ORIGINAL ARTICLE

Personhood and future belief: two arguments for something like Reflection

Simon J. Evnine

Abstract This paper offers two new arguments for a version of Reflection, the principle that says, roughly, that if one knew now what one would believe in the future, one ought to believe it now. The most prominent existing argument for the principle is the coherence-based Dutch Strategy argument advanced by Bas van Fraassen (and others). My two arguments are quite different. The first is a truth-based argument. On the basis of two substantive premises, that people's beliefs generally get better over time and that being a person requires having knowledge of this fact, it concludes that it is rational to treat your future selves as experts. The second argument is a transcendental one. Being a person requires being able to engage in plans and projects. But these cannot be meaningfully undertaken unless one has Reflection-like expectations about one's future beliefs. Hence, satisfaction of Reflection is necessary for being a person. Together, the arguments show that satisfaction of Reflection is both rational and necessary for persons.

Keywords Reflection · Epistemology · Personhood · Belief · Plans

1 Introduction

In this paper, I want to present two distinct, but complementary, arguments for the view that

(SLR) one should treat one's future selves as experts.

By treating someone as an expert I mean that if one came to know what the expert believed about something in her area of expertise and did not come to know any

S. J. Evnine (✉)
Department of Philosophy, University of Miami, P.O. Box 248054, Coral Gables
FL 33124-4670, USA
e-mail: sevnine@miami.edu

reason why one should not, one would adopt that belief oneself (if one did not already have it). So SLR holds that if one were to come to know one's future beliefs and not learn of any reason why they should not, those beliefs should become one's current beliefs. Satisfying SLR means giving a vote of confidence to your future selves *qua* epistemic agents.[1]

''SLR'' stands for ''something like Reflection.'' This is because SLR is, indeed, something like Reflection, a principle introduced into the philosophical literature and defended by Van Fraassen (1984 and 1995):

$$(\text{Reflection}) P_{a,t}(Q|P_{a,t'}(Q) = r) = r.$$

This says that a person a, at a time t, should accord a probability r to Q conditionally on the hypothesis that a will accord r to Q at some later time, t'. Just how like Reflection is SLR?[2] I would like to think of SLR simply as an informal interpretation of Reflection. Suppose one satisfies Reflection. In that case, if one came to be certain what one's future opinion would be concerning a given proposition Q, and learnt nothing that led one to revise the conditional probability determined by Reflection, then it would have to be one's current unconditional opinion with regards to Q (if it were not already). This would seem to merit the description of treating one's future self as an expert, glossed as I glossed it above.[3]

There are, however, some differences between SLR and Reflection. First and most obviously, is the fact of SLR's informality. This will be important. I shall argue, in the course of this article, that some of the most interesting philosophical issues that arise in a discussion of SLR cannot arise with the official version of Reflection precisely because it is couched in terms of a formal apparatus (second-order conditional probability) that does not have the resources to express various qualifications that SLR will require. Secondly, Reflection is stated in terms of degrees of belief, represented as probabilities, whereas SLR, through my gloss of the notion of taking as an expert, deals simply with belief. Reflection will extend to full belief if one treats full belief as probability above a certain threshold.[4] If one does not, then a separate but analogous principle would be needed to cover the case of full belief. In my gloss of SLR, I intend ''belief'' to cover both full and partial belief.[5] Despite these differences, I shall proceed on the assumption that even if SLR is not simply an informal interpretation of Reflection, the two principles are close enough to make a discussion of one potentially relevant to a discussion of the other. In fact, because the defense of my principle, SLR, is so closely intertwined

---

[1] I say ''future selves'' without meaning thereby to imply any metaphysics. ''Oneself at future times'' would have done as well. See Evnine (2005) for further discussion of this choice.

[2] I am grateful to a referee for making me think about this more carefully.

[3] The idea of treating Reflection in terms of the notion of expertise derives from Gaifman (1988), especially, 200–204.

[4] This has been dubbed the Lockean Thesis by Foley (1993), 140ff.

[5] I do not, however, commit myself to seeing partial belief in terms of probability. So that is another difference between Reflection and SLR, though not one that is material to the arguments of this paper.

with discussions of Reflection in the literature, it will be cumbersome to keep distinguishing the two principles. Points raised in connection with the official version of Reflection are answered in the context of my defense of something like Reflection. Consequently, I shall simply use ''Reflection'' through the rest of this essay to refer to either or both the official version and my principle which is something like it. It will be clear from context how it is to be taken.

A variety of arguments as to why one should satisfy Reflection have been offered but these arguments have been vigorously contested.[6] For a number of reasons, however, some of the most interesting philosophical issues connected to a defense of Reflection have remained unexplored. In this paper I want to give two arguments in favor of (a qualified version of) Reflection which are quite distinct from the kind of argument previously given for the principle. These arguments, while distinct, are connected to each other. At the broadest level they are connected because both of them derive from basic features of what it is to be a person.[7] Issues of personhood and personal identity have cropped up occasionally in discussions of Reflection but their true significance has not hitherto been properly or fully understood.

The most prominent argument advanced for Reflection is the Dutch Strategy argument (also known as the diachronic Dutch Book argument).[8] It shows that someone who violates Reflection is open to a series of bets, offered over time, each of which she will find fair by her odds at the time the bet is offered, but which together guarantee a net loss. This argument, like the better-known (synchronic) Dutch Book argument for probabilistic constraints on degrees of belief, is a coherence-based argument. The guarantee of a net loss is supposed to indicate a certain internal incoherence in the beliefs of someone who fails to satisfy Reflection. What distinguishes it from the synchronic version of the argument is that the coherence it enjoins is of a diachronic variety. This has made the argument seem suspect to some who nonetheless have sympathy with the synchronic version of the argument.[9] By contrast, the arguments I will give here are of quite different characters. The first is what I call a truth-based argument, an argument that shows that satisfaction of Reflection is rational because it will promote true beliefs. The second argument is a transcendental argument. It shows that satisfaction of Reflection is a necessary condition of making plans, which is itself an essential part of being a person.

---

[6] The following is a sampling of the literature. For the principle are Van Fraassen (1984); Sobel (1987); Green and Hitchcock (1994); and Skyrms (1997). Against the principle are Christensen (1991); Talbott (1991); and Foley (1994). More equivocal positions are taken by Maher (1993), chapter 5; and Vickers (2000).

