ful, detailed study. It contains valuable insights on a host of topics, including Hegel's understanding of contingency (142–44), immediacy (176), and judgment (186–200). To my mind, however, it offers an overall interpretation of Hegel's *Logic* that is unconvincing, even though it is undoubtedly original and thought-provoking.

References

Hegel, G. W. F. 2007. *Philosophy of Mind*, translated by W. Wallace and A. V. Miller, with revisions and commentary by M. J. Inwood. Oxford: Clarendon Press.

Hegel, G. W. F. 2010. *Science of Logic*, translated and edited by George di Giovanni. Cambridge: Cambridge University Press.

Kant, Immanuel. 1997. *Critique of Pure Reason*, translated and edited by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.

Kant, Immanuel. 2000. *Critique of the Power of Judgment*, edited by Paul Guyer, translated by Paul Guyer and Eric Matthews. Cambridge: Cambridge University Press.

Stephen Houlgate
University of Warwick

Philosophical Review, Vol. 131, No. 2, 2022
DOI 10.1215/00318108-9554743

Richard Pettigrew, *Choosing for Changing Selves*.
Oxford: Oxford University Press, 2019. 245 pp.

Choosing brings change. A major life choice, a "big decision," as Edna Ullmann-Margalit termed it, brings the possibility of changing who you are. Richard Pettigrew's important new book, *Choosing for Changing Selves*, explores theories of decision-making for choices that change us, with a particular focus on life-changing decisions.

Life-changing decisions are decision cases where a persisting agent, through their choice, creates a new self, through replacing their values—that is, replacing that self's utility function. (Take the persisting agent to be constituted by a series of appropriately related selves, and selves to be defined by their values, and by extension by their utility functions.) Such cases are of deep philosophical and practical interest, involving questions about knowledge, evidence, and experiential value, and concerning real world choices such as choosing to have a child, determining one's future medical care, or getting a divorce (Bykvist 2006; Ullmann-Margalit 2006; Paul 2014).

The central focus of the book is on cases where an agent, through her life-changing choice, replaces her utility function. If your values change as the result of your choice, which values govern your action? Your values at the time of your choice (your *ex ante values*)? Or your values at the time of the outcome of your choice (your *ex post values*)?

Spelled out in terms of the metaphysics of selves: at t_1, the agent is realized by her *ex ante self*, whose values are defined with utility function U_1. This self deliberates, and chooses to act in a way that leads to changes in U_1. By changing herself in this way, she realizes her *ex post self* at t_2, a new self with new values defined by utility function U_2. If these value changes go "all the way down," there is no way to choose that is consistent with both U_1 and U_2. Take a person who, at t_1, highly values being child-free, as defined by her utility function U_1, and does not value being a parent. However, if she becomes a parent at t_2, she will highly value that state (using utility function U_2). Traditional decision theory requires that, to choose rationally, she must choose in accordance with her values. In such a case, which self's values should determine her choice?

As Pettigrew demonstrates with clear, elegant prose and mathematical precision, given a small set of reasonable assumptions, making these choices rationally requires one to modify orthodox expected utility theory. Blending formal epistemology with the metaphysics of selves and persisting agents, he proposes a new theory of rational choice for persisting agents.

His solution, in brief, is that one should choose in accordance with a weighted general value function (U_G) that is the amalgamation of the local functions U_1, U_2, ..., U_n of one's n selves—that is, an amalgamation of the value functions of all the past, present, and future selves that compose the life of the agent. Inspired by the work of Derek Parfit and others, he develops and defends an account of U_G as a weighted average of the local utility functions of the different selves that compose the persisting agent over time, where the weights are determined by sacrifices made by local selves, psychological connectedness, and the degree of shared values.

A key move in Pettigrew's argument for U_G is to treat the rationality of choosing for changing selves as a judgment aggregation problem, where the credences and utilities of distinct selves are aggregated. This move is modeled explicitly on the literature for collective decision-making. Chapter 5 nicely develops the way the weak reflection principle applies in cases of credal (or value) change, showing how principles for credences about objective chances depend on the structure of the situation. In particular, in certain types of collective knowledge and peer disagreement contexts (Titelbaum and Kopec 2009), rationality can dictate how a person should incorporate facts about the credences of other persons into their own credences.

