Erschienen in: Philosophy of Science. Cambridge University Press. 2022, 89(2), pp. 398-401
https://doi.org/10.1017/psa.2022.13

Review of Richard Pettigrew's *Choosing for Changing Selves*

Pettigrew, Richard, *Choosing for Changing Selves*. Oxford: Oxford University Press (2020), 272 pp., $72.00 (hardcover; also available as an e-book).

How we evaluate different options tends to change over the course of our lives, and frequently predictably so. At age 20, you might consider yourself lucky to land a job that involves frequent travelling; at age 40, you might strongly prefer to travel for work only occasionally. That our evaluations need not be stable is a problem for all those versions of expected utility theory that tell us to choose rationally by picking an option that we have calculated will maximise our expected utility. Such versions of expected utility theory regard an option's expected utility for an agent as the sum of the agent's suitably quantified evaluations ("utilities") of the option's possible outcomes, with each utility discounted by the probability with which the agent thinks the respective outcome will occur if she chooses the option. (Pettigrew sides with such "realist" versions of expected utility theory, favouring them over "constructivist" versions that impose rationality constraints on an agent's preferences over outcomes, and that regard an agent's expected utility as a mathematical construct that we can derive from the agent's suitably constrained preferences; see 14–15). If evaluations change over time, it is unclear how we should determine the utilities of possible outcomes. Should we base them on the evaluative outlook we hold at the time of making the decision? Or should we take into account, in some way, the evaluative outlook we ascribe to some of our past or future selves?

In *Choosing for Changing Selves*, Richard Pettigrew proposes a carefully argued and insightful answer to these questions. His key move is to assume that if our evaluations evolve, this presents us with a problem of *judgement aggregation* as familiar from social choice theory in the context of individual decision-making. According to the Aggregate Utility Solution that Pettigrew defends, the different evaluations of our past, present, and future selves are aggregated by first constructing "group utilities" over all possible outcomes, where a group utility is simply the weighted sum of all the group members' utilities. The group utilities are then combined with the agent's current credences, or subjective probability assignments, to give the group expected utilities. As Pettigrew clarifies (especially in chapter 5), an exclusive focus on current credences is justified not because other credences are irrelevant, but because defensible current credences will adequately take other credences into account.

In the first and much longer part of the book, Pettigrew develops and defends his Aggregate Utility Solution as a theory of *rational* decision-making for changing selves.

He starts by discussing existing solutions, highlighting their main flaws. Building on insights in formal epistemology and social choice theory, he then lays out the Aggregate Utility Solution (which avoids the identified flaws). In chapters 8 and 10, Pettigrew responds to the worry that we cannot use the Aggregate Utility Solution to actually make decisions, as it presupposes access to information that decision-makers typically lack. Pettigrew considers, in particular, how we might ensure that the utilities of different selves are measured on the same scale, and he discusses L. A. Paul's and Sarah Moss' respective objections that we can never adequately predict how our evaluative outlook would change if we underwent an epistemically transformative experience.

In the second part of the book, Pettigrew considers what weights it might be *reasonable, even if not rationally required* to assign to the utilities of different selves when constructing group utilities. Most importantly, he uses Parfitian ideas about psychological connectedness to argue that we should increase the weights that we assign to the utilities of evaluative outlooks more similar to our current one. He also claims that we should give weight to the diverging evaluations of past versions of ourselves if we are currently morally justly benefiting from their sacrifices.

Choosing for Changing Selves is an exceptionally rich and rewarding book. Seemingly effortlessly, Pettigrew lays out a host of complex ideas in an accessible and engaging manner. Pettigrew's writing is rigorous and careful; formalism is introduced whenever this helps clarify ideas or advance an argument. At the same time, Pettigrew ensures that his ideas are accessible also to readers who wish for a more informal exposition of key issues. The book should thus prove an illuminating read for decision theorists, social choice theorists, formal epistemologists, and ethicists alike.

Despite the book's evident strengths, there are some points I nevertheless take issue with. In the remainder of this review, I will argue that despite Pettigrew's efforts to the contrary, his Aggregate Utility Solution is of limited practical use only. From the outset, Pettigrew is clear that he wants his theory to help us make better decisions in difficult cases. As he puts it: "[. . .] I'm interested in providing a decision theory that we might actually use to deliberate about choices we face, and which makes clear the reasons that motivate and justify the decisions we make as a result" (15; see also xiii and 158–61). Pettigrew deserves much credit for never losing sight of this goal, and it speaks in his favour that he measures the value of his proposed theory partly by asking whether it helps improve actual decision-making. The practical usefulness of Pettigrew's theory is nevertheless limited.