[7] By a person I mean a rational, finite agent of the kind of which normal adult human beings are the best and perhaps the only example with which we are acquainted.

[8] The argument is given in Van Fraassen (1984) and is adapted from one given by David Lewis in favor of conditionalization (which, as Van Fraassen (1995) shows, implies a slightly more generalized version of Reflection). Lewis's presentation of the argument was not published until (1999), but it had entered the literature through a presentation of it in Teller (1973).

[9] See especially Christensen (1991) and (2004), chapter 5.

2 The first argument for Reflection: Ameliorism and Self-Knowledge

My first, truth-based argument for Reflection depends on two substantive premises about fundamental features of what I call our epistemic situation and our knowledge of those features. The first premise is that, in general, people's beliefs get better over time. Call this principle Ameliorism. The second is that it is an essential feature of being a person that one have some knowledge of Ameliorism. Call this Self-Knowledge. (Strictly speaking, the argument requires only justified belief in Ameliorism. It does not require that that belief be true. Hence it does not require Ameliorism as a premise and Self-Knowledge might be more accurately called Self-Belief. Nonetheless, since I think Ameliorism is true, or at least some qualified version of it is, the argument is more convincing for relying on it, since its truth helps make Self-Knowledge seem more plausible.) I take up the elaboration and defense of these principles in turn.

Time, I maintain, is generally the friend of belief. This manifests itself in several ways. First of all, we are located, limited creatures, whose conceptions of the world have to be put together bit by bit, largely on the basis of experience. The more time we have, the more experience we have, the more complete our picture of the world becomes. Not only does increased experience enlarge the scope of our beliefs. It provides materials for their refinement and correction. The more experience we have, the less likely it is that a false belief will endure unchallenged. Time has other salutary, if somewhat vaguer, effects on our beliefs. We gain in wisdom and understanding. Whatever exactly these are, they are likely to improve our beliefs. There are also problems that take time to think through, or opinions on which we can benefit from a second thought. The picture I am painting is hardly unfamiliar. Societies have typically treasured their elders as experts. They have seen more of the world, experienced its vicissitudes, and had time to see the effects of their own and others' actions, and to ponder the meaning of it all.

It is, admittedly, hard to make Ameliorism precise. I have talked vaguely about improving one's beliefs. This includes at least two dimensions: increasing the ratio of true to false belief, and increasing the number of true beliefs. But the question of whether and how beliefs can be quantified, thus enabling precise measurement in both these dimensions, is itself controversial. Nor are all beliefs equal in their value. Some account ought to be given of relevance, or importance, of beliefs acquired lest age be seen, ultimately, as an opportunity to stockpile trivialities that are bound, one way or another, to get in the way of more important things. Nevertheless, experience, wisdom and understanding, universally admitted as the prerogatives of age over youth, must themselves bring discernment as to what is important. And the sheer accumulation of experience is an ungainsayable and impressive (though perhaps easily overlooked) fact of life.

The Ameliorist idyll, of course, is only part of the story. Time has other, less salutary effects on belief that must also be taken account of. Ultimately, these will lead us to a qualified version of Ameliorism. But for the present, let us pretend that an unqualified Ameliorism characterizes the shape of our epistemic lives. As it happens, literature provides us with a case with which further to explore Ameliorism. The wizard Merlin, in T.H.White's *The Once and Future King*, lives his life backwards,

getting younger with each passing day.[10] How does Ameliorism stand with respect to his situation? In getting younger, one supposes, he loses, along with his wrinkles, wisdom and understanding and, at a certain point, even cognitive skills that, for the rest of us, come with growing up. To the extent that the overall quality of one's opinion depends on cognitive skills, wisdom and understanding, his later (younger) opinion is not superior to his earlier (older) opinion. In this sense, Ameliorism is false for Merlin. However, the case is not as simple as that. For we have to deal with the issue of those beliefs about the world that come simply with the increased opportunity to experience it that time affords. At his origin, did Merlin come into the world with a stock of beliefs about all sorts of empirical things (the address of that witch who could supply liver of blaspheming Jew, etc.) and does he then gradually lose these beliefs throughout the course of his life? If so, the story borders on incoherence. For all his wizardly qualities, Merlin has the form of a human being. He has eyes and ears with which he goes through time, seeing and hearing more and more as he gets younger. Why then, should he not be gaining empirical beliefs? Does the world make no impact on his senses as he goes through it? If, on the other hand, he begins his existence with little or no empirical information, and acquires it in due course as he progresses through the world, then despite his getting younger all the time, he will still be acquiring more and more beliefs through ordinary empirical means. To this extent, Ameliorism would still apply. Only when he became too young to have the beliefs (i.e. when his cognitive skills became too immature to support the beliefs), would his later beliefs on empirical matters cease to be superior to his earlier beliefs. He would fade into the imbecility of a first childhood.

What this fantasy reveals (if it is coherent at all) is that time actually plays two distinct roles in improving our beliefs. The first role is that our cognitive skills, our neural hardware, etc., develop rather than senesce over time. This is clearly a contingent fact about human persons. Other kinds of persons could be, as it were, born at their cognitive peak and thereafter suffer a steady cognitive decline. Instead, we grow into our cognitive peak just as we grow into our physical peak. Cognitive decline may occur later in life (I will return to this issue below), but it is worth remembering that what faces us all with such high probability now is a recent feature of human life. Until only a short time ago, life expectancy would have most of us dying long before our cognitive equipment could go into decline. Let us call this aspect of the improvement of our beliefs over time Contingent Ameliorism. That is to be contrasted with the much more fundamental way in which beliefs can improve owing merely to our being in time. The passage of time affords an increase of experience. Experience provides the basis for additional beliefs and for the correction and systematization of existing beliefs. To contrast this sense of Ameliorism with what I have just called Contingent Ameliorism, I will use the name Necessary Ameliorism. But it is important to take this in the right way. My claim is not that it is

---

[10] White's conceit is actually more complicated than this and clearly incoherent for reasons unrelated to those I discuss in the text. Merlin's living backwards is supposed to explain his knowledge of the future. For that to work, he must not be moving forwards in time but getting younger, as I pretend in the text, but getting older, like everyone else, while going backwards in time. But that, in turn, would be inconsistent with how he is represented by White: the order of the phonemes in the English sentences he speaks, for example, is the same as for everyone else.

a necessary truth that people's beliefs improve owing to the way in which the passage of time enables the increase of experience. There are clearly many ways in which this can fail to happen. I will discuss some of the more significant ways in due course. Here we may just mention that one might simply be highly unlucky in having lots of misleading experience. A variant of this would include cases in which a malevolent demon, for example, destroyed the world but continued to produce in us experiences as if things were going on just the same. The necessity in Necessary Ameliorism lies not in amelioration itself, but rather in the way in which amelioration is tied, under normal conditions, to very deep features of belief, experience and time. Experience occurs over time and provides an important avenue for continuing acquisition and refinement of beliefs. These links are of a conceptual, or quasi-conceptual nature, even if it is contingent whether conditions are indeed normal in any given case.