Drawing heavily on an analogy between a person's relations to other people and one's current self's relations to one's past and future selves, Petti-

grew then claims that facts about the utilities of one's different selves should be taken into account in contexts of self-change, using collective norms to calculate the local utilities of past, present, and future selves and incorporate them into a general utility function.[1] The result is Pettigrew's solution to the problem of choosing for changing selves: credences and utilities of distinct selves (on analogy to cases of collective decision-making involving distinct persons) are weighted and aggregated to determine a general value function (U_G), and when an agent makes a life-changing decision, they should maximize their utilities in accordance with U_G.

I found this systematic approach enlightening and fascinating. The mathematical treatment of utility functions as applied to the question of self-change, set in the context of contemporary formal epistemology, is especially valuable. The book is packed with interesting ideas about how to treat utilities for persisting agents, and does an especially good job of weaving these insights into extant discussions of credal change. However, I am not convinced by Pettigrew's solution. Space is limited, so I will only discuss two objections.

First, why think that the relations and requirements for how we treat other people ("collective norms") should govern the way we treat our other selves? Pettigrew takes the premise to be obviously true. I am skeptical. In fact, there is excellent evidence that, in important contexts, we treat our own selves more harshly than we treat other people (Crockett et al. 2014).

A related question concerns the weights we assign to selves. Pettigrew defends a "principle of minimal mutilation," placing special emphasis on the sacrifices one's past selves have made. But why should my past selves be treated with such deference? I see no reason now to place *any* weight at all on my five-year-old self's dislike for well-aged tawny port, or for that matter, on my twenty-year-old self's facile romantic attractions, no matter how many sacrifices those selves made for me. I want the freedom to dispense with my past mistakes, to start anew, to throw off the shackles of my former selves.

The second objection runs deeper. Pettigrew's approach requires us to make meaningful intraself utility comparisons. The utility function (U_1) of the ex ante self must be comparable to the utility function (U_2) of the ex post self, or if U_1 and U_2 are incompatible, we must be able to scale them in order for utility changes to be meaningful. (By analogy, if I tell you that it was 25 degrees outside at t_1, but that it is now 35 degrees at t_2, this is meaningful only if (1) you know the temperature scales I am using, and (2) you can convert values in one to values in another. Warming from 25°F to 35°F is not the same as warming from 25°F to 35°C.)

Pettigrew addresses this problem by drawing on extant models of interself utility comparisons, and develops a very beautiful solution for intraself comparisons: the "matching intervals solution." This method allows one to set

1. The order of aggregation is "ex post," as defined on pp. 50–51.

the zero and unit of the scale for different utility functions of one's different selves in order to ensure that the functions are comparable. Assuming (!) for the sake of discussion that all relevant incompatibilities can be resolved by scaling, we can make meaningful comparisons between our current and future (or counterfactual) selves as long as we can compare the right intervals. The solution requires an individual to be able to determine and compare differences between their current utilities for pairs of outcomes (o_1, o_2) to differences between a current utility for o_2 and their future (or counterfactual) utility for o_2.

Unfortunately, the problem of incomparable value functions resurfaces when Pettigrew considers transformative choices. An agent makes a transformative choice when their ex ante self, with utility function U_1, chooses to act in a way that replaces their current utility function, in effect, creating an ex post self at t_2 with a new (incompatible) utility function U_2.

When you choose to transform, your choice will create a new self, with a new utility function. To evaluate the desirability of this choice, you must compare the utilities assigned to the outcomes by your new self with the utilities assigned to the outcomes by your old self. For the comparison to be meaningful, the utilities of these selves must be comparable. (For example, the way you'd value parenthood when you're child-free [using U_1] is different from the way you'd value parenthood as a parent [using U_2]. To meaningfully interpret this difference, U_1 and U_2 must be comparable.)

The intuitive way we try to compare utilities across selves is by prospectively assessing the utilities of our future selves. If you know what it would be like to be some future self, you can imagine yourself this way, empathically assess your utilities (and scale if needed), and meaningfully compare this utility to your current utility.

The trouble is that, in transformative cases, we cannot simply imaginatively evolve ourselves forward in order to empathize with our future selves in this way. We lack the necessary prospective abilities. Pettigrew's solution to this problem is to substitute other-based testimony about utility changes, drawn from social-scientific research on people who are relevantly like you.

However, if we are to be justified in relying on other-based testimony about utility change when we are considering a class of cases involving the possibility of incomparable utility functions across our different selves, we must know that the reported utilities are comparable. Unfortunately, the problem of incomparable value functions infects the interpretation of existing empirical results (Paul and Healy 2018).

