation of human mind and behaviour. Folk psychology characterizes human behaviour as something intelligible by attributing propositional attitudes, such as beliefs and desires, to agents. And the attribution is made on the assumption that the agent is rational, so that the attitudes attributed are coherent with each other and with actions. This principle of rationality is a fundamental guiding principle of the enterprise of folk psychological explanation, because, without it, the attribution of propositional attitudes to agents will not be possible.

Akrasia presents a threat to the practice of folk psychology. In akratic action, there is a breakdown somewhere in the rational process of having reasons for action, making judgments as to which action is the best, intending to perform the action, and the actual performance of the action. If akrasia amounts to a breakdown of rationality, and if rationality is the guiding principle of folk psychological explanation, then akrasia presents us with a breakdown of explanation. This is the source of the problem, which poses the question, "how is weakness of the will possible?"

Is it possible somehow to patch up the breakdown and make sense of akratic action?

* Fukui Prefectural University

\(^1\) Davidson (1970).
as akratic action? Or do we need to conclude that human agents are bound to be perfectly rational and there can be no such thing as akratic action?

Many of the proposed solutions to the problem of akrasia have this feature in common, that they introduce some further elements into the folk psychological framework than belief and desire. Davidson’s original paper made the distinction between all-things-considered judgment, which is a prima facie, conditional judgment, and unconditional judgment, to which intentional action is directly geared\(^2\). He later went on to argue that apparent cases of irrationality can be made intelligible by representing the agent’s mind as partitioned into more than one semi-autonomous structures\(^3\). Meanwhile, Richard Holton argued for the existence of an irreducible faculty of will-power, which is distinct from belief, desire, and intention, and which can be exercised rather like muscles\(^4\). Alfred Mele paid attention to the distinction between evaluative and motivational aspects of desire, the former of which leads to judgment and the latter of which leads to action, and then argued that it is the role of self-control that makes these two aspects coherent with each other\(^5\). Ronald de Sousa and Christine Tappolet focused on the function of emotion in guiding intentional action\(^6\), while Amélie Rorty emphasized the role of attention and custom\(^7\). All these proposals are based on the recognition that the simplest folk psychological framework consisting only of belief and desire will not be capable of accounting for the existence of akratic action.

However, it would not be appropriate to think of these proposals as recommending a replacement of folk psychology with something else. Rather, they are attempts at enriching the conceptual apparatus of folk psychology in a way that makes it possible for it to describe akratic action as akratic action. For example, the will-power of Holton, which is one of those candidate apparatus, is something we are supposed to be familiar with from our first-personal perspective, and at the same time something which we regularly employ in explaining the behaviour of others from the third-personal perspective. It is true that the introduction of these conceptual elements will make the explanatory framework more complicated than it has been, and it will become more difficult for us to capture it in simple formal models of practical logic or decision theory. Nevertheless, these attempts are intended to be an improvement on folk psychology, and not a suggestion for demolishing it. Whether they are meant to be an improvement on folk psychology per se, or an improvement on the characterizations of an already existing folk psychology, is not a question that I will be

---

\(^2\) Ibid., pp.38ff.

\(^3\) Davidson (1982), Davidson (1986).


\(^5\) Mele (1987).


\(^7\) Rorty (1980a), Rorty (1980b).
concerned with.

II

What I am going to propose in this paper shares the same general direction, and it particularly relies on Mele's approach to akratic action\(^8\). The conceptual apparatus that I am going to bring in is time preference. The idea of incorporating time preference in an explanation of akratic action has been explored, apart from Mele, particularly by George Ainslie and Jon Elster\(^9\). Here, first, I will explain the notion of time preference, and review two competing mathematical models that are being offered as correct representations of the phenomenon. Then, I will introduce Mele's distinction between the evaluative and the motivational aspects of desire, and explain how it can be related to time preference. From there, an overall picture of what an akratic action is will follow. Finally, I will close this paper by indicating a particular difficulty that this picture may face.

Time preference is one of the basic concepts in economics, and appears in any textbook of micro-economics. It represents the idea that if an agent evaluates some future reward now, the value of the reward will be discounted to a certain degree. For example, if one is offered a choice between receiving the reward of 1 million yen right now or the same reward one year later, one will choose to receive the reward now rather than later. In order to balance present and future reward, the future reward must contain some extra on top of the present reward. If one finds the options of 1 million yen right now and 1.1 million yen one year later just equally attractive, one is said to have the "discount rate" of 1.1 m/1 m - 1 = 10\% per year. And the "present value" of 1.1 million yen one year later is said to be 1 million yen. Rewards do not have to be in terms of money. If one finds the options of a dinner at a local Italian restaurant this evening and a dinner at a luxurious French restaurant next month just equally attractive, again one exhibits a positive rate of time preference.

