Prudence for Changing Selves

KRISTER BYKVIST
Jesus College, Oxford

What is the prudentially right thing to do in situations in which our actions will shape our preferences? Suppose, for instance, that you are considering getting married, and that you know that if you get married, you will prefer being unmarried, and that if you stay unmarried, you will prefer being married. This is the problem I will deal with in this article. I will begin by explaining why preferences matter to prudence. I will then go on to discuss a couple of unsuccessful theories and see what we can learn from their mistakes. One of the most important lessons is that how you would have felt about a life had you never led it is irrelevant to the question of what you prudentially ought to do. My theory takes this into account. What counts is how you feel about a life when you are actually leading it.

I. INTRODUCTION

What is the prudentially right thing to do in situations in which our actions will shape our preferences? This is the problem I am going to discuss in this article. It is neatly illustrated by the following examples:

The career choice. Suppose that you are a philosopher who has been offered a job: a teaching position in Oxford. You must now choose between moving to Oxford and moving to Sweden where you will become a professional folk fiddler. Moreover, suppose that if you choose to take up the position in Oxford, you will come to prefer this life to being a fiddler in Sweden – playing intricate polska tunes on the fiddle would not be for you! If you choose to live in Sweden, however, then you will come to prefer living in Sweden as a fiddler to living in Oxford as an academic philosopher.[1]

Here is another example:

The unmarried person's dilemma: 'To wed or not to wed'. You are considering getting married. The problem is, however, that you know that if you get married, you will prefer being unmarried to being married. If you get married, you will adopt certain perfectionist ideas about marriage and think that your marriage does not live up to the standards. However, if you stay unmarried, you will accept less exacting requirements and prefer being married to being unmarried.

[1] Similar examples are presented in Philip Bricker, 'Prudence', *Journal of Philosophy* 77:7 (1980), pp. 381–401, and Allan Gibbard, 'Interpersonal Comparisons: Preference, Good, and the Intrinsic Reward of a Life', *Foundations of Social Choice Theory*, ed. J. Elster and A. Hylland (Cambridge, 1992).

© 2006 Cambridge University Press
doi:10.1017/S0953820806002032
Utilitas Vol. 18, No. 3, September 2006
Printed in the United Kingdom

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

As far as your own well-being is concerned, what do you have most reason to do in these situations? In other words, what is the prudentially right thing to do in these situations? To answer this question we seem to need to find a vantage point from which we can judge which life is best. But the problem is exactly how to identify this vantage point, since what is the best life seems to depend on what you choose to do. In the first example, whatever life you choose you will prefer the chosen life to the alternative life, and, in the second example, whatever life you choose you will prefer the alternative life to the chosen life.

Any complete theory of prudence must solve this problem. In this article, I am going to give the bare bones of a theory of prudence that will deal with this question in a satisfactory way. I should stress that I am not concerned with the more general question about what is the most *rational* thing to do in cases in which your actions shape your preferences. The considerations that are relevant to practical rationality need not be confined to considerations of well-being.

II. PREFERENCES MATTER

One might object that the problem of contingent preferences is spurious, because it is a mistake to think preferences or desires determine well-being. It is not true, the objection goes, that things are good for us just because we desire them, or better for us just because we prefer them. Once we have realized that desires and preferences do not play these roles, the problem vanishes. We should simply choose the better life. And we can decide which life is better for us without consulting our desires and preferences. So, the desires and preferences I will form if I move to Oxford and the ones I will form if I move to Sweden are irrelevant when it comes to deciding which life is best for me. (For ease of exposition, I will often use 'preference' and 'desire' to cover a broad class of attitudes, including wants, desires, and favourings. This is to stretch their normal usage a bit: 'preference' is usually reserved for *comparative* attitudes, preferring something *over* something else. This more restricted comparative sense will be invoked in section V.)

This response is not convincing. It is a mistake to think that the problem of contingent preferences arises only if we assume that they are the sole standard of well-being. It will arise for all theories that claim that preferences are *relevant* to an account of well-being. Since no one could reasonably deny that pleasure and displeasure are important components of well-being, the problem will arise for all theories that assume that our desires for present experiences *make* these experiences pleasant, and that our preferences for certain present experiences over others *make* the former experiences more pleasant (or

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

less unpleasant).[2] More generally, the problem will arise for all theories that claim that preferences determine the value of the *experiential* aspects of our lives – the feel and flavour of our lives. According to these theories, I should consult my preferences when I compare the value of the experiential aspects of a life as a philosopher in Oxford – the peculiar experience of marking twenty-five end-of-term exam scripts over the weekend or the experience of igniting someone's interest in philosophy in a one-to-one tutorial – with the value of the experiential aspects of being a fiddler in Sweden, e.g. the experience of fiddling until your fingers are numb.

The problem will also arise for theories that claim that preferences should have a say in deciding the relative values of *non-experiential* aspects of our lives. In particular, the problem will be a live issue for those who think that preferences should act as tie-breakers in a choice between two equally worthwhile ways of life.[3] Since the life as an academic philosopher in Oxford and the life as a fiddler in Sweden seem to be equally worthy of concern, this account would let your preferences decide which life is better for you. Similarly, since being unmarried can be as worthwhile as being married, this account would say that what life is best for you will depend on your preferences.