How much, exactly, of our cognitive improvement is owing to Contingent Ameliorism? I spoke above of cognitive skills and neural hardware. A baby is born unable to see and requires some short time to develop sufficiently to process information in this way (though this all happens before beliefs begin to be acquired at all). Other forms of development of hardware are more subtle. A normal six-year old, while able to have beliefs, presumably cannot have them about quantum mechanics. Perhaps before puberty, one cannot have beliefs involving concepts of sexual desire or romantic love. (In this case, it is not the *neural* hardware the needs developing.) In describing Merlin's case, however, I also suggested that he would lose wisdom and understanding. I have no analysis or empirical theory of these phenomena. While they may depend on developments in neural or other hardware, they may also (or only) depend on experience and other features associated with Necessary Ameliorism. If they depend on factors of both these kinds, then the story of Merlin will not be coherently describable in terms of these concepts. If they depend only on features associated with Necessary Ameliorism, then even Merlin will gain in wisdom and understanding as he gets younger.

The second premise of the argument is Self-Knowledge. The idea that being a person brings with it some requirements on self-knowledge is common. Sometimes, for example, a difference between consciousness and self-consciousness is used to mark the difference between (mere) animal and personal existence. Coming closer to our point, Donald Davidson has argued that to have beliefs at all (surely a necessary condition of being a person) entails that one have the concept of belief and hence can attribute beliefs to oneself and others; that one have, in other words, beliefs about beliefs.[11] Assuming this much to be true (I shall be happy to stipulate that the arguments in this paper apply only to such beings for whom it is true), one can, I think, plausibly go further. Recognition of Necessary Ameliorism will be inevitable. The argument for this is based on the conditions of belief-attribution. If one is to attribute beliefs to people in general, there must be guidelines, supplied by such things as knowledge of their environments, their experiences, and so on, that play a role in such attributions. The whole point of attributing beliefs is that it helps explain people's actions by connecting those actions to certain ways in which the

---

[11] This idea occurs throughout his work. An extensive discussion is in (1984).

world impinges on the agents. The knowledge necessary for belief attribution to work must include at least the following: that there is a default condition that absent special circumstances, if a belief is attributed at one time, it is available at a later time for use in explanation and interpretation; that new experience leads to new belief; that experience is capable of leading to rational revision of belief; and so on. These framework conditions of belief (as we might call them) imply Necessary Ameliorism. Thus, being a person (and so an attributer of belief) entails having knowledge of Necessary Ameliorism.

Although ordinary human experience provides us with knowledge of Contingent Ameliorism fairly easily I do not think one can argue merely from framework conditions of belief to knowledge of it, as one can with Necessary Ameliorism, even supposing we restrict ourselves to beings, such as humans, for whom Contingent Ameliorism is true. One could engage in belief attribution, with knowledge of the framework conditions that imply Necessary Ameliorism, without having knowledge of what effects differences in people's ages had on what kinds of beliefs they could have. Below, we shall return to issues of belief attribution to show that such a practice requires knowledge of more than Necessary Ameliorism. Those considerations will show, I think, that it would be very strange (though not impossible) for creatures for whom Contingent Ameliorism was true not to have knowledge of it. But for the moment, we may note that as long as one had no good reason to believe in the falsity of Contingent Ameliorism (to believe in Contingent Senescence, for example) the argument for Reflection about to be given would go through with knowledge of Necessary Ameliorism alone.

Given Self-Knowledge and Ameliorism, the rationality of satisfying Reflection is obvious. If the beliefs of one's later selves are, by and large, superior to those of one's current self, and if one has good reason to believe this, then it is rational to treat those later selves as experts in the sense defined above. This defense of Reflection sees the principle in general terms. It is rational to satisfy the principle because of our general expectations of what our future beliefs will be like. This attitude towards our future beliefs will still be rational even if we have good reason to believe, as we do, that in *some* cases, future beliefs will be wrong or otherwise inferior to current beliefs. This observation helps to disarm the most common kind of counter-examples made to Reflection, cases in which one has reason to believe that some particular future belief will be wrong by current standards. In order to provide such cases, resort is almost always made to such things as mind-altering drugs, hypnosis or other 'external' disrupters of rational belief formation. To give one example, suppose you believe that you have just taken a drug that will make you believe that you can fly. You will naturally not want to adopt the belief now that you can fly—you will not, that is, treat your future self as an expert in this case.[12] But as I have just been emphasizing, this is perfectly compatible with treating your future self as an expert in general.[13]

---

[12] The example is from Christensen (1991).

[13] It is also, it should be pointed out, consistent with treating oneself as an expert in this particular case. The definition of treating someone as an expert, recall, required us to adopt her belief only if we did not, in learning of that belief, also learn some reason why we should not adopt it. This condition is clearly activated in the case described.

Can this kind of counter-example be extended to undermine the argument for Reflection I have given? Could I not suppose that I have taken a drug (Universal Brain Scrambler—UBS) that will not simply provide me with one belief that I now consider false, but that will, by current standards, massively and comprehensively mess up my beliefs? Here is where considerations of personal identity enter the picture. Van Fraassen supplemented the Dutch Strategy defense of Reflection with a line of argument that when I fail to satisfy Reflection with respect to some future belief, either the subject of that belief is not me or the belief is not really mine.[14] Concerning this kind of claim, in the context of a case in which one foresees getting drunk and believing, as a result, that one can drive safely, one critic of Reflection, Maher (1993, 107–108), writes:

> this is a desperate move. Nobody I know gives any real credence to the claim that having ten drinks, and as a result thinking he or she can drive safely, would destroy his or her personal identity.