It seems to be a striking fact that many of the favourite examples given by philosophers of akratic action fit so nicely with the idea of time preference. For instance, a man may choose to continue his smoking habit, although he is aware that the present smoking habit is not compatible with his future health. This man may also be aware that the future utility of future health is greater than the present utility of present smoking. Yet, the time at which he has to make the decision is right now, not the future. Then, the utility of future health will be discounted due to time preference, while the utility of present smoking will not be. If we suppose the discount rate to be 3\% per year, the present value of health 25 years from now will be the

\(^8\) Mele (1987).
\(^9\) Ainslie (2001). Elster has explored this idea in a number of writings, such as Elster (1999), Elster (2000).
result of discounting by the 25th power of \((1/1.03)\), which will amount to less than 50% of its original value. In that case, if the utility of present smoking is sufficiently high, the desire for health will not be able to compete with the desire for cigarettes.

Of course, the above description of the case does not make it automatically into an instance of akritic action. The concept of akritic action which I would presuppose rather than argue for in this paper is that of an action against a judgment, all-things-considered or unconditional, that a contrary action is better, or against an intention to perform a contrary action, where a contrary action is an action that is incompatible with the actual action. To anticipate my conclusion, akritic action occurs when one’s judgment made from a “comprehensive” perspective on one’s own life is not in agreement with one’s motivation that reflects a “cognitive-condition-bound” perspective.

We need to know a little more about time preference. As we saw in the smoking example, according to standard theories, if the discount rate is \(n\) per unit of time, the present value of utility \(U\) after \(t\) units of time will be \(U\) times the \(t\)th power of \((1/(1 + n))\). This is Paul Samuelson’s DU model, or exponential discounting. Thus, if present 1 million yen can be balanced by 1.1 million yen one year later, then they both will be balanced by 1.21 million yen two years later.

Ever since Samuelson formalized his model, the question of why the phenomena of time preference in general, and of exponential discounting in particular, exist, and the question of whether it is rational for agents to discount future utility in this manner, have not been seriously addressed in the literature. Exponential discounting has been widely accepted, not because it was well understood why it is true of human behaviour, but because it was useful and easily applicable as a mathematical model. However, for our purposes, the issue of whether it is rational for agents to discount future utility is very important. And although the explanation of why it is important has to wait until later, we will explore the issue a little further at this stage.

The question of why there is this phenomenon of time preference in the first place has finally come into the spotlight very lately. Unexpectedly, studies into this area owe much to the works of pre-Samuelson 19th and early 20th century economists, such as John Rae, Nassau William Senior, William Stanley Jevons, and Eugen von Böhm-Bawerk. These economists pointed out such factors as the following, among many others, as constitutive determinants of time preference. First, people’s lives are uncertain with respect to the future. Second, as people tend to get wealthier over time, due to the law of diminishing marginal utility, the same goods will have less utility in the future than in the present. Third, people are not good at imagining future happiness. Fourth, people are less concerned about the future than about the

---

\(^{10}\) The historical details about the treatment of time preference are reviewed extensively in Loewenstein (1992) and Frederick, Loewenstein and O’Donoghue (2002). I depend heavily on these papers in the following paragraphs.

---
present. Fifth, present goods can be used to produce more goods in the future. It is far beyond the scope of this paper to enumerate and examine all the alleged constitutive elements of time preference that these economists suggested, but a brief look at some of them is in order here.

Some of the items listed above can be excluded as irrelevant on certain conceptions of time preference. For example, the point that, as people tend to get wealthier over time, the same goods will have less utility in the future than in the present, thus making it more rational to consume the goods now rather than later, is irrelevant under the conception of time preference according to which we discount future utility rather than future goods.

In connection to the question of whether time preference is a rational phenomenon, the point about uncertainty of the future is the most relevant among the items listed above. In our example of the smoker, on the assumption that he discounts future utility by 3% per year, the present value of his health 25 years from now turned out to be less than 50% of its original value. But suppose he thinks that, even if he kicks the smoking habit right now, the probability of his being alive 25 years later is less than 50% anyway. We can imagine that he is bound to live in a region where warfare constantly takes the lives of the local residents. In this case, it will be perfectly rational for him to discount the future utility in the way he does, because, even if he chooses to stop smoking now, the expected utility due to future health is less than 50% of the utility of the health.

In fact, the notion of time preference strictly characterized is a notion according to which future utility will be discounted by time preference after whatever reductions due to uncertainty of the future are already taken into account. Thus, uncertainty of the future and time preference are independent of each other, and the former is not a constitutive determinant of the latter. That said, however, in psychological experiments as well as in economic field studies, it is true that separating out these officially independent factors is no easy thing to do.\(^\text{11}\)

Would there then be any constitutive element of time preference that makes it rational for agents to discount future utility? That is, can time preference be explained away by rational considerations that the agent should take into account in evaluating future utility? It is far beyond my capacity to examine all the candidates for rational grounds of time preference, but Shane Frederick, George Loewenstein and Ted O'Donoghue make a startlingly interesting remark in this connection.