To see what I mean, consider Pettigrew's case of Fernando (5): "[If Fernando chooses a particular pension scheme], they will donate 10% of his pension payments to effective charities once he retires. If he opts in now, there is no way to reverse this decision [. . .]. Considering it now, he would like to do this. [. . .] However, he also knows that, when he retires, his values will have changed and he'll prefer to keep the money for himself to spend on commissioning portraits of him and his pets."

What should Fernando do? Pettigrew argues that Fernando should choose based on his current values, thus binding his future self to give to charity (35). The rational part of Pettigrew's Aggregate Utility Solution cannot explain or justify this verdict. Whether the Aggregate Utility Solution favours opting in or out depends on the weights that Fernando assigns to the utilities of his different selves, and the solution puts no rational constraints on this (other than that the weights should sum to one).

The rational part of Pettigrew's theory thus fails to provide Fernando with practical guidance.

Admittedly, once Fernando reaches a decision in whatever way he does, the fact that his decision can subsequently be recast as maximising his group expected utility in the sense specified by Pettigrew's solution indicates that his decision thereby satisfies desirable rationality considerations (see, in particular, chapter 9). In the rare cases where we arrive at a practical conclusion that cannot be so recast, it will thus be worthwhile to investigate why this is the case (cf. chapter 15). There is, in this sense, a quality assurance role to play for the rational part of the Aggregate Utility Solution. If we turn to Pettigrew's book with the hope of receiving practical guidance, this role is, however, disappointingly auxiliary.

The problem just raised is one that Pettigrew's theory shares with other formal approaches to rational decision-making. It is, moreover, an issue that Pettigrew is well aware of. He motivates the second part of his book by noting that while there are few "absolute constraints on the weights we use to aggregate our past, present, and future selves, [. . .] there are nonetheless more reasonable and less reasonable ways to choose among the possible weights" (159). According to ideas presented in the second part of the book, Fernando reasonably assigns more weight to the utilities of his current self than to those of his retirement self, roughly because Fernando is more psychologically connected to his current self than to his retirement self he shares more memories with the former, his values are more aligned with the former, and so on (215–6). Giving more weight to psychologically closer versions of ourselves helps guard against alienation, and it ensures that we are sufficiently steadfast in our values, not leaving them behind just as soon as some other set of values would be better satisfied (chapters 13 and 14).

This argument in favour of opting in is rather unsatisfactory, however. Consider Fernanda, who faces a decision problem similar to Fernando's, but whose values are reversed. Fernanda would currently prefer to spend 10% of her retirement savings on quirky portraits, but she knows that, upon retiring, she would prefer to give the money to charity. If she has to make a binding choice now, what should she choose? According to Pettigrew's argument, she should spend the money on portraits. At some point in the book, however, Pettigrew seems to suggest that Fernando should give to charity because this is the morally superior choice (35–6). Whatever Pettigrew's considered view, it seems at least reasonable to argue that the values of Fernanda's retirement self are morally superior to her current values, and that the utilities of her retirement self are for this reason appropriately assigned a higher weight.

The crux is now the following. Even if I am right that it is reasonable for Fernanda to opt in, this is something that the Aggregate Utility Solution can easily accommodate. Pettigrew makes clear that he is presenting no complete theory of decision-making for changing selves. The second part of the book simply discusses *some* salient considerations that Pettigrew argues bear on the reasonableness of our choices. We are thus free to extend the Aggregate Utility Solution with further reasonableness considerations including the consideration that, other things equal, we should assign higher weights to morally superior preferences.

In my view, this means that there are important limits to the practical guidance that Pettigrew's book ultimately offers. By discussing how a number of salient

considerations bear on the reasonableness of our choices, the second part of the book doubtlessly offers helpful food for thought for struggling decision-makers. At the same time, it is undeniable that a myriad of (possibly interacting) considerations bear on the reasonableness of difficult choices and that the guidance offered by *Choosing for Changing Selves* will thus generally be limited to providing some potentially useful pointers. The main reason why I nevertheless recommend Pettigrew's book to those looking for practical guidance is simple: I believe that no work in decision theory can ultimately do much better.

Susanne Burri
Department of Philosophy, University of Konstanz, Germany
Email: susanne.burri@uni-konstanz.de