One might object here that even if we allow desires to play this limited role, we have to maintain that our desires determine the value of the thing, at least to some extent. But this would be to get the order of explanation wrong: things are not valuable because they are desired; they are (or should be) desired because they are valuable.[4]

[2] The view that subjective attitudes such as desires determine the pleasantness or unpleasantness of experiences is widely held. Among the most famous defenders are Gilbert Ryle, *The Concept of Mind* (New York, 1949), ch. 4; J. C. B. Gosling, *Pleasure and Desire: The Case for Hedonism Reviewed* (Oxford, 1969), chs. 3 and 10; R. B. Edwards, *Pleasures and Pains: A Theory of Qualitative Hedonism* (Ithaca, 1979); Robert Nozick, *The Examined Life* (New York, 1989), pp. 103–4; and Wayne Sumner, *Welfare, Happiness, and Ethics* (Oxford, 1996), ch. 4, sect. 3. Some of these authors would also like to define *pain* in terms of subjective attitudes. I doubt that that this is right, since, as Sumner, p. 103, points out, there are experiences that can be unpleasant without being painful, e.g. nausea, hiccups, sneezing, and itching. But even if we accept that not all experiential aspects can be defined in terms of subjective attitudes, they seem to play a crucial role in deciding whether a painful experience followed by a pleasurable experience is on the whole a pleasant sequence of experiences. To echo Mill, what else is there to decide whether the sequence of experiences is pleasant on the whole except the desires or subjective attitudes of the person having the experiences?

[3] See, for instance, Thomas Scanlon, *What We Owe to Each Other* (Cambridge, MA, 1999), pp. 47–8; Warren Quinn, *Morality and Action* (Cambridge, 1993), pp. 228–55; Joe Raz, *The Morality of Freedom* (Oxford, 1986), p. 142; and E. J. Bond, *Reason and Value* (Cambridge, 1983), pp. 16–18.

[4] This objection can be found in *Utilitarianism and Beyond*, ed. A. Sen and B. Williams (Cambridge, 1990), pp. 12–13; Thomas Scanlon, 'Moral Basis of Interpersonal Comparisons', *Interpersonal Comparisons of Well-Being*, ed. J. Elster and R. E. Roemer

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

I think there is something to this criticism even if it is not perfectly clear what this 'because' means. But we need not succumb to this criticism just because we are desire-sensitive. The objection assumes that what has value for me is the *object* of my desire. So, if I desire to eat fermented herring, then eating fermented herring is good for me. But we do not have to say that. We can say instead that what has value for me is the fact that my desire for eating herring is satisfied.[5] In other words, what has value for me is this compound state of affairs:

I want to eat fermented herring, and I eat fermented herring.

More generally, states of affairs of the following form have value for me:

I want p, and p.

Of course, if we think that only desires about things worthy of concern are relevant to well-being, we would not say that *all* states of affairs of this form are good for me. Bare desire satisfactions have no value; only the satisfactions of desires with the right content are good for me.

This theory will avoid the charge of reversing the order of explanation, because it will not say that things are good for us because, or partly because, we desire them. If the compound state of affairs that I want to eat herring and I eat herring is good for me, then it is good for me no matter whether I want this compound state of affairs to obtain. This means that you can be desire-sensitive and still deny that desires in themselves provide reasons or values.

Of course, it is still possible to deny that desires have any relevance to well-being. I think this is a mistake, but I am not going to enter the debate about whether desires have any relevance to well-being. Throughout the article, I am going to assume that desires at least play one of the limited roles explained above. This makes the results of this essay relevant to a large class of well-being theories, ranging from pure desire theories to the mixed theories that recently have gained so much support.[6] Common to all these theories is the assumption that

(Cambridge, 1993), p. 25; James Griffin, 'Against the Taste Model', *Interpersonal Comparisons of Well-Being*, ed. J. Elster and J. E. Roemer (Cambridge, 1993), pp. 48–52.

[5] In drawing this distinction, I am much influenced by W. Rabinowicz and J. Österberg, 'Value Based on Preferences. On Two Interpretations of Preference Utilitarianism', *Economics and Philosophy* 12 (1996), who distinguishes between the satisfaction interpretation and the object interpretation of preferentialism concerning intrinsic values. Ingmar Persson, 'Peter Singer on Why Persons are Irreplaceable', *Utilitas* 7 (1995), p. 64, makes a similar distinction between what he calls the relational view and the object view.

[6] Bengt Brülde, 'The Human Good' (PhD dissertation, University of Gothenburg, 1998), ch., 8; Ronald Dworkin, *Sovereign Virtue*, (Harvard, 2002), ch 6; Stephen Darwall, 'Valuing Activity', *Human Flourishing*, ed. E. Paul, F. Miller, and J. Paul (Cambridge, 1999); Richard Kraut, 'Desire and the Human Good', *Proceedings and Addresses of the American Philosophical Association* 68:2 (1994), pp. 39–54, p. 45 and n. 13; Derek Parfit, *Reasons and Persons* (Oxford, 1992), Appendix I, p. 502; Susan Wolf, 'Happiness and

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

something can be good or better for a person only if it is endorsed by the person.

III. THE PROBLEM SIMPLIFIED

The examples I presented at the beginning of this article share a common feature: my future actual and counterfactual selves will disagree about what is the best world to live in. Now in real life we might have more conflicts to take into account. There might, for instance, be a conflict between what my present self wants and what my future selves want. There might also be a conflict between my present self and my past selves. My past selves might even disagree between themselves. Finally, my choices might affect not just the preferences of my future selves, but also how many future selves I will have. This means that the problem of changing selves gets really messy if we take into account every possible kind of preference change. To make things a bit more manageable, I will deal with a simplified version of the problem. I will make the following assumptions:

(a) Same number case. The agent's choice will not affect the number of preferences.
(b) Past indifference. The past selves are indifferent or undecided between the outcomes available to the present self.
(c) Present indifference. The present self is indifferent or undecided between the outcomes available to the present self.
(d) Future invariance. The preferences of the future selves in an outcome do not differ.