The most compelling argument supporting the logic of positive time preference was made by Derek Parfit . . ., who contends that there is no enduring self or “I” over time to which all future utility can be ascribed, and that a diminution in

\(^{11}\) The points in this paragraph are explicitly made in Frederick, Loewenstein and O'Donoghue (2002), p.382.
psychological connections gives our descendent future selves the status of other people—making that utility less than fully “ours” and giving us a reason to count it less . . . .\textsuperscript{12}

According to Parfit, the identity of the present self with some future self is a matter of degree. It depends on how much of our present hopes, plans, tastes and personalities survive into the future. The reason, then, that an agent discounts future utility is that the agent that enjoys the utility is not strictly identical to the present agent. It would be perfectly rational to value future utility only to the degree to which the present self is identical to the future self.

I am not going to discuss here the Parfitian thesis of personal identity. I rather assume that the thesis is extremely unpopular and nobody will want to take it seriously. Thus, if the most plausible argument for the rationality of time preference is the Parfitian thesis, then many people will agree with me in concluding that it is so much the worse for the rationality of time preference. In addition, fortunately, there is one consideration that is likely to demolish all the allegedly rational grounds of time preference, including uncertainty of the future and the Parfitian thesis, in one fell swoop. This leads us to the introduction of hyperbolic discounting, which is an alternative mathematical model of time preference in competition with Samuelson’s exponential discounting.

In exponential discounting, future utility will basically be discounted at a fixed rate per unit of time. Therefore, if an agent finds 1 million yen right now and 1.1 million yen one year later equally attractive, then this agent exhibits the discount rate of 10\% per year, and consequently should find 1 million yen ten years from now and 1.1 million yen eleven years from now also equally attractive. But this is not the way in which most people behave. Most of us will exhibit lesser discount rate if the choice is between ten years and eleven years from now. Hyperbolic discounting captures this phenomenon nicely. According to this model, the present value of utility \(U\) after \(t\) units of time will be \(U\) divided by \((1 + n \times t)\). Then, supposing \(n = 0.1\) and the unit of time to be one year, while present 1 million yen will be balanced by 1.1 million yen one year later, 1 million yen ten years later will be balanced by 1.05 million yen eleven years later.

Which of these two models, exponential or hyperbolic, more properly reflects actual human behaviour is still an open question. Thus, some writers, such as Elster, seem to go a bit too far in assuming the hyperbolic model to be the correct one\textsuperscript{13}. Nevertheless, it is true that hyperbolic discounting gives us a plausible explanation of various phenomena such as the above that cannot be appropriately accounted for by exponential discounting.

\textsuperscript{12} Frederick, Loewenstein and O’Donoghue (2002), p.359.
\textsuperscript{13} Elster (2000), p.25.
If hyperbolic discounting is true of human behaviour, however, the idea that time preference has a rational ground seems much less persuasive. I said earlier that uncertainty of the future is not strictly a constitutive element of time preference, but even if it were, it would not be compatible with hyperbolic discounting, for, there is in general no reason to suppose that the probability of the occurrence between now and one year later of something unexpected is greater than the probability of the corresponding occurrence between ten years and eleven years from now. In the same way, the explanation of time preference in terms of the Parfitian thesis of personal identity is not compatible with hyperbolic discounting either, for, there is in general no reason to suppose that the degree of continuity between my present self and myself one year later will be less than the degree of continuity between myself ten years from now and myself eleven years from now.

Hyperbolic discounting also has the advantage of accounting for the following phenomenon, which seems quite pervasive in human behaviour. Sticking to the previous example of discounting future utility by \( n = 0.1 \), if we are to choose between 1 million yen 10 years later and 1.08 million yen eleven years later, the later reward will seem better. However, as time goes by, and as it approaches the point at which the reward will be received, the attractiveness of both will get closer to each other, and after seven and a half years, the preference will be reversed to the earlier reward. This phenomenon of preference reversal is what is called “time inconsistency”. Exponential discounting cannot account for this phenomenon.

It might perhaps be thought that akrasia is to be identified with time inconsistency. The identification may seem plausible on the ground that time inconsistency has an agent who at first prefers a larger, later reward but then myopically cancels this preference to opt for the smaller, earlier reward. However, ak rasia should be distinguished from changes of mind. So far, the agent who exhibits time inconsistency and eventually chooses the earlier 1 million yen has only changed his mind at some point in the process. On the other hand, a truly akratic agent must, at the very moment of making the final choice, be holding a judgment that the other alternative is better. Thus, time preference and hyperbolic discounting are not yet sufficient as an apparatus with which to describe akratic action as akratic action.

III

Here we will temporarily change our subject, and introduce Mele’s distinction between the evaluative and the motivational aspects of desire.

Gary Watson ... has suggested that ‘wants more’ has both an evaluative and a motivational use. In the evaluative use, to say that ‘\( S \) wants \( x \) more than he wants \( y \)’ is to say that \( S \) ‘prefers \( x \) to \( y \) or ranks \( x \) higher than \( y \) on some
scale of values or "desirability matrix," whereas in the motivational use, it is to say simply that 'S is more strongly motivated by considerations of x than by considerations of y'.