Maher is quite right about the kind of case he describes; but as we have seen, such a response as he envisages the defender of Reflection making is not actually necessary. However, when one considers the UBS case, the response is, it seems to me, not a desperate move at all. In such circumstances, it would be quite natural to say that one would become a different person.[15] What lies behind such judgments is the idea that there are epistemic constraints on personal identity over time such that, at least in some cases, facts about beliefs and their formation can be sufficient for the non-identity of a person at one time and a person at another. The most likely form for such constraints is in terms of methods of belief acquisition. For example, one might hold that a necessary condition of a person Q's (at time t') being the same person as P at some earlier t is that Q's beliefs at t' should have been acquired from P's beliefs at t in ways that P at t finds reasonable. In the UBS case, such constraints will obviously be flouted on a grand scale.[16]

---

[14] This theme is sounded in Van Fraassen (1995) and also in Jeffrey (1992), especially 126–132.

[15] Like Rovane (1998), 59–64, I think that our ordinary conception of what a person is is probably too confused to avoid conflicts of intuition over analytical claims such as the one I go on to make in this paragraph. But I cannot adequately defend here either the claim in the text or the meta-claim in this footnote.

[16] Above, I mentioned as a way in which amelioration of beliefs might fail to obtain the possibility that a malevolent demon might destroy the world but continue to produce in us the kinds of experience we would expect to have anyway. If one had reason to believe such a hypothesis might obtain in the future, it might be thought that this would provide a counter-example to the epistemic constraint on personal identity suggested here since there is, unlike in the UBS case, little intuitive support for the judgment that personal identity would not be preserved. (This worry was urged on me by Elia Zardini.) I am not sure what the best response is to this objection but I offer the following somewhat tentatively. If the hypothesis is that the demon puts the beliefs straight into one's head, so to speak, then there is, if not transformation into a different person, at least some kind of loss of self since there is no autonomy in the belief-forming process. If, on the other hand, the hypothesis is that the environment is so manipulated that our normal interactions with it become radically misleading, then it is not the case that our beliefs are being formed in ways we would not consider reasonable. The fault would lie elsewhere than with the method of belief acquisition. No doubt there are further variants of the hypothesis that may raise further problems.

The way in which I formulated the epistemic constraint on personal identity raises the question of whether we can specify what is meant by ''reasonable ways'' in which beliefs can be acquired, and whether we can do so independently of appeals to the rationality of Reflection. Maher (1993), whose response to invocations of personal identity was quoted above, thinks that it might be plausible to say that Reflection is rational when applied to beliefs formed in certain ways, but goes on to suggest that the only way of specifying the relevant ways will turn out to be as ways which preserve the rationality of Reflection, thus rendering such a move useless as a defense of Reflection. In fact, however, I think it is feasible to come at the notion of ''reasonable ways'' of forming beliefs independently of Reflection by elaborating further what is involved in the Self-Knowledge premise of the argument. This will emerge in the next section.

3 Qualifications to Ameliorism and Self-Knowledge

I noted above that Ameliorism will certainly be thought to paint too rosy a picture of our epistemic lives. There are all sorts of ways in which our beliefs do not get better over time. However, although a reasonable person believes this, she does not typically anticipate specific lapses in her future beliefs. Her lapses are unpredictable in type and in content. As we have noted, recognition of such likely future deficiencies is compatible with a general satisfaction of Reflection. The situation here with respect to our future selves is analogous to our relation to any fallible expert.[17] If my friend knows a lot more about the history of 16th century Florence than I do, I will do well to defer to her beliefs about that, even while I recognize that she will sometimes be wrong. Unless there is some predictable pattern to her failings (she always gets things wrong about the military history), if her mistakes are few enough, a general policy of deferring to her opinions will still be rational.

There are some fairly predictable ways in which future beliefs will be inferior to current ones. The two biggest serpents in the garden are senility and forgetfulness. Neither, however, poses a problem for Reflection. Reflection tells one to defer to one's future opinions, whereas forgetfulness induces a lack of opinion. The fact, therefore, that I will undoubtedly forget many things I now know does not make it irrational to defer to those opinions I will have. Only if I have reason to think that my forgetfulness will be of so great an extent as to undermine the reliability of whatever opinions remain will future lack of opinion impugn the rationality of Reflection. Again, the same goes for my friend the Florentine historian. That she does not know some things about 16th century Florence (perhaps even things that I do know) does not, by itself, undermine her expertise about the things about which she does have opinions. Only if her blanks are so great as to undermine her reliability in what she does have opinions about should I cease to take her as an expert. Normal memory loss induced by time and increased by age does not undermine our general reliability. The fact that, as an old man, I will no longer remember many details of my earlier life

---

[17] I explore the analogies between Reflection and taking other people as experts in Evnine (2003) and (2005).

does nothing to undermine my accumulated wisdom with respect to the things I do have opinions about. Hence, knowledge of the usual kind of memory loss that we suffer as we progress through time does not make Reflection irrational. Nor is senility a problem. Even if we had (as we usually do not) reason to believe that that was our destiny, still, the reasons given in the UBS case for violations of conditions on personal identity would tell equally here against identifying one's present self with the senile person one will become.

The kinds of cases that put pressure on Reflection in fact come from a very different source. Roughly put, different methods of belief acquisition are sensitive to temporal considerations in different ways. Out of these differences, counter-examples to Reflection can be constructed. One kind of case is as follows. I now believe, on the basis of perception, that the walls of the room I am in are blue. Suppose I know, or hypothesize, that in a week's time someone will ask me about the color of the walls of the room I was in a week previously and I will form a belief as to their color based on memory. My predicted or hypothesized memory-based belief will not be superior, in reliability, to my current perception-based one, for two reasons. First, a memory is a doxastic presentation of a fact that occurs at a temporal distance from the point at which that fact was initially registered. The greater the temporal distance, the more opportunity for degradation of information preserved. But secondly, the initial registration of the fact was the perception-based belief itself; hence, it cannot be less reliable than the memory-based belief which derives from it (and in fact it must be more reliable unless memory is not thought to add any unreliability itself). Thus, even if I now knew what my belief would be one week from now, it would be irrelevant to what I should now believe about the color of the walls. Another example:[18] suppose I am now eating spaghetti for dinner. I believe this on the basis of perception and perhaps other sources (such as knowledge of my own actions and intentions). If asked whether I ate spaghetti for dinner today two years from now, having no memory at all, I might estimate its probability on the basis of, say, my memory that at that time in my life, I was not eating a lot of pasta. Once again, this future belief, even if I knew now what it would be, would be irrelevant when set against my current belief. One could put the message of these examples crudely by saying that, other things being equal, perception trumps memory or retrospective statistical estimates. Since perception is always of things present, whereas memory and retrospective estimates are always of things past, here is a way in which present selves are more expert than future selves.