By assuming (a), we can put aside the problem of whether we should create preferences in order to satisfy them. Assumptions (b), (c), and (d) make sure that we do not have any intertemporal conflicts. We can therefore focus exclusively on the problem of preference change across available outcomes. A complete theory of prudence must of course have something to say about preference change within an outcome. But this is the topic for another article.

(e) Only global preferences count.

A global preference is a preference about a part of one's life taken as a whole, or a life as a whole. It seems reasonable to give priority to these preferences since they are more comprehensive than local preferences about particular states of affairs. Global preferences concern the way these states of affairs make up bigger wholes, for instance, the way

Meaning: Two Aspects of the Good Life', *Social Philosophy and Policy* 14 (1997), pp. 207–25.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

they unfold in time and make up temporal wholes. To simplify the discussion, I will assume that *only* global preferences count, and only those that are about our whole lives.

(f) The diachronic identity of the agent is not at stake.

A sharp enough break in the psychological connections between selves at different times may make us wonder whether we are still talking about the same person. It is of course crucial to prudence, seen as a theory about the normative significance of the agent's *own* well-being, that the future selves will still be 'her'. I will therefore only talk about cases in which it is clear that the diachronic identity of the agent is not threatened. It should be stressed that my talk about selves here is metaphorical. I do not wish to take a stand on how diachronic identity should be analysed, whether, for instance, persons are endurers or perdurers. So, for instance, when I say that my future self would do this or that or prefer this or that, I am only saying that I will at some point in the future do this or that or prefer this or that.[7]

IV. POSSIBLE SOLUTIONS

A theory of prudence comes in two parts: an axiological theory that tells you when an outcome is better or worse for you, and a criterion of prudential rightness that tells how these outcome-evaluations determine what you prudentially ought to do. On the most common interpretation, the maximizing theory of prudence, the normative component will simply tell you to choose the action whose outcome is better for you than that of any alternative action. More exactly, the criterion of rightness breaks down into the following conditions.

An action is prudentially obligatory for me iff its outcome is better for me than that of any alternative action.[8]

[7] Another simplifying assumption I make is that each action has a determinate world associated with it, the world that would be the case if the action were performed. This is of course in general false. Often there are many worlds that *might* be the case if an action were performed, but no one world that *would* be the case. However, my discussion can be transformed into an indeterministic framework. It is just that this requires more complicated formulations of the various positions.

[8] Roger Crisp has pointed out to me that it seems odd to say that you have an obligation to be prudent. One possible reason why this might sound odd is that 'obligation' is usually used in moral contexts. But we could instead talk about what we prudentially ought to do. 'Ought' is widely used in non-moral contexts. Another reason why 'obligation' as well as 'ought' might seem out of place in the prudential context is that we seem to be naturally inclined to do what is best for us, and obligations and oughts are only meaningful in contexts where what we ought to do does not coincide with what we naturally are disposed to do anyway. Obligations and oughts have a corrective or restrictive function. There might be something to this idea about the proper role of obligations or oughts. But note that prudence, on most accounts, tells you to promote your well-being throughout your

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

An action is prudentially right for me iff its outcome is not worse for me than that of any alternative action.

An action is prudentially wrong for me iff it is not right for me.

I will begin by looking at a couple of solutions that assume this normative framework but vary the standard for when one outcome is better than another.

A. Ideal preferences

One could argue that the problem of preference change is evaded if we claim that our ideal self should decide the matter. What I should do in a case where my preferences are contingent on my actions is to consult my ideal self and find out which outcome he would desire. How do I find this out? I ask myself what I *would* desire if I had all relevant information and reasoned with the greatest possible care.[9] Idealism, as we might call it, would then imply the following principles:

- If my ideal self prefers my life in w to my life in w', then w is better for me than w'.
- If my ideal self is indifferent between my life in w and my life in w', then w is equally as good for me as w'.

I have nothing against consulting my ideal self, provided we are not too different. It is just that often there is no unique ideal self to consult. Since what I would desire if I had full information and reasoned carefully depends on my actual psychological make-up, what is my ideal self will change from one situation to another. In particular, my ideal self will differ from one time to another and from one world to another. So, the question, then, is which ideal self I should consult in a situation where my actions will shape my future selves, including my ideal selves.

One possibility is to turn to the *present* ideal self and accept his verdict. In other words, I should ask myself what I, with my present psychological make-up, would want if I were fully informed and reasoned carefully. This brings us to the next theory: presentism.

whole life and give equal weight to your well-being no matter when you enjoy it. This is not something that we are naturally inclined to do. We often give more weight to what happens to us in the near future. Thus, obligations and oughts can fulfil their corrective function even in the prudential domain.

[9] Views of this type are developed in Richard Brandt, *A Theory of the Good and the Right* (Oxford, 1979), chs. 6–7; John Rawls, *A Theory of Justice* (Harvard, 1971), pp. 417ff.; and John Harsanyi, 'Morality and the Theory of Rational Behaviour', *Utilitarianism and Beyond*, ed. A. Sen, and B. Williams (Cambridge, 1990), p. 55; Peter Railton, 'Facts and Values', *Philosophical Topics* 14 (1986), has developed a slightly different view according to which an individual's good consists in what his ideal self would want his actual self to want or pursue. This version is still vulnerable to the objections stated in the main text.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

B. Presentism

This theory would judge outcomes by the preferences of the present self, either the actual or the ideal. More exactly, it would entail the following principles:

- If my present self prefers my life in w to my life in w', then w is better for me than w'.
- If my present self is indifferent between my life in w and my life in w', then w is equally as good for me as w'.