The point here is not as if there are some desires that come under our evaluation and other desires that motivate us. Still less is the distinction between moral reasons and selfish reasons. Rather, any desire, broadly conceived to include the moral, the selfish and anything else, can receive an evaluation by and at the same time be a motivation for an agent.

According to Mele, what our practical reasoning is concerned with is the evaluative and the motivational aspect of desire. One evaluates each of one's own desire to determine how desirable it is to satisfy it. Then one weighs them against each other and arrives at an all-things-considered judgment as to which course of action is the best. This in turn leads to the formation of an unconditional judgment or intention to perform the action. And, in non-akratic cases, the intended action will follow.

On the other hand, Mele holds that it is the motivational aspect of desire that determines which action the agent will eventually perform. Thus, if the desire to do an action x is stronger than the desire to do an incompatible action y in terms of motivational force, the agent will do x rather than y. Of course, it does not mean that the motivational force of the desire directly causes the action. Doing x can be an intentional action, in which case there can be an intention to do x in the process that culminates in the performance of the action.

According to Mele, the evaluative aspect and the motivational aspect can become out of line with one another. This is precisely the condition that can lead to akraic action. The agent who evaluates doing y higher than doing x will, all things considered, judge that it would be better to do y rather than x. This may lead to the formation of an unconditional judgment or even an intention to do y rather than x. Nevertheless, it may be that the motivational force of doing x is stronger than that of doing y. In that case, the agent will form the intention of doing x rather than y, and akraicly perform the action of doing x.

It might be instructive to have a bit more detail about Mele's description of the process that leads to the performance of akraic action. He distinguishes unconditional judgments from intentions, and argues that the agent can retain the unconditional judgment that it would be better to do y rather than x, even at the time of doing x. But with intentions, the story is a little different. On the basis of the unconditional judgment, the agent might form the intention of doing y rather than x.

---

15 Ibid.
16 Ibid., pp.43–44.
x, and it might even be an intention to do y rather than x here and now. However, any action takes some time to perform, and somewhere in the course of performing the intended action, the original intention is abandoned and replaced by a contrary intention of doing x rather than y. The original intention cannot survive the formation of the contrary intention, because, unlike desires, one cannot have incompatible intentions at the same time. Thus, according to Mele, akratic actions satisfying the following conditions are both possible. First, an agent can hold an unconditional judgment that it would be better to perform a certain action, and yet, exactly at the same time, intentionally perform a contrary action. Second, an agent can form a here-and-now intention to perform a certain action, and, a moment later, yet perform a contrary action.

We might contrast this idea to Davidson’s view of akratic action in his “How is Weakness of the Will Possible?”. Davidson argued that it is possible for an agent to have an all-things-considered judgment that it would be better to do y rather than x, and at the same time have an unconditional judgment that it would be better to do x rather than y. However, he did not in principle allow the possibility of an action that is contrary to the unconditional judgment. Later participants in the debate were not satisfied with this, and tried to argue for the possibility of action against unconditional judgments and intentions. Mele is one of them. And if we opt for Mele’s position, the distinction between all-things-considered judgment and unconditional judgment will no longer play any particularly important role. Therefore we will hereafter talk about them inclusively as “better judgment”.

Akratic action arises in Mele’s framework in cases where the better judgment or intention that results from the evaluative ranking of desires on the one hand and the actual action that reflects the relative motivational force of desires on the other come apart. Our task will then be to explain how this can happen. Even more fundamentally, we need to explain whether the evaluative and the motivational aspects have anything to do with each other in the first place, and if they do, how they are related. So far, we only know that they are in agreement in cases of non-akratic action.

To this question, Mele points to two major factors that can engender the disagreement, namely, proximity and attention. Among them, proximity is the factor that we have already talked about, which is time preference. The motivational force of a desire gets enhanced when the desire can be satisfied in the near future. It will also be enhanced when the agent is paying attention to the object of the desire. Temporal proximity and attention can thus result in the motivational force of the desire becoming disproportionate to its evaluation. This is the way in which the two aspects can become out of line with each other, which can result in akratic action.

It is thus Mele’s idea to introduce time preference in the explanation of the

\[17\] Ibid., pp.84ff.
disagreement between the evaluative and the motivational. However, he does not seem to develop this idea fully into an overall account of what an akratic action is. Moreover, since he takes attention rather than proximity as the primary factor in the enhancement of motivation, he does not examine in detail such questions as whether time preference is rational at all or whether the contrast between exponential and hyperbolic discounting has any further consequences. Therefore I would like to discuss this idea more extensively in what follows, and then in the last place point out a serious difficulty that it might face.

IV

First, let me reformulate Mele's idea in my own terminology. The evaluation of a desire is not dependent on whether its satisfaction is immediate or distant. Therefore, we can think of the evaluation to be its utility without any discount due to time preference. It can be grasped when one contemplates one's entire life, and thinks about what is good for oneself, without being biased by what can be gained soon and what lies far ahead. I will call this standpoint the "comprehensive perspective". On the other hand, the motivation of a desire is dependent on the temporal point of view from which one envisages various courses of action. Naturally, something temporally closer will be more powerful, while something more distant will be weaker. Hence, the motivational force is under the effect of discounting, and can be regarded as identical to what has so far been called the "present value" of the utility. I will call this standpoint the "present-bound perspective".