Perception is one kind of present authority. Another is derived from the privileged access we have to our current mental states. It has often been alleged that such access yields infallible belief. If so, beliefs based on privileged access are clearly more expert than any beliefs we can have in the future about the same things since beliefs about past mental states are never plausibly thought to be infallible. But even if privileged access does not lead to infallible belief, it is hard to deny that it leads to belief with greater authority than any other methods of forming beliefs about past mental states.

---

[18] Adapted from Talbott (1991). Talbott was, as far as I know, the first to point the way to these kinds of issues.

The point about such examples as these is not that they present anomalous exceptions to Ameliorism and hence idiosyncratic counter-examples to Reflection. Rather, in combination with Ameliorism they provide a more nuanced conception of our epistemic situation. Ameliorism is one aspect of that situation. The epistemic lives of persons have a genuinely progressive, accumulative character (both necessarily and, in the case at least of human beings, contingently, as discussed above). At the same time, however, because our beliefs come to us through particular methods, and because those methods have particular strengths and weaknesses with respect to time, there will inevitably be certain predictable disturbances to the Ameliorist picture. This too is an important aspect of our epistemic situation. Taken all together, we get a refined version of Ameliorism, namely, that future belief is superior to current belief, with those exceptions that arise from the ways in which different methods of belief acquisition are sensitive to time.

The disturbances to Ameliorism stemming from the particular ways in which methods of belief acquisition are sensitive to time straddle the divide between Necessary and Contingent Ameliorism. Unlike the framework ways in which time favors later belief over earlier belief, the features of belief formation that are relevant to refining this view are somewhat contingent in nature. It is, that is to say, contingent precisely which methods of belief formation we have and what their strengths and weaknesses are. Yet there are certain framework-like conditions in place. Persons, rational but limited agents, must have ways of acquiring empirical information, and therefore must have faculties of perception; and they must have ways of retaining at least some of that information, therefore they must have some form of memory. It is in the nature of such faculties that perception is more reliable than memory, for the reasons suggested above. Other methods of belief acquisition, such as introspection and inference, will have their characteristic temporal profiles as well. I do not pretend to have accomplished, with these meager remarks, an account of the necessary features of methods of belief-formation for persons. I merely seek to indicate that there will be general features of those methods that are bound up with the very nature of the types of faculties at issue. Which types of belief-forming faculties persons have, and what their temporal profiles are, is therefore not a completely contingent matter by any means.

As I noted at the beginning of Sect. 2, the truth-based argument for Reflection depends primarily on a belief in Ameliorism (the necessary existence of which in persons was claimed by Self-Knowledge). The chief reason for defending Ameliorism was that its truth helped to make more plausible the relevant version of Self-Knowledge. How does Self-Knowledge stand in the light of our refined Ameliorism? I think one can argue that rational beings like us must have knowledge not simply of the general truth of Ameliorism but also of those ways in which we have seen Ameliorism should be complicated. The argument for this is an extension of the argument for Self-Knowledge given above based on the practice of attributing beliefs. Beliefs arise, and must be seen to arise by creatures with them, as part of our interactions with the world. They cannot, therefore, be attributed where we have no conception of the modes by which they are affected and acquired. Knowledge of these modes, needless to say, will not take the form of a scientific psychology; it is

knowledge that must be available to all creatures, however undeveloped, that have and attribute beliefs. This knowledge will include, I conjecture, (a) a basic classification of the avenues by which beliefs can be acquired into categories like perception, memory and inference; and (b) some conception of the conditions of operation of these different methods. Such knowledge amounts to knowledge of precisely the ways in which Ameliorism must be refined. In the absence of such knowledge, it is hard to see how the attribution of beliefs would be constrained in such a way as to make beliefs plausible elements of rational explanations of behavior. Take perception. The folk-psychological notion of perception as a means of acquiring beliefs will recognize the existence of appropriate sensory organs and contain the knowledge of the obvious and basic ways that such organs can be impeded in their operations (a person in the dark or with eyes closed cannot form beliefs on the basis of visual perception). It must also recognize the ways in which memory-based beliefs can derive from perception-based ones.

It is not necessary that the recognition of the ways in which methods of belief-formation are sensitive to time be codifiable in some set of principles. This kind of knowledge is not meant to be able to settle all difficult questions about the comparative worth of the epistemic warrants of different beliefs. Rather, such knowledge is required as a background against which the attribution of beliefs acquired by the different methods is anchored to the circumstances in which these beliefs are attributed and used to explain the actions of their owners. To serve this purpose, the conception of our epistemic situation, the existence of which is claimed by Self-Knowledge, can be somewhat vague. This, indeed, is why I could not above state precisely how Ameliorism ought to be refined to take account of these issues.

The refined versions of Ameliorism and Self-Knowledge serve as premises for a refined version of Reflection that manifests this same vagueness. Rather than the general

One should treat one's future selves as experts,

we have

One should treat one's future selves as experts, subject to those restrictions flowing from the temporal idiosyncrasies of various methods of belief acquisition.

Lest it be thought that Reflection will now have become desperately ad hoc and hence drained of any interest, note that similar, and similarly vague, restrictions will apply to taking anyone as an expert about anything, and not just to taking one's future selves as general experts. An expert astronomer who tells me, without leaving her darkened study, on the basis of complex calculations, that there is currently an eclipse of the sun will not be heeded in the face of my being able to look up at the daylight sky. Her calculation is superseded by my perception precisely because, with respect to such things as whether the sun is shining, perception is a more reliable method of belief formation than is calculation. The doctor who tells me that since my test is negative, there is nothing wrong with me, and so I cannot be in pain, should be laughed at. My knowledge of my own pain, afforded by a method that has

privileged access to its objects, is superior to any other method.[19] Yet one cannot say precisely under what circumstances an expert's opinion is trumped by one's own. Thus, although Reflection turns out to be a messier principle than it first appeared, it would be unrealistic to expect anything else. One cannot adequately deal with issues of expertise in epistemology without making some reference to methods of belief acquisition. Experts are experts about their fields of expertise in virtue of being able to form beliefs about it on the basis of appropriate methods. They may form beliefs about the field on the basis of other, less reliable methods (the astronomer falls to tea-leaf reading to predict an eclipse) and in that case, their beliefs will not carry any special authority. Or, there may be different ways of forming beliefs about the same subject-matter that are available to others, which beliefs will then supersede the experts' (I judge about my headache by a method with privileged access, which the pain specialist cannot do). Thus, no principles about the way in which expert opinion should influence us can afford to ignore issues of the methods of belief acquisition. And no such principles can be qualified with complete precision. It follows that Reflection, itself such a principle, must also take account of these issues and will also itself be somewhat vague.