This theory does not seem to be especially attractive, however. It would entail that in the career-example and the marriage-dilemma it does not matter what I do, and this holds irrespective of how my future selves will feel about their lives. So, for instance, suppose that if I were to move to Sweden, I would hate the life as a fiddler but hate it less than the life as an academic philosopher in Oxford. If I were to move to Oxford, on the other hand, I would love these two options but love the life as a philosopher more. Now, presentism would entail that it does not matter what I do, since my present self is indifferent between these options. More generally, presentism will not give any weight to what happens to our future selves. Presentism only looks at the interests of our present selves.

One could object that presentism can give some weight to the well-being of future selves since a person would care about his future selves if he were fully informed. But if he values spontaneity and thinks that the moment itself is all that counts, then there seems to be no guarantee that his obsession with the present moment would disappear if he were fully informed and reasoned carefully.[10]

Presentism might look attractive if we fail to distinguish between *what is rational for an agent to do* and what is good for him.[11] It might be argued that what is rational for a person to do depends crucially on his present attitudes. For instance, whether it is rational for me *now* to prepare to go to a concert a year from now may depend decisively on my present attitudes towards this future event. This dependency is uncontroversial if we talk about what it is rational to do *given our present beliefs and desires*. But whether it will benefit me to go to the concert does not seem to depend on my present attitudes. So, if the prudentially right thing to do is determined by all prudential

[10] It would not help to define 'careful reasoning' in terms of a proper concern for one's future well-being, since this would give us a circular account of well-being. Well-being is defined by present ideal desires, which in turn are defined by a proper concern for well-being.

[11] Here I follow Stephen Darwall, *Welfare and Rational Care* (Princeton, 2002), pp. 34–5.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

values, including the future ones, it is quite possible that prudence and practical rationality will give us conflicting prescriptions in many cases.

No matter whether presentism is seen as a theory about prudence or a theory about practical rationality, accepting presentism makes it difficult to plan your life according to what is prudentially or rationally best for you, since what is best for you will vary with time. You may end up implementing a plan that you will have no reason to complete at a later time. Assume that you have a desire to write a PhD-thesis in philosophy and that you therefore start preparing for this. You may then find yourself with no reason to complete your thesis once you have started working on it. For by that time you might have changed your mind about philosophy. This change need not be irrational, e.g. based on a mistaken belief about what it means to do philosophy. It might just be that you acquired a new and stronger interest in something else, investment banking, for instance. Now, assume that you know that you will change your mind about your philosophy career. On presentist assumptions, how should you then plan your life?

C. Actualism

Presentism seems to be too restrictive. Future preferences seem to be important as well. We could take them into account by postulating that all and only past, present, and future preferences that exist in the *actual world* should count.[12] This theory would imply the following principles:

- If my actual selves – past, present, and future – weakly prefer my life w to my life in w', and at least one actual self strictly prefers my life in w to my life in w', then w is better for me than w'.[13]
- If my actual selves are indifferent between my life in w and my life in w', then w is equally as good for me as w'.

One might find this theory attractive because of our natural bias towards the actual. As Barry points out:

> It seems rational in a quite straightforward sense to want to satisfy the preferences I have. But there is nothing to suggest that I would be equally rational to want to become some entirely different kind of person who would

[12] This theory is discussed in Bricker, 'Prudence', and something very close to this theory is defended in Brian Barry, 'Utilitarianism and Preference Change', *Utilitas* 1 (1989), pp. 272–82, and R. M. Adams, 'Existence, Self-Interest, and the Problem of Evil', *Noûs* 13 (1979), pp. 60–1.

[13] A person weakly prefers one thing to another if he either prefers the first thing or is indifferent between the two things.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

be more satisfied with the things that would be available to that person than I am now with the things available to me.[14]

A similar point is made by Adams who claims that it does not seem reasonable for Helen Keller, as an adult, to wish, for her own sake, that she had never been blind or deaf. A life without her actual handicaps would have contained very little of the experiences and achievements she cared about in her actual life. Why should she wish for such a radically different life?

I very much doubt that these considerations support an actualist *well-being* theory. Barry and Adams talk about what is 'rational' or 'reasonable' to want or wish, but this need not coincide with what is good or better for a person. What is rational or reasonable to want is not exclusively determined by considerations of well-being. To see this point, we do not have to consider a counterfactual life in which your concerns are radically different. Consider a possible life which is far better for you than your actual life, but in which your actual concerns are retained. It does not follow that it is reasonable or rational *all things considered* to want or wish for this life. If it is not easily available to you, it may not be reasonable to want. But even if it is readily available, it may not be reasonable to want or wish for this life, since it might be worse in terms of non-prudential values. It might, for instance, be worse in terms of moral and non-moral achievements.

There are two serious objections to an actualist theory of prudence. First, the theory will entail that whether an act is prudentially right might depend on whether it is performed. To see this, go back to the career-example. If I were to move to Oxford, then my actual preferences in this scenario would favour my move. Since my actual desires determine the values of outcomes, in this scenario moving to Oxford is better than moving to Sweden. But this means that I ought to move. On the other hand, if I were to move to Sweden, a different scenario would be realized, and my actual preferences in this scenario would not favour my move to Oxford. So, in this alternative scenario it is better for me to move to Sweden and thus I ought not to move to Oxford. So, the conclusion is that I ought to move to Oxford only if I will move to Oxford.