The relation between evaluation and motivation will now be clear enough. We get the latter by discounting the former by time preference. The better judgment will reflect the evaluation of the various desires that one has, while the actual action will be determined by the relative strengths of the motivational forces. If they come apart and one thereby acts against one's better judgment, that will constitute an akratic action. Thus, in my words, akratic action occurs when one's "comprehensive perspective" and "present-bound perspective" disagree with each other.

It seems to me that this way of understanding akratic action goes very well with our everyday intuitive grasp of what weakness of will is. We often think that we should really do y rather than x, but due to weakness of will, do x rather than y. It seems fair to say that what this "should really do" amounts to is the idea of "should do if we are to act on our deliberations which reflect our comprehensive perspective". The case of the smoker who sticks to his smoking habit despite his higher evaluation of his future health is a typical example.

Now we may remind ourselves that, when I introduced the notion of time preference, I went into a rather lengthy discussion of whether discounting future utility is a rational thing to do. In particular, I made a lot of effort for denying that time
preference might be explained away by uncertainty of the future. I also tried to deny that time preference might be deemed rational by the Parfitian thesis of personal identity. I can now explain why these issues are important.

Although I have claimed that evaluation reflects the comprehensive perspective from which we consider our entire life without any bias due to our temporal location, it is not really possible for human agents to assume a completely comprehensive perspective. That is because the future is uncertain for us, and we cannot assess our future courses of life as if we have a clear view of the consequences of each option that is open to us. What is achievable for us is an "as-far-as-possible comprehensive perspective", and in taking this perspective, we make probabilistic expectations of our future, and evaluate future utility within these expectations. We had the example of the smoker who is bound to live in a war-torn region, and who thinks that, even if he chooses to kick the smoking habit right now, the probability of his being alive 25 years later is less than 50%. If he were to make a better judgment now as to whether he should continue smoking now, and take an as-far-as-possible comprehensive perspective, it would not be right for him to attach the full weight of the utility of future health to the non-smoking option. He should rather reduce its utility in his evaluation, in proportion to its uncertainty. In other words, what a rational agent should consider from the as-far-as-possible comprehensive perspective is the expected utility of future courses of life.

Now, if the phenomenon of time preference is to be explained away by uncertainty of the future, there can never be any gap between the as-far-as-possible comprehensive perspective and the present-bound perspective. From the present-bound perspective, we discount future utility by time preference, and on this assumption, the result will be exactly the same as what comes out from the as-far-as-possible comprehensive perspective. Akratic action of the sort I am envisaging will then not be possible.

Accounting for time preference with the Parfitian thesis of personal identity will generate the same sort of problem. Parfitian thesis makes the identity of my present self with some future self into a matter of degree. Therefore, even when one is taking a comprehensive perspective, it will be rational to adjust the evaluation of future utility by the degree of identity between oneself and the future self that will enjoy the utility. Again, if the existence of time preference is the result of the Parfitian thesis, taking the present-bound perspective will give one the same result. And that will certainly make it impossible for akratic action to arise.

Fortunately, time preference is standardly treated as independent of uncertainty of the future, and the Parfitian thesis of personal identity remains hugely unpopular. Moreover, if hyperbolic discounting is right, neither of these purported explanations of time preference will work. Would there be any other explanation of time preference, which makes it rational for agents to discount future utility, and which leaves no possibility of a gap between the comprehensive and the present-bound perspectives?
I cannot think of any, particularly if it is demanded that the explanation should be coherent with hyperbolic discounting. Therefore, it will be reasonable to conclude at this point that there will always be the possibility of disagreement between the two perspectives.

Earlier, I introduced the notion of "time inconsistency", a consequence characteristic of hyperbolic discounting. I there emphasized that akrasia should not be identified with this notion. Time inconsistency is a phenomenon where a preference for a later, larger reward over an earlier, smaller reward will be reversed as the earlier reward approaches in time. After the preference reversal, the comprehensive perspective, which will stick to the judgment that taking the later reward is better, gets into disagreement with the present-bound perspective, which guides our action into taking the earlier reward, thereby engendering a case of akratic action. However, what makes this into a case of akratic action is not the preference reversal per se, but the disagreement between the two perspectives. Preference reversal alone would at most constitute a change of mind.

This may become clearer by pointing out that not every case of akratic action need involve time inconsistency. In my view, the frequently asserted statement that hyperbolic discounting is irrational while exponential discounting is rational is false. As far as I can see, any discounting due to time preference, apart from those resulting from such "rationalizing" factors as uncertainty of the future, is liable to generate akratic action, and to that extent irrational. Our hypothetical smoker, who exponentially discounted future utility by 3% per year, acts akratically as long as he smokes against his judgment that it would be better not to do so, without showing any sign of preference reversal.