One reason why these issues are ignored in discussions of Reflection is that, as noted above, Reflection originated within the context of formal epistemology. The formal apparatus deployed there (usually probability theory) makes no space for method of belief acquisition even to be represented.[20] The formalism is designed to express the most minimal constraints on rational belief, constraints that obtain at a much higher level of abstraction than those imposed by the considerations adduced so far. Thus, the original version of Reflection can be accurately represented in the formalism as

$$P_{a,t}(Q|P_{a,t'}(Q) = r) = r.$$

Our qualified but more realistic version cannot be represented in the formalism at all.[21] The formal apparatus is simply not designed, or intended, to deal with these kinds of questions. Consequently, defenses of the original version of Reflection have

---

[19] As noted, I do not insist that methods with privileged access must be infallible, nor that it is impossible for me to be wrong and the doctor to be right in such cases. A similar caveat applies to the astronomer case.

[20] Skyrms (1983) suggests another dimension along which formal, probabilist epistemology might want, and be able, to incorporate more information that is external to one's current degrees of belief.

[21] The only way to represent facts about how a hypothesized future belief is acquired in the probabilistic formalism is to specify them explicitly as part of the hypothesis. Reflection would thus look like this: $P_{a,t}(Q|P_{a,t'}(Q)=r$ and S$)=r$, where S can specify a method on the basis of which a assigns a probability of r to Q at t'. In fact, the standard kind of counter-examples to Reflection are all of this form since their plausibility depends on hypothesizing, for example, not only that I will believe later tonight that I can drive safely, but that I will believe this as a result of having had ten drinks. (Some cases are not exactly like this, but they involve future beliefs, such as that I am the Messiah, that by current lights I could only believe if I were crazy. Hence the hypothesis implies such an extra claim even if it is not explicit.) However, as Vickers (2000) points out, a demand that one satisfy Reflection cannot be confused with a demand that one satisfy the above expansion in full generality. If S states simply that Q is false, then it would obviously be irrational to satisfy the expanded schema even in cases in which it would not be irrational to satisfy the original one. Thus, to allow information to enter about method of belief acquisition, one would have to express Reflection by the expanded schema, restricted to cases in which S says something only about method of belief acquisition. This, I believe, is pretty much equivalent to where I have ended up in the text.

all had to rely on abstract considerations of coherence. My first argument, by contrast, is a truth-based defense that draws on substantive epistemological principles to show why we should accord our future opinion a special status: we have reason to believe that it is more likely (subject to the restrictions noted) to be accurate than our present opinion.

4 The second argument for Reflection: agency

The argument offered and refined in the preceding two sections rested on considerations of personal identity and what is necessary for being a person at several points. In this section, I want to look at a different, though not unrelated, argument for Reflection that is also based on these considerations.

One essential feature of being a person is the ability to engage in temporally-structured, goal-directed behavior; persons are essentially, in Michael Bratman's phrase, planning beings.[22] Minimally, plans are likely to have the following structure. A goal is set at one time for which action at one or more subsequent times is taken by the agent to be necessary. This is what distinguishes plans from intentions that immediately manifest themselves in action. Where, for a goal adopted at one time, action is needed at some later time, one effectively delegates responsibility for bringing about the goal, by performing the relevant action, to one's later self. Given that action is guided by the beliefs of the agent, it is very important for us to ensure that the people, including our future selves, to whom we delegate actions that are necessary for our current goals, have belief systems that we judge are up to the job of doing what is important by our current standards. Of course, *in extremis*, we may delegate responsibilities to people about whom we are very doubtful. (I will return to this point later.) But one cannot routinely engage in meaningful planning without reasonable confidence in the abilities (including the cognitive abilities) of those to whom actions necessary to the plan are delegated. Thus, we can say, in general, that insofar as we are planning beings, we must have some confidence in the beliefs of our future selves. We can, however, be a little more specific about how this works.

There are many reasons why we may delegate actions necessary to accomplish present goals to future selves. We may not currently have time to perform them; circumstances may not be appropriate; we may simply be lazy. But one reason for such delegating that is endemic to persons is that we currently lack pertinent information. The reason I don't book my train ticket now is because I haven't yet consulted the time-table. I don't yet have the relevant beliefs to guide my action. I delegate the purchase of the ticket to a future self who will have such beliefs. That is to say, I intend to look at the time-table and then book the ticket. Thus, plans often come into being precisely because we judge that our future selves will be superior, with respect to the relevant beliefs, to our current selves.

I say that this reason for planning is endemic to persons because, by definition, persons are epistemically limited creatures. We always operate under conditions of partial knowledge. Under favorable circumstances, this partial knowledge may be

---

[22] I draw extensively on Bratman (1987) in the following.

sufficient not to require us to defer action to accomplish present goals because we do not yet have a pertinent piece of information. But even where deferred action is called for for some other reason (perhaps the circumstances are not yet in place for the necessary action), I know now that I will later need beliefs that I do not yet have, beliefs about a whole host of concrete details of the situation in which I will be acting. I must also have confidence that I will still have many beliefs that I now have that are relevant. When I do come to book my train ticket, I must be able to form beliefs about where the phone is, what the number is, etc.; nor can I have forgotten what I now know about when I must arrive by, or who may be available to meet me. There are no doubt many present beliefs which it will not matter whether I still have when I come to book my ticket. But, along with my necessary confidence that I will have acquired the pertinent new information and that I will be able to take account of the circumstances in which I will have to act, I must also have confidence in the general integrity of my future beliefs about a large but indefinite range of things, including such things as what a train is and that people cannot be in two places at once.