This normative variance is troubling since it is difficult to use a prudential theory as a guide to action if whether you ought to do something depends on whether you will do it. When you use a theory as a guide to action, you use the theory in your deliberations about what to do. In particular, you use it to decide which options are obligatory or permissible. On the basis of this deliberation you then make up your

[14] Barry, 'Utilitarianism and Preference Change', p. 280.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

mind and decide what to do. But if an action's rightness depends on whether it is performed, then in order to decide whether an action is right you first have to know whether or not you are going perform it. But there is no point in deliberating about whether to perform an action if either you believe that you will perform it, or you believe that you will not perform it. If you believe that you will perform the action, the issue is settled for you, and there is no point in deliberating about it further. If you believe that you will not perform the action, the action is no longer a serious possibility, i.e. something that is compatible with what you believe (even if it might be something you can do); so again there is no point in deliberating about whether to perform it.[15]

The claim here is *not* that once you have formed the belief (or disbelief) that you are going to do A, then you are no longer able to deliberate about whether to do A. If you give up the belief or the disbelief, you may start deliberating again. The claim is rather that *while* you are in the grip of the belief or disbelief that you will do A, it is not possible for you to deliberate about whether to do A. Or at least, this is not possible if you are *rational*. For rational agents, belief or disbelief about what they are going to do excludes wondering about whether to do it. The rational wondering whether to do A cannot cohabit in one's mind with either the belief or the disbelief that one is going to do A.[16]

A second and related objection is that actualism generates dilemmas: some situations involve unavoidable wrongdoing in the sense that whatever you were to do, you would do something that would be prudentially wrong. This is brought out by the marriage example. Suppose that you get married. Since you will then regret this choice and prefer not being married to being married, in this situation not being married is better for you than being married. Hence it is wrong for you to get married. Suppose instead that you do not get married. Then you will prefer being married to not being married and thus it is

[15] This argument is spelled out in Erik Carlson, *Consequentialism Reconsidered* (Dordrecht, 1995), pp. 101–2. Bricker, 'Prudence', p. 395, touches on a similar argument. The general idea that the prediction of one's actions crowds out deliberation has widespread support. See, for instance, Alvin Goldman, *A Theory of Human Action* (Princeton, 1970), p. 194, and Richard Taylor, *Action and Purpose* (Englewood Cliffs, 1966), p. 174. Cf. also Carl Ginet, 'Can the Will Be Caused?', *The Philosophical Review* 71 (1962); Isaac Levi, 'Consequentialism and Sequential Choice', *Covenant of Reason* (Cambridge, 1997), pp. 70–101; and Frederic Schick, 'Self-knowledge, Uncertainty, and Choice', *Decision, Probability, and Utility*, ed. P. Gärdenfors and N.-E. Sahlin (Cambridge, 1988).

[16] In Neil Tennant, *The Taming of the True* (Oxford, 1997), pp. 252–4, this is generalized to hold for all rational wonderings no matter what they concern. The rational wondering whether p excludes both the belief and the disbelief that p.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

better for you to get married. But this means that it is wrong for you not to get married.

What we have here is a kind of prudential dilemma, but one that differs from the usual kind. Normally, a dilemma is seen as a situation in which all available actions *are* wrong. This is not the situation here. No matter how you act, there is an available act that *is* right. If I marry, not getting married is right. If I do not get married, getting married is right. But this is not much comfort, for you cannot act in such a way that you comply with the theory: there is no action such that if you were to perform that action you would act rightly. You *would* be damned if you *were* to get married, and you *would* be damned if you *were* not to get married.

Why is it important that a normative theory be possible to comply with? Again, the answer has to do with the practical role of a normative theory. A normative theory is supposed to guide action and thus be usable at least by agents who are ideal in the sense that they lack all cognitive and motivational defects.[17] But if we cannot comply with a theory in a situation, we cannot be guided by it in that situation, since to be guided by a theory is to use that theory when deliberating about what to do and on the basis of this deliberation act in compliance with the theory.

D. Ratificationism

Perhaps this shiftiness of actualism could be remedied if we modified the normative component of the theory. One possibility is to define outcome-values by your actual preferences, but change the definition of prudential rightness so that an action is prudentially right iff is ratifiable.[18]

[17] All plausible normative theories are sometimes not usable by agents, since their application often requires information the agent lacks and might not be able to gather and process in a limited time. We all know that utilitarians have this problem, because they need information about how their actions will affect *future* well-being. But deontologists will face similar problems. For instance, they may need to know a lot about *past* suffering and wrongdoings in order to know whom to compensate, and how to judge the present distribution of goods and burdens in society. More generally, since any reasonable moral theory must incorporate a principle of non-malevolence, the problem about knowing the well-being of future people will reappear in the deontological framework. Only clearly implausible theories, such as one that tells us to do whatever you feel like doing, will be easy for any agent in any situation to apply correctly.

[18] The notion of ratifiability is adopted with slight alterations from decision theory, which defines a ratifiable option as one that has maximal expected utility on the assumption that the decision-maker has already *decided* to perform the action. I depart from this usage by talking about the *actual* maximal *prudential* value given the *performance* of the agent's action. For more on the decision-theoretical notion of ratifiability, see Richard Jeffrey, *The Logic of Decision*, 2nd rev. edn. (Chicago, 1983), p. 16, and 'The Logic of Decision Defended', *Synthese* 48 (1981), pp. 473–92.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

> An action is *ratifiable* for an agent iff, were the outcome of the action realized, this outcome would be better for the agent than that of any alternative actions.