I mentioned earlier that Mele suggested proximity and attention as the main factors that may result in the evaluative and the motivational aspects of desire becoming out of line with one another\(^{18}\). We have focused on proximity so far, but having understood how things work out, it is now relatively easy to incorporate the other factor into our scheme, if we prefer to do so. Attention is a phenomenon where the cognitive power of an agent is directed exclusively towards some particular object. Given that people's attention tends to be directed towards something that is temporally proximate, we might say that temporal proximity is actually one of the determinants that underlie the agent's attentional condition. And there may also be other determinants, such as spatial proximity, habit, and other psychological conditions.

In any case, paying attention to some particular object to the exclusion of others will mean getting remoter from a comprehensive perspective. Attention is a matter of getting immersed in whatever cognitive condition that is prevalent for the agent at that moment. Therefore, we might generalize the contrast between the comprehen-

\(^{18}\) Ibid.
sive and the present-bound perspectives into a contrast between the comprehensive and the cognitive-condition-bound perspectives, where the latter incorporates time preference and any other factors that can make the motivational force of a desire disproportionate to its evaluation.

The assertion that agents' evaluation of desires and the better judgment that depends on it are formed from the "comprehensive" perspective may be challenged by asking whether and why they would be so. In responding to this challenge, it may be better separately to consider, first, whether and why better judgments "should" reflect the comprehensive perspective, and second, whether they actually "do" reflect such a perspective.

First, an argument for the point that the agent should form better judgments from the comprehensive perspective can be based on the non-Parfitian thesis of personal identity. John Searle makes a similar claim, appealing to the following example of a female smoker in Denmark\(^\text{19}\). Searle says that it is irrational of this woman that she smokes now, at the age of twenty, knowing that that will kill her by the age of sixty. At present, at twenty, this smoker is completely indifferent to her health at sixty. But on the other hand she is now very much aware that as she approaches sixty, her health at sixty will become increasingly important for her. Nevertheless, she now judges that it would be better for her to smoke now, and smokes accordingly. In my terminology, her time preference has a relatively high rate of discounting, which results in her giving little weight to her future health. However, as long as she is now aware that her future health will become important for her in the future, she is in possession of the capacity of taking the comprehensive perspective. Yet, her better judgment is on the side of smoking rather than non-smoking. Therefore, she in effect has formed a better judgment that reflects the cognitive-condition-bound perspective rather than the comprehensive perspective. Of this woman, Searle categorically says that she is irrational, and points out, "The irrationality derives from the fact that the same self that makes the decision now is the self that will die by the age of sixty"\(^\text{20}\). It seems to me that Searle is quite right in saying this. That one should form better judgments from the comprehensive perspective follows directly from the non-Parfitian thesis of personal identity. Given that myself at twenty and myself at sixty are exactly the same self, there is no reason to give a favourable treatment to the former at the sacrifice of the latter. Thus, better judgments that reflect a comprehensive perspective are rational, and those that do not are irrational.

As to the other question of whether agents' better judgments in fact do reflect a comprehensive perspective, it would be difficult to insist that they necessarily do, as long as Searle's smoker in Denmark is a sufficiently realistic example. Rather,

---

\(^{19}\) Searle (2001), pp.128–131. I am not suggesting that the point Searle is making there is exactly the same as mine. In particular, time preference as such is not on his agenda.

\(^{20}\) Ibid., p.131.
we might have to be content with saying that actual better judgments are generally formed from perspectives which lie somewhere between the comprehensive and the cognitive-condition-bound.

Then, strictly speaking, there are four grades of perspectives. First, there is the completely comprehensive perspective, from which our entire life will be viewed at one single glance. This, obviously, is something beyond the capacity of human agents like ourselves, because the future will inevitably remain uncertain for us. Second, we have the as-far-as-possible comprehensive perspective, which grants us only probabilistic expectations of our future courses of lives. This, I am claiming, is the perspective from which we “should” form better judgments, and such judgments are the rational ones. Third, there is the perspective that is below the as-far-as-possible comprehensive, but still can claim to be “to-some-extent comprehensive”, in the sense that it is above the cognitive-condition-bound. This is the perspective which most of our better judgments in fact “do” reflect. Finally, we have the present-bound or the cognitive-condition-bound perspective, which will determine the motivational forces of our desires, and which will eventually lead to actual actions. Akratic action, by its definition as action against better judgment, arises as a result of disagreement between the third and the forth. Therefore, my thesis fully spelled out is that *akrasia* is a phenomenon that is generated by an agent’s “to-some-extent comprehensive” perspective getting out of line with the agent’s “cognitive-condition-bound” perspective.

The disagreement between the second and the third perspectives can also lead to irrationality, though of a different kind. Searle’s smoker in Denmark is a case in question. Or perhaps, for her, the third and the fourth perspectives were exactly identical with each other. At any rate, her case is not an example of akratic action, because her action is in accordance with her better judgment. Her irrationality lies rather in her judging as she did. How the second and the third perspectives can come apart and how the resulting irrationality should be accounted for are interesting questions indeed. It appears to me that the irrationality is closer to self-deception than *akrasia*\(^\text{21}\), but defending this thesis is far beyond the scope of this paper.