These considerations amount to a limited, transcendental defense of Reflection. By calling it a transcendental defense, I mean that it shows that satisfaction of Reflection is a necessary condition for doing something, engaging in planning, in temporally-structured goal-directed behavior, that is itself a necessary condition of being a person.[23] But I call the defense limited for two reasons. First, as I noted above and will explore further below, we can sometimes plan with little confidence in those (including future selves) to whom we delegate actions. Secondly, the argument as given does not yet show that at any given time, I must satisfy Reflection with respect to my future selves in general, but only that I must satisfy it with respect to those of my future selves to whom my current plans delegate some responsibility.

The second limitation is not, I think, very serious. A person's life cannot very well consist of a number of plans formed and executed serially, with no overlapping. Such an unstructured arrangement of plans would not be adequate to most of the characteristic features of personal life. And in fact, as Bratman emphasizes, plans are typically arranged in complex patterns of hierarchy and interlocking. Some have thought that a person's life must involve a single plan or goal by which all others are hierarchically subsumed.[24] If this were true, then the second limitation on the transcendental argument for Reflection would be decisively removed. But even if it is not true, leading the life of a normal person means that one must always be involved, as it were, in a continual warp and woof of plans at different levels of generality. This means (assuming an answer to our first limitation) that one must have a general confidence in one's future selves as epistemic agents. One must suppose that they will, generally, preserve a vast background of beliefs that one already has, that they will have the means of adequately determining, at future times

---

[23] Strictly, the transcendental defense requires also the premise that we are persons in the relevant sense. I take this to be obvious.

[24] Aristotle, *Nicomachean Ethics* I,i and Sartre (1969), 433–556 may be of this view.

of action, the situations in which they will be acting, and that they will acquire those particular bits of information that are now lacking but that are necessary for the execution of many actions if those actions are to serve one's current goals.

I now come to the first, more interesting limitation. We do sometimes plan even in circumstances in which the people to whom we delegate important parts of the plan are not thought to be reliable. Consider first the case in which I must act through others. To some extent, others are always unreliable in this sense: my intention that they do something is not, in general, sufficient to bring it about that they do it. With others, one must always take certain measures to get them to do what one intends them to do. At a minimum, one must inform them of one's intention. I can intend for A to make the salad for tonight's dinner, but unless I take the step of informing her of my intention, there is no reason to think she will. Even when informed, others can be more or less reliable, and further engineering may be required to get them to act as one intends. If I know A is forgetful, I may phone her in the afternoon to remind her about making the salad. Or I may tell not only A but also B, with whom she lives, that I intend for A to make the salad. People may be unreliable not only for epistemic reasons (they don't know of our intentions or they forget them) but also for non-epistemic reasons: they are disinclined to be helpful, they don't care, they are easily distracted, they overcommit themselves, and so on. In all these cases, we may take appropriate measures to forestall anticipated failures. The general point here is that where our plans require the actions of others, we must always, to a greater or lesser extent, take measures to ensure fulfillment of our intentions. This is to be contrasted with the case of our own actions. Under favorable, but not unusual, circumstances, no measures need to be taken to bring about action in line with our intentions. Merely to intend to do something is sufficient for us to do it (or at least, to try to do it).[25] This distinction should feature in an account of personal identity. It is an important fact about someone's being me that I can control his behavior through my intentions in this special way. Conversely, it is an important fact about someone's being distinct from me that my attempts to control his behavior through my intentions must always rely on what I have called taking measures. My intentions cannot control the behavior of others in the same way that, under normal circumstances, they control my own behavior.

Now it often happens, in a host of small ways, that I predict that my current intention to do something at a later time will not, by itself, bring about my doing that thing. Every time I write an appointment in a diary we are dealing with just such a case. I predict that I will very likely forget, when the relevant time comes, that the meeting is scheduled for just that time. I therefore act now to lay down a clue that will remind me, at a relevant point in the future, that the meeting is at hand. In a small way, the relation between my current intention and the future action of mine towards which it is directed becomes more like the relation between my intention and the action of someone else towards which it is directed. I must, as it were,

---

[25] This way of putting things may be too crude. Perhaps one ought to say something like this: if A intends to x, then that provides a *prima facie* reason for thinking A will x. In any case, it is not necessary in this context to characterize exactly the way in which intentions influence our future actions.

inform myself (i.e. remind myself) of my intention just as I must inform A of my intention that she make the salad for dinner. I must, in short, take measures to see to the fulfillment of my intention.

When the extent to which I cannot control my future actions merely by intending them becomes too great, we have a breakdown in personal identity. In Parfit's (1984, 327–328) example of The Nineteenth Century Russian a young, idealistic nobleman intends to give land he will inherit in middle-age to the peasants. But he predicts that as a middle-age man he will no longer share his youthful ideals and so will not act then in accordance with his current intention. He thus seeks to bind himself by taking such measures as we have been looking at. He makes a contract now that will give away the land then and that can only be revoked with his wife's permission. He then makes his wife promise that she will never give that permission. Parfit, rightly, argues that it is quite acceptable to characterize the case in terms of different selves. The young man feels the older one will be a different person; the wife will argue that the older man cannot release her from the promise made to the younger one precisely because the promise was made to a different person.

This is a case concerning values. Reflection is an epistemic principle.[26] But we can doubt the future fulfillment of our current intentions for epistemic reasons, and not just for the already-noted reason that we may forget we had the intention. Suppose I intend now to sell my house when I retire in order to move to smaller accommodations. I predict, however, that I will be suffering from dementia at that point. The problem that I may forget that I had the intention can be remedied by, say, writing it down in a way that I will find at the relevant time. But my dementia is likely to affect much more than that. There may be cognitive deficiencies concerning how to sell a house, where the title is, what my name is, and so on. The only way now to ensure against such large-scale future epistemic deficiency is analogous to the Nineteenth Century Nobleman's recourse. I must take serious measures, say by now handing over power of attorney to someone who will be cognitively able to execute my current intention. As in Parfit's case, I reveal the otherness of the future self for whom I act precisely by having to act through another person who stands in an appropriate relation both to me and that future self. The argument here thus confirms, though by a different route, the point made in the previous argument that serious breakdowns in Reflection are associated with disruptions of personal identity over time. Not only does personal identity over time require that I see the beliefs of my future self as being acquired in ways I now consider reasonable; it also requires that I be able to exert a certain kind of control over my future actions, a kind of control which does not involve having to externalize my current intentions and knowledge in someone who will carry them forward for action relative to a future person who may not have either the intentions or the knowledge to carry them out.