Since actual preferences are supposed to decide what outcome is better or worse than another, ratificationism in effect says that an act is prudentially right if it is self-supporting in the sense that were it to be performed it would be supported by the preferences that would *then* be actual. In the career-example, moving to Oxford and moving to Sweden are both ratifiable actions. If you moved to Oxford, then this option would be sanctioned by the preferences you were to develop in this situation. Consequently, this move would be better for you than moving to Sweden. Moving to Oxford is thus ratifiable. If you moved to Sweden, then this move would be sanctioned by the preferences you were to develop in this alternative situation. The move to Sweden would therefore be better for you than the move to Oxford, and moving to Sweden is therefore also a ratifiable option. The conclusion is thus that moving to Oxford and moving to Sweden are both permissible options.

This may seem reasonable, but only if we assume that I would be equally interested in these different lives. Suppose that if I moved to Oxford I would love my academic life in Oxford, and love it more than my musical career in Sweden. If I moved to Sweden, on the other hand, I would hate my life there but hate it less than an academic life in Oxford. Then it does not seem reasonable to say that it is prudentially right for me to move to Sweden. If I move to Sweden I will find myself leading a life I hate, whereas if I move to Oxford I will find myself leading a life I love. Surely, a theory of prudence should be sensitive to these differences in attitudes.

Another problem with ratificationism is that sometimes there is no ratifiable action. The marriage-example shows this. If I get married, I will prefer not to get married, and therefore this outcome will be worse than the outcome of not getting married. If I do not get married, I will prefer getting married; so, this outcome will be worse than getting married. Here ratificationism will imply that no action is right.[19] But surely there may still be a right thing to do in this kind of case. Assume, for instance, that if I were to marry, I would hate this and hate it more than not getting married, whereas if I were not to marry, I would love this but love it less than getting married. Surely, I should stay unmarried and lead a life I will come to love instead of choosing a life I will come to hate.

[19] That no action is right does not mean that all actions are wrong. It might only mean that they lack normative status. Ratificationism will only generate a prudential dilemma if it claims that an action is prudentially wrong iff it is not ratifiable.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

E. Necessitarianism

On this account I should take into consideration only my necessary selves, i.e. those of my selves that will exist no matter how I act.[20] This theory will of course include my past and present selves, but how do I identify future selves across different worlds? For my purposes here it is enough to say that two person-stages from two different worlds are the same self iff they are parts of the same person and share the same time and preferences.

- If all my necessary selves weakly prefer my life in w to my life in w', and at least one necessary self prefers my life in w to my life in w', then w is better for me than w'.
- If my necessary selves are indifferent between my life in w and my life in w', then w is equally as good for me as w'.

This theory differs from actualism in that it considers non-actual selves and it differs from presentism in that it considers future and past selves. However, in practice, this theory will come very close to presentism. For instance, it will give the same results for the career-example and the marriage-example. In both these examples my future selves are contingent, which means that only the past and present preferences should rule. More generally, necessitarianism and presentism will have the same implications in practice since in most cases we have the opportunity to kill ourselves. Even if we cannot do it immediately there is often an available course of action that will lead to this outcome. But if this is so, then most future selves will be contingent on our actions; they will exist only if we do not kill ourselves. So, necessitarianism would imply that they should not be taken into account. But the mere fact that a person can kill himself should not let him off the hook. He still has prudential reasons to take care of his future self.

F. 'The survival of the strongest preference'

Both actualism and necessitarianism are too restrictive. Not only actual or necessary preferences count. The most inclusive theory of prudence would of course count all selves no matter whether they are actual, necessary, contingent, past, present, or future. I think this is a good starting point for an acceptable theory of prudence. However, what shall we do when our selves disagree? In particular, what shall we do

[20] This theory is defended in Richard Arneson, 'Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare', *Philosophy and Public Affairs* 19:2 (1990), p. 174: 'What is prudent for a person to choose at a given time depends on her current actual preferences (the preferences she had, has, and will have independently of the choice to be made)'. A similar idea applied to the rationality of actions is discussed in Ingmar Persson, 'Rationality and Maximization of Satisfaction', *Noûs* 22 (1998), pp. 537–54 (p. 543).

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

when our future selves disagree, which is the case we are focusing on? One option would be to go for the self with the strongest preference. In other words,

- If my past and present preferences are indifferent between my life in w and my life in w', and my future w-self prefers my life in w to my life in w' more strongly than my future w'-self prefers my life in w' to my life in w, then w is better for me than w'.

What does it mean to say that one preference is stronger than another? If we assume that the preferences of the different selves can be represented by utility functions, and that it is meaningful to compare utility differences from different selves, then we could say that self i's preference for x over y is stronger than self j's preference for y over x iff $U^i(x) - U^i(y) > U^j(y) - U^j(x)$. Of course, this is a big 'if'. However, I will show that even if we assume that we make sense of these assumptions, this theory is not acceptable. Going for the strongest preference might sometimes mean that I should act so that I end up leading a life which I will hate. To see this, consider the following version of the marriage-case.

I do marry	I do not marry
I prefer being unmarried to being married.	I prefer being married to being unmarried.
U(being unmarried) − U(being married) = 5.	U(being married) − U(being unmarried) = 10.
I hate being unmarried, but I hate being married more.	I love being unmarried, but I love being married more.

Now, if we should go for my strongest preference, then I should get married even though I will hate that life once I am married. This cannot be right. If I were to remain unmarried I would love that, which suggests that this is exactly what I should do. What counts is how I feel about getting married when I am married. How I would have felt about being married had I never got married is irrelevant to the question of what I prudentially ought to do. To think that it should be relevant would be to accept a kind of *cross-world intervention*. This intervention is not acceptable when the alternative selves are both concerned with worthwhile objects and equally rational. I will therefore try to formulate an inclusive theory that counts all my selves but does not have this implication.