V

It is not necessarily the case that agents will simply pursue the course of action that is prompted by a desire with stronger motivational force. There are various measures that one can employ in order to guide one’s behaviour in a more satisfactory man-

---

\(^{21}\) The gap between the second and the third may correspond to what David Pears in Pears (1984) called “self-deceptive *akrasia*”, while the gap between the third and the fourth may correspond to his “last-ditch *akrasia*”. This point has been hinted to me by Koki Asano’s draft of a paper prepared for a joint presentation on irrationality at the Philosophy of Science Society Japan Annual Conference in 2007.
ner, and they are called “precommitment” by Elster, “interpersonal bargaining” by Ainslie, and “self-control” by Mele. In what follows I shall sketch them through an example, and in the course of doing so, a serious problem will emerge that originates from our introduction of the distinction between evaluation and motivation.

As we have already seen, hyperbolic discounting and the resulting time inconsistency are not essential ingredients of akratic action. However, just for the sake of vividness, I will present an example incorporating these elements. Mr. A has a desire to pursue his studies of Kant after returning home this evening. But at the same time, he has a desire to have a drink at a bar near N Station, where he usually gets off his train on his way home. From a comprehensive perspective, the utility of Kant studies is greater than that of the drink. Therefore, Mr. A, from this perspective, forms the judgment that it would be better to engage in Kant studies at home rather than to go to the bar. In addition, at the time when he leaves his office, since both Kant studies and the drink lie relatively far ahead in the future, the motivational force of Kant studies is also stronger than that of the drink. However, as his train approaches N Station, the motivational force of the drink gets closer to that of Kant studies, and just at the moment when the train arrives at N Station, the relative strengths of the motivational forces are reversed. This, as should by now be evident, is the consequence of hyperbolic discounting and the fact that the bar is located a bit closer than home. Mr. A goes into the bar, has some drinks, and spoils his desire for pursuing Kant studies. On the other hand, he has never revised his judgment that it would be better to go home and study Kant, even at the moment of stepping into the bar. Therefore, his action is a clear example of akratic action against a better judgment.

Now there are quite a few ways in which Mr. A can arrange things so that his determination to study Kant at home will not be defeated. I will present four of them, following Elster. First, he may leave his wallet in his office. This in effect eliminates the option of going to the bar. Second, he may take a pill which has the effect of making him very sick if he has any drink during a certain period of time. Disulfiram is exactly this kind of drug, which has been developed in order to combat alcoholism. Taking this pill will generate a very heavy penalty for going to the bar, and the option will cease to be attractive as a whole. That is, as Mr. A walks past the bar, the discounted utility of Kant studies will be greater than the non-discounted utility of the drink minus the drug effect. Third, he may buy himself a piece of cake on his way home, as a reward for himself that he can enjoy if he successfully gets back home and engages in Kant studies. The effect is that, as he goes by the bar, the discounted utility of Kant studies plus the cake will remain greater than the non-

\[\text{References}\]


discounted utility of the drink. Finally, he may take a different route this evening, getting off at M Station, from which he needs to take a long walk if he is to go to the bar instead of going directly home. In this case, the utility of the drink will always be more heavily discounted than the utility of Kant studies, thus enabling the latter never to be defeated by the former.

Mr. A, wise and sensible as he is, employs some or the other of these measures of self-control, and successfully engages in his favourite Kant studies. The important point to emphasize is that what eventually decides his course of action is the relative strengths of the motivational forces of his competing desires. Self-control techniques, by manipulating the options, arrange things in such a way that the motivational force of the desire which agrees with the agent’s better judgment will not be exceeded by the motivational force of any contrary desire. In this sense, self-control is an “indirect” way of controlling one’s own behaviour.

It is here that a serious predicament begins to manifest itself for the entire approach we have been pursuing so far. In what follows, I will try to present the problem clearly, without attempting to meet the challenge. I will be content with merely indicating a direction in which it may be solved, and leave it to another occasion to develop the solution fully. Now, in cases in which the self-control measures have worked properly and one has chosen the course of action that is in line with one’s better judgement, we will naturally want to say that it was the role of the better judgment to utilize the techniques of self-control and indirectly generate the agreeable course of action. It is generally desirable that evaluation and motivation do not come apart, and if self-control is the way to achieve the desirable condition, it should be the better judgment, which reflects the evaluation, that is responsible for exercising self-control. If better judgments did not play any part in bringing about the desirable consequence, there would not be any other part for them to play.

That is precisely the problem, however. Let us concentrate on the second of the self-control measures above, namely, that of taking the anti-alcohol pill, and ask this apparently peculiar question, “Can Mr. A take the pill after the train has arrived at N Station?” As far as better judgments are concerned, Mr. A never revises his judgment that it would be better to go home and study Kant. But in terms of motivational force, drinking at the bar has already exceeded Kant studies on his arrival at N Station. Now, the reason he has for taking the pill is that he wants to study Kant. That is, the motivation for taking the pill is identical with the motivation for Kant studies. Then, given that actions result from relative strengths of motivational forces, it would appear that Mr. A will not be able to take the pill after arriving at N Station, notwithstanding the better judgment he is holding at that moment. If he is to employ the second kind of self-control technique to prevent himself from going to the bar, it appears that he should have taken the pill before arriving at N Station.