One could bring out further the relation between the two arguments I have given by saying that planning itself has implicitly, and often explicitly, an epistemic dimension. It has an epistemic dimension implicitly in that we must

---

[26] I briefly discuss a value-based analogue of Reflection in (2003).

work on the assumption that our future selves will be epistemically capable of performing the actions that are delegated to them. Without this assumption, one part of the contrast between the way in which our intentions control our own actions and the way in which they can control the actions of others would be lost. For that distinction relies on (among other things) the way in which, for my intentions to be effective in controlling the actions of others, I must take measures to remedy their expected epistemic deficiencies (something I can do in my own case as well, but not too much without threatening personal identity). The need for epistemic planning becomes explicit when the reason for delegating actions to future selves is precisely to acquire necessary information that I do not now have. In that case, we must have knowledge of our epistemic situation in order to plan how to acquire that information. My plan must take into account what avenues I have for acquiring knowledge and under what circumstances they can operate effectively. If I cannot see, my plan must ensure that the information I need to acquire will be available non-visually. If some crucial information will only be available in Chinese, I must learn the language or provide myself with an interpreter. Thus, in a variety of ways, the kind of epistemic self-knowledge the existence of which is asserted by Self-Knowledge, the self-knowledge that makes Reflection rational, also plays an integral role in planning, an activity for which Reflection is a necessary condition.

5 Conclusion

We have looked at two arguments for Reflection, a truth-based argument and a transcendental one. The truth-based argument shows why it is rational for persons to satisfy Reflection but says nothing about whether persons do satisfy it. The transcendental one shows that persons must satisfy it, but says nothing about whether it is rational to do so. Together, they show that a necessary epistemic feature of personal life is also rational. This is strong support for Reflection.

I noted at the outset that the prevalent argument for Reflection is the Dutch Strategy argument, which is based on considerations of coherence. Critics of this argument have maintained that the kind of diachronic coherence demanded by Reflection is, unlike synchronic coherence, not a desideratum on belief at all. And it is indeed hard to see why the beliefs of a person at one time ought to cohere with her beliefs at another time. A person who believes something at one time and its negation at another is surely exhibiting some kind of diachronic incoherence. But it would be foolish to suppose that one should never change one's mind.[27] Diachronic coherence as such, then, is no virtue in a person's beliefs.

The arguments presented in this paper make no appeal to the value of diachronic coherence in belief. But we are perhaps in a position to see why coherence should have seemed to some as connected to Reflection. I have twice invoked conditions on

---

[27] Indeed, one argument deployed by Christensen (1991), 240–241 against the Dutch Strategy argument for Reflection is that a version of it could be used equally to require that a person never change her beliefs at all.

personal identity in the course of the arguments. I claimed that it was a condition on a person Q's (at t') being the same as a person P (at t) I) that Q's beliefs at t' be reached from P's at t by methods considered reasonable by P at t; and II) that, under normal circumstances, P's intentions at t concerning actions by Q at t' be sufficient to bring about those actions. Both of these conditions suggest that a person's life exhibits a certain coherence over time. This coherence is not the formally characterizable type of logic or probability theory but something vaguer and mushier, and that applies to more than just beliefs. Since condition I) is part of the argument that makes Reflection rational and condition II) is part of the argument that shows Reflection is necessary for planning, it is no wonder that there appears to be a link between Reflection and diachronic coherence. But that link is not so much the basis for the argument for Reflection as a by-product of it.

References

Bratman, M. (1987). *Intention, plans and practical reason*. Cambridge: Harvard University Press.

Christensen, D. (1991). Clever bookies and coherent beliefs. *Philosophical Review, 100*, 229–247.

Christensen, D. (2004). *Putting logic in its place: Formal constraints on rational belief*. Oxford: Clarendon Press.

Davidson, D. (1984). Thought and talk. In D. Davidson (Ed.), *Inquiries into truth and interpretation*. Oxford: Clarendon Press.

Evnine, S. (2003). Epistemic unities. *Erkenntnis, 59*, 365–388.

Evnine, S. (2005). Containing multitudes: reflection, expertise and persons as groups. *Episteme, 2*, 57–64.

Foley, R. (1993). *Working without a net: A study of egocentric epistemology*. New York: Oxford University Press.

Foley, R. (1994). How should future opinion affect current opinion? *Philosophy and Phenomenological Research, 54*, 747–766.

Gaifman, H. (1988). A theory of higher order probabilities. In B. Skyrms & W. Harper (Eds.), *Causation, chance and credence* (Vol. 1, pp. 191–219). Dordrecht: Kluwer.

Green, M., & Hitchcock, C. (1994). Reflections on reflection: Van Fraassen on belief. *Synthese, 98*, 297–324.

Jeffrey, R. (1992). Conditioning, kinematics, and exchangeability. In R. Jeffrey (Ed.), *Probability and the art of judgment*. Cambridge: Cambridge University Press.

Lewis, D. (1999). Why conditionalize? In D. Lewis (Ed.), *Papers in metaphysics and epistemology* (pp. 403–407). Cambridge: Cambridge University Press.

Maher, P. (1993). *Betting on theories*. Cambridge: Cambridge University Press.

Parfit, D. (1984). *Reasons and persons*. Oxford: Clarendon Press.

Rovane, C. (1998). *The bounds of agency: An essay in revisionary metaphysics*. Princeton: Princeton University Press.

Sartre, J.-P. (1969). *Being and nothingness*. trans. Hazel Barnes, London: Methuen.

Skyrms, B. (1983). Three ways to give a probability assignment a memory. In J. Earman (Ed.), *Minnesota studies in the philosophy of science*, (Vol. 10, pp. 157–161). Minneapolis: University of Minnesota Press.

Skyrms, B. (1997). The structure of radical probabilism. *Erkenntnis, 45*, 285–297.

Sobel, J. H. (1987). Self-doubts and Dutch strategies. *Australasian Journal of Philosophy, 65*, 56–81.

Talbott, W. (1991). Two principles of Bayesian epistemology. *Philosophical Studies, 62*, 135–150.

Teller, P. (1973). Conditionalization and observation. *Synthese, 26*, 218–258.

Van Fraassen, B. (1984). Belief and the will. *Journal of Philosophy, 81*, 235–256.

Van Fraassen, B. (1995). Belief and the problem of Ulysses and the sirens. *Philosophical Studies, 77*, 7–37.

Vickers, J. (2000). I believe it, but soon I'll not believe it any more: Scepticism, empiricism, and reflection. *Synthese, 124*, 155–174.