G. Dominance

How will this theory look? One might think that any reasonable inclusive theory must satisfy the following dominance principle:

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

- If my past and present selves are indifferent between my life in w and my life in w', and both my future w-self and my future w'-self prefer w to w', then w is better for me than w'.

This principle looks very attractive – it tells us to go by the comparative preferences of our future selves if they agree and no other self disagrees. But appearances are deceiving; this is not an acceptable principle. Too see this, consider the following version of the marriage-case. Suppose that if you were to get married, you would hate being married, but hate being unmarried more. Suppose further that if you were not to get married, you would love being unmarried but love being married more.

Here your future selves agree: they both prefer being married. However, if you were to go by your comparative preferences, you should get married even though you will hate that life once you are married. Again, if you were to remain unmarried you would love that life. Surely, not getting married is therefore the best option for you. This shows that cross-world intervention might be problematic even in cases where the intervening preferences agree.

V. MY SOLUTION

I will not give a complete theory of prudence here, just the basic principles that can handle the cases at hand. Roughly put, my theory states that one should effect a reconciliation between all of one's selves (past, present, future contingent, future necessary) and the worlds these selves inhabit. No cross-world intervention is allowed: a life should be judged only by the desires one would develop if one were to lead that life.[21] More exactly, this theory will embrace the following principles:

- If my past and present selves are indifferent between my life in w and my life in w', and my future w-self wants my life in w more

[21] Here I am in agreement with Bricker, 'Prudence'. However, my solution differs from his in many crucial respects. Whereas Bricker defines the prudentially right action for a person in terms of his judgements about what is good, I define it in terms of the person's non-cognitive attitudes (and possibly the worthiness of the desired objects). Bricker's definition has the absurd implications that I can act prudentially right without making myself better off in terms of my own well-being. Obviously, if I act in accordance with my evaluative judgements there is no guarantee that I promote my own well-being. My judgements may be about impersonal goodness rather than my own good, and even if they are about my own good, they may be false. Another difference is that Bricker focuses on preferences (two-place attitudes) whereas I focus on one-place attitudes (see below in the text). Finally, Bricker does not allow for any discrimination in terms of the value of the desired objects. This makes his approach vulnerable to the objection that prudence may tell us to adjust our preferences to things that are not worth pursuing. For more on this objection, see the last section of my article.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

than my future w'-self wants my life in w', then w is better for me than w'.

- If my past and present selves are indifferent between my life in w and my life in w', and my future w-self wants my life in w as much as my future w'-self wants my life in w', then w is equally as good for me as w'.

Now, these principles will not say much if I cannot explain what it means to say that one self wants one thing more than (or as much as) another self wants another thing. Note that what we are comparing here are *one-place* attitudes (one self's attitude towards x versus another self's attitudes towards y), rather than *two-place* attitudes (one self's preference for x over y versus another self's preference for y over x). To explain this comparison, we need to focus on the *polarity* of attitudes. There are three kinds of (one-place) attitudes we can take towards a state of affairs: we can favour it, disfavour it, or be neutral towards it. Roughly put, to favour x is to be positively oriented towards x in your actions, emotions, feelings or evaluative responses. So, if you favour x, you tend to be motivated to bring it about, be glad and happy when you think it obtains, have pleasant thoughts about it, or see it in a good light. To disfavour x is then to be negatively oriented towards x in your actions, emotions, feelings or evaluative responses. You tend to be motivated not to bring it about, be sad and unhappy when you think it obtains, have unpleasant thoughts about it, or see it in a bad light. Finally, to be indifferent towards x is to show indifference towards x in your actions, emotions, feelings, or evaluative responses.

With these notions at hand we can now explain the comparison of attitudes in the following way:

My x-self wants x more than my y-self wants y iff

(i) my x-self *favours* x *more* than my y-self *favours* y, or
(ii) my x-self *disfavours* x *less* than my y-self *disfavours* y, or
(iii) my x-self *favours* x and my y-self *disfavours* y, or
(iv) my x-self *favours* x and my y-self is *indifferent* towards y, or
(v) my x-self is *indifferent* towards x and my y-self *disfavours* y.

My x-self wants x as much as my y-self wants y iff

(i) my x-self *favours* x *as much as* my y-self *favours* y, or
(ii) my x-self *disfavours* x *as much as* my y-self *disfavours* y, or
(iii) my x-self is *indifferent* towards x and my y-self is *indifferent* towards y.

One might object to this account on the grounds that it seems to presuppose that absolute attitudes are primitive and cannot be reduced to comparative ones. But this is not so. I could happily define favouring,

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

disfavouring, and indifference in terms of preference. Here is how:

> S favours x iff S prefers x to something he is indifferent towards.
>
> S disfavours x iff S prefers y to x and y is something he is indifferent towards.
>
> S is indifferent towards x iff S is indifferent between x and the negation of x.[22]

Of course, I do have to assume that it makes sense to compare attitudes of different contingent selves of the same person. I see no problem in comparing absolute attitudes with different polarity: favourings with disfavourings, favourings with indifferent attitudes, and disfavourings with indifferent attitudes. What creates a possible problem are comparisons of absolute attitudes that have the same positive or negative polarity. It is here that one might think that it is better to rely on preferences only, since then one only needs to make sense of comparisons of preferences. What does it mean to say that one contingent self favours x more than another contingent self favours y?