The problem, stated in its fully generality, is that better judgments after all have
no role to play in the generation of behaviour. Let us suppose that Mr. A has taken the pill before arriving at N Station. What is it, however, that brought about this action? It seems that this case can be accounted for in the following manner. Before arriving at N Station, the motivational force of Kant studies was still stronger than that of the drink. However, the former motivation involved an awareness that there is likely to be a preference reversal in the near future. Finding this likelihood to be a grave threat to its realization, the motivation of Kant studies went ahead and took a precautionary measure, namely, that of taking the pill, which has the consequence of significantly reducing the competing motivation and thereby eliminating the possibility of preference reversal. In this story, there has never been a role for the better judgment. Motivations did all the work.

Mele discusses in detail a paradox concerning self-control, which is deeply related to this problem\(^\text{24}\). But as far as I understand him, he does not seem to be fully aware of the seriousness of the issue. He presents a sophisticated argument for establishing in effect that Mr. A can perhaps take the pill after arriving at N Station. It is an argument to show that the motivation for Kant studies, which has already become weaker than the motivation for the drink, can somehow manage to bring about an exercise of self-control, of taking the pill, that will spoil the attractiveness of the drink. This argument, if successful, will demonstrate that there may be ways to combat one’s stronger motivations. Unfortunately, however, it is clear that it does not really address our critical issue. Even if Mr. A can take the pill after arriving at N Station, according to Mele, it is an effort made by the weaker motivation to overcome the stronger one. And as such, it does not tell us anything about the role of the better judgment, which has always favoured Kant studies. Again, motivations are doing all the work, and there is no place for the better judgment in achieving the self-controlled behaviour.

As I see the situation, there are three alternatives that are open to us here. I will close this paper by sketching each of them, and indicating which of them is to be recommended. First, we might give up all talk of evaluation, better judgment, and the comprehensive perspective, and let motivations do all the work. On this view, a human agent is a composite of various desires, each of which struggles against each other to attain a stronger motivational force and secure its own realization. There will be a number of strategies that long-term desires can employ in order to prevent a preference reversal, but they have nothing to do with the long-term desires’ being more highly evaluated. This, in essence, is Ainslie’s picoeconomic model.

Picoeconomic model is a coherent option, but it is not really a very attractive position to take. If we accept this model, we will surely have to give up the definition of akratic action as action against better judgments. This may at first seem like a def-

 initial affair of no importance, but it in fact has a quite far-reaching consequence. We have a strong intuition that, in performing an akratic action, the agent goes against something that should have been respected at the time of the performance of the action. In my view, the "something" is the better judgment. The picoeconomic model, on the other hand, has no conceptual apparatus for characterizing akratic action in this manner. It is true that, with the notion of preference reversal, the model may define akratic action as action against a desire that has previously had stronger motivational force. However, thus characterized, the most that can be said about akratic action is that it involves a change of mind. At the time of action, it goes against nothing significant, and therefore we would not be able to find any serious irrationality about it. It seems to me that this drawback represents a clear limitation of the picoeconomic model.

The second alternative is to be an epiphenomenalist about evaluation, better judgment, and the comprehensive perspective. It grants them reality without allowing them to play any part in the generation of behaviour. There is little to be said for this option. If we are to be an epiphenomenalist about them, the only rationale for introducing these notions would be to characterize akratic action as action against better judgment. That would seem incredibly \textit{ad hoc} indeed.

The third alternative is to try to attribute some substantive role to evaluation, better judgment, and the comprehensive perspective. This is the option that I recommend, and I believe that it is what Mele really had in mind. The most plausible way in which this idea can be developed, in my view, is to think of self-control as a \textit{sui generis} faculty of the mind. It is a faculty that has as its input the better judgement of the agent, and that can generate some extra motivation as its output. If self-control functions successfully, the extra motivation generated will tip the motivational balance so that it will be in agreement with the better judgment. If self-control fails, akratic action may result. The crucial point here is that the exercise of self-control is not a kind of action. It is rather one element in the process that culminates in the eventual performance of the action. And unlike actions, self-control as a \textit{sui generis} faculty is not in need of its own motivation. This will enable us to avoid the problems that I discussed in relation to the example of Mr. A's attempt at pursuing his Kant studies.

Self-control, thus conceived, may come to look rather like what we ordinarily understand as "the will". Developing this idea is a challenging task indeed, but if successful, it will give a proper substantive role to evaluation, better judgment, and the comprehensive perspective, which I think are hugely important in illuminating the phenomenon of \textit{akrasia}. However, I will leave it to another occasion to work on what I have here only briefly hinted at as the most satisfactory approach.
References