In brief: in randomized controlled trials (RCTs), utility changes are measured by comparisons (of average results) between a treatment group and a control group. In the cases of interest, over a fixed period of time, members of the treatment group undergo transformative experiences. Members of the control group do not. All else is held fixed. After the experience, utility

changes of members of the treatment group are measured and compared to utility changes (if any) of members of the control group. On the assumption that the relevant utility comparisons are meaningful, any differences in (average) utility between the two groups can be seen as a measure of the impact of the transformative experience.

But we cannot assume meaningfulness. A version of the problem of intrapersonal incomparability arises for the selves of the treatment group before and after their treatment (before and after the transformative experience). In these studies, the utilities of treated subjects are compared at t_1, before (ex ante) the "treatment" (the transformation) and at t_2, after the treatment (ex post). By definition, transformation involves utility function replacement. If the subjects' ex post utility functions are not comparable to their ex ante utility functions, comparisons of their ex ante utilities with their ex post utilities will not be meaningful.

You might hope that the problem can surmounted, for the important comparison in these studies is at t_2, between the ex post utilities of the treatment group and the ex post utilities of the control group. But the problem remains. For an RCT, the meaningfulness of the ex post comparison between the control group and the treatment group relies on matching the two groups before the treatment is applied, and then assuming this match carries forward throughout the experimental procedure. (The need for such a match stems from empirical constraints concerning the use of counterfactuals, and derives from what social scientists term "the fundamental identity problem.") The change in utility functions destroys the match, destroying inferences relying on comparability.

In brief, *ex post comparability between the utilities of treatment and control groups can fail, and for precisely the same reason as comparability between the utilities of one's current and counterfactual selves can fail.*

There is much more to engage with in Pettigrew's absorbing book than I can treat in this short review. My objections notwithstanding, there is no question that it makes a major contribution to the literature. It is filled with insightful ideas and new connections. It breaks new ground by developing a systematic mathematical theory of the role and relationships between the utility functions of selves (the existing literature focuses almost solely on credence functions), and by applying this mathematical approach to the range of metaphysical and epistemological issues that arise when choosing for changing selves. It is a model of clarity and precision. As such, *Choosing for Changing Selves* is essential reading for anyone interested in decision theory, practical reasoning, agent persistence and change, temporal discounting, and personal identity.

References

Bykvist, Krister. 2006. "Prudence for Changing Selves." *Utilitas* 18, no. 3: 264–83.

Crockett, Molly J., Zeb Kurth-Nelson, Jenifer Z. Siegel, Peter Dayan, and Raymond J. Dolan. 2014. "Harm to Others Outweighs Harm to Self in Moral Decision Making." *PNAS* 111, no. 48: 17320–25.

Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press.

Paul, L. A., and Kieran Healy. 2018. "Transformative Treatments." *Noûs* 52, no. 2: 320–35.

Titelbaum Michael G., and Matthew Kopec. 2019. "When Rational Reasoners Reason Differently." In *Reasoning: New Essays on Theoretical and Practical Thinking*, edited by Magdalena Balcerak Jackson and Brendan Balcerak Jackson, 205–31. Oxford: Oxford University Press.

Ullmann-Margalit, Edna. 2006. "Big Decisions: Opting, Converting, Drifting." *Royal Institute of Philosophy Supplement*, no. 58: 157–72.

L. A. Paul
Yale University

Philosophical Review, Vol. 131, No. 2, 2022
DOI 10.1215/00318108-9554756

Linda Trinkaus Zagzebski, *Epistemic Values: Collected Papers in Epistemology*. Oxford: Oxford University Press, 2020. 364 pp.

Linda Zagzebski has made wide-ranging and influential contributions in epistemology and philosophy of religion over her career. Within epistemology specifically, Zagzebski is rightly associated with her pioneering book *Virtues of the Mind*, which is a core classic in virtue epistemology (Zagzebski 1996). Importantly, though, Zagzebski has also carved out notable positions in a range of other areas of epistemology, including in debates about the nature of understanding, the value of knowledge, religious epistemology, intellectual autonomy and authority, and skepticism and the Gettier problem. *Epistemic Values: Collected Papers in Epistemology* brings together twenty of Zagzebski's epistemology essays, divided into six sections (with usually three or four essays each) that span all of the above general themes.

I will make a general comment about the book as a whole and then use the remainder of this space to critically focus on a limited selection of the specific essays included. The general comment is that while the book contains no new articles—everything here has been previously published (the most recent in 2019)—it nonetheless helps readers of her work to better appreciate her various interventions in epistemology as part of a wider kind of vision. Recurring