In reply, I would say that if we assume that it is meaningful to compare preferences across different selves of the same person, and if favourings can be defined in terms of preferences along the lines presented above, then a comparison of favourings boils down to a comparison of preferences. To decide whether my x-self favours x more than my y-self favours y, we should compare my x-self's preference for x over something my x-self is indifferent towards with my y-self's preference for y over something my y-self is indifferent towards. Comparisons of favourings will then be comparisons of preference differences. The same reasoning can of course be applied to comparisons of disfavourings. I cannot, therefore, see that the friend of preferences has an advantage. We are in the same boat. We both need to make sense of comparisons of preference differences.

I hope that it is reasonably clear what my theory states. It is now time to look at the implications of my theory. First of all, it will give the right implications in the marriage-case above in section IV.G. Since I would favour being unmarried if I stayed unmarried and would disfavour getting married if I got married, my unmarried self wants to stay unmarried more than my married self wants to get married. And my theory will therefore prescribe not getting married. It will also give a plausible solution to the problems I stated in the beginning, *The Career Choice* and *The Unmarried Person's Dilemma*. Whether I should move to Sweden cannot be decided by looking at the preferences

[22] Here I follow Roderick Chisholm, 'The Descriptive Element in the Concept of Action', *The Journal of Philosophy* 61 (1964), pp. 620–3.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

of my contingent future selves. Whether I should move depends on what my fiddler-self feels about his life as a folk musician and what my philosopher-self feels about his life as a philosopher. So, for instance, if I would hate my life as a fiddler if I moved to Sweden, but love my life as an academic philosopher if I moved to Oxford, I should move to Oxford. Similarly, whether I should get married depends on what my unmarried self feels about his bachelor life and what my married self feels about his life in matrimony. For instance, if I would love my life if I stayed unmarried but hate my life if I got married, then I should stay unmarried.

My theory will also do better than the other theories discussed. It will take account of the desires of future selves and thus do better than presentism. Since, according to my theory, what is the best life for you does not vary over time, we can avoid the problems about moral planning that afflicted presentism. We can also avoid the problems about normative variance and action-guidance that afflicted actualism. Since my theory will not allow that what is the best life for you will depend on what option you choose, it will not imply that the rightness of an action depends on whether the action is performed.

Finally, it will not allow cross-world interventions, since an outcome is judged only by the desires one would develop there, and not by the desires or preferences one would have developed in an alternative outcome.

VI. AN OBVIOUS OBJECTION: THE STOIC CONCLUSION

The obvious problem with my account is that it seems to imply the Stoic conclusion: if you cannot make the world conform to your desires, make your desires conform to the world. Even if prudence is about finding a match between oneself, or one's selves, and the world, there are limits to the extent which we can adjust to the world in the name of prudence. The problem with my theory is that it does not seem to set any limits. Whenever you cannot change some aspect of the world so that it matches your current desires you should, if possible, try to favour that aspect. These adjustments are especially problematic when the states of affairs we adjust to are things that are not worth pursuing, either because they are worthless or because they are worthy of aversion. Clear examples would be Rawls's example of counting the blades of grass and various kinds of abuse, degradation and humiliation.

It is true that my theory implies this kind of adjustment if we assume that all desires count, no matter what content they have, and how they are formed. But we do not have to assume this. If we count only rational desires and weed out those that are based on false beliefs or faulty reasoning, we will be able to avoid some of the most troubling

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press

adjustments, since, typically, in cases where people accept abuse and degradation they have false beliefs – they think they do not deserve a better life or that they ought to endure this suffering. It is easier to avoid the problematic implications if we adopt a discriminative version of my theory that tells us to satisfy desires that concern the right kind of things, e.g. things that are worthy of concern. Since abuse and degradation are not things worthy of concern, favouring these things does not make us better off. However, when the fixed state of affairs is something worthy of concern, it does not seem unreasonable to think that taking an interest in it will make us better off. For instance, it does not seem to be unreasonable to think that if you are unable to change some of your good character traits that are worthy of concern, it will make you better off if you start to favour them instead of being indifferent towards them. Hence, once we make a distinction between rational and irrational desires, and between objects that are and are not worthy of concern the problematic aspects of the Stoic conclusion will disappear.

VII. CONCLUSIONS

We have thus solved the problem of finding a vantage point from which to judge which life is best for a person in cases where his choices shape his preferences and desires. The mistake was to look for a single vantage point identified with the comparative preferences of one of the person's many future selves. Instead, each of the person's contingent selves should have a say. In order to decide whether a life x is better for him than another life y, we should consider his x-self's attitudes towards x and compare those with his y-self's attitudes towards y. If his x-self wants x more than his y-self wants y, then x is better for him than y (at least if we assume that both x and y are objectively desirable).

Of course, my solution does not address all pressing problems concerning preference change. Most importantly, it does not deal with preference conflicts across time and the creation and satisfaction of new preferences. But I hope to have shown that the theory defended in this article is one important building block in a complete theory of prudence.[23]

krister.bykvist@jesus.ox.ac.uk

[23] Earlier versions of this article were presented at the Oxford Philosophical Society, the Moral Sciences Club, Cambridge, and the philosophy departments at the University of Edinburgh and the University of Birmingham. I would like to thank the audiences at these seminars for helpful discussion of my paper. I am especially grateful to Gustaf Arrhenius, John Broome, Roger Crisp, Tim Lewens, and Derek Parfit for very useful comments and criticisms.

https://doi.org/10.1017/S0953820806002032 Published online by Cambridge University Press