University of Groningen

Self-Control, Decision Theory, and Rationality - New Essays

Grayot, James D.

Published in:
Journal of Economic Methodology

DOI:
[10.1080/1350178X.2020.1729531](https://doi.org/10.1080/1350178X.2020.1729531)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Grayot, J. D. (2020). Self-Control, Decision Theory, and Rationality - New Essays. *Journal of Economic Methodology*, 27(2), 184-189. <https://doi.org/10.1080/1350178X.2020.1729531>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Self-Control, Decision Theory, and Rationality – New Essays

edited by J. L. Bermúdez, Cambridge University Press, Cambridge, UK, 2018, xii + 268 pp., £75.00 (hardcover), ISBN: 978-1108329170

James D. Grayot

To cite this article: James D. Grayot (2020) Self-Control, Decision Theory, and Rationality – New Essays, Journal of Economic Methodology, 27:2, 184-189, DOI: [10.1080/1350178X.2020.1729531](https://doi.org/10.1080/1350178X.2020.1729531)

To link to this article: <https://doi.org/10.1080/1350178X.2020.1729531>

Published online: 18 Feb 2020.

Submit your article to this journal [↗](#)

Article views: 81

View related articles [↗](#)

View Crossmark data [↗](#)

- CORE Team. (2017). *The Economy*. Oxford: Oxford University Press. Retrieved electronically at <https://www.core-econ.org/the-economy-download-glossary/?Lang=En>
- Komlos, J. (2015 [2014]). *What every economics student needs to know and doesn't get in the usual principles text*. New York, NY: Routledge. first published M.E. Sharpe.
- Komlos, J. (2019). *Foundations of real-world economics: What every economics student needs to know*. New York, NY: Routledge.
- Mankiw, G. N. (2017[1998]). *Principles of economics* (8th ed.). Boston, MA: Cengage Learning.
- Tieleman, J., de Muijnck, S., Kavelaars, M., & Ostermeijer, F. (2018). *Thinking like an economist: A quantitative analysis of economics bachelor curricula in the Netherlands*, Rethinking Economics Netherlands.

Merve Burnazoglu
Applied Economics, Utrecht University School of Economics.
 Kriekenpitplein 21-22, 3584 EC, Utrecht, Netherlands
m.burnazoglu@uu.nl

Francis Ostermeijer
Department of Spatial Economics, Vrije Universiteit Amsterdam,
De Boelelaan 1105, 1081HV Amsterdam, Netherlands
francis.ostermeijer@vu.nl

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

<https://doi.org/10.1080/1350178X.2020.1729530>

Self-Control, Decision Theory, and Rationality – New Essays, edited by J. L. Bermúdez, Cambridge University Press, Cambridge, UK, 2018, xii + 268 pp., £75.00 (hardcover), ISBN: 978-1108329170

The history of Western philosophy is awash in platitudes championing rationality and self-control. Since Plato, philosophers have cautioned against the iniquities of a weak will and have proffered ever-evolving solutions for circumventing temptation. Today little has changed. Conventional wisdom still holds that practical rationality consists in the judicious exercise of self-control. Yet, the formal study of self-control has progressed from a purely philosophical inquiry to an interdisciplinary endeavor, implicating the cognitive and behavioral sciences as equal stakeholders in the investigation.

Much like the concept of rationality itself, the concept of self-control is a moving target – and the space through which it moves is both normative and descriptive. This is perhaps what makes it such a fascinating if perplexing phenomenon. On the one hand, self-control gives rise to normative paradoxes insofar as it requires one to forego a shorter, sooner reward for a presumably larger, later one. This contradicts classical views in decision theory which identify rational choice as consisting in the maximization of (expected) utility at every node or juncture where choice is possible. On the other hand, self-control gives rise to descriptive paradoxes because it requires persons to make sacrifices for their future selves. The question as to whether different time-slices of a person constitute motivationally disconnected agents depends in equal parts on psychological and meta-physical considerations, neither of which are uncontroversial.

In this collection, J. L. Bermúdez (ed.) draws from a coterie of interdisciplinary experts in philosophy, psychology, and economics to present the newest developments in decision theory about

the rationality of self-control. The book contains eleven original essays, each responding to the prompt: “How is the rationality of resolute choice to be defended within the instrumental perspective of decision theory?” and “How is resolute choice even possible?” The chapters are organized according to three sub-sets of questions which serve as *de facto* themes for the book – they are: (i) *How does decision theory provide a standard of rationality for decision problems of the basic type of the paradigm case?* (ii) *What are the mechanisms responsible for basic preference reversals that give the decision problem its force?* (iii) *How does precommitment happen? – what formal, physiological, and normative implications derive from different self-control technologies?* Navigating these thematic questions is made considerably easier thanks to a remarkably concise introduction by Bermúdez. In fact, I would count the introduction as one of the highlights of the book.

Chapters (1)–(3) go a long way toward clarifying what exactly is normatively paradoxical about self-control, showcasing how decision theorists have been and remain divided over its proper characterization when construed as a dynamic choice. According to McClennen (1990), instrumental rationality requires an agent to be resolute in the face of temptation: to be resolute, an agent justifies their plan of action if it yields Pareto improvements to each time-slice of the agent, leaving some time slices better off but none worse-off. Thoma (Ch. 1) rejects this sort of argument on two fronts. First, she argues that conceiving of dynamic choice as an intertemporal prisoner’s dilemma (as portrayed and defended by Gauthier, 1987, 1994) fails to capture what is compelling about temptation. Whereas agents in a prisoner’s dilemma have constant preferences (their goals are to achieve the best state without ceding advantage to the opponent), agents who face temptation typically have inconstant preferences which lead them to violate previous resolutions. Second, the concept of resolute choice as conceived by McClennen relies on a metaphysically vague and psychologically problematic conception of time-slice agents. If time-slices are interpreted to be autonomous, then it demands that each agent carry out resolutions it did not make; and, if agents are semi-autonomous, say because of the continuity of memory and other conative attitudes, then the notion of time-slices becomes redundant and temptation vanishes.

The baton is then passed to Peterson & Vallentyne (Ch. 2) and Weirich (Ch. 3). Peterson & Vallentyne defend a modified version of *sophisticated choice*, building upon Rabinowicz’s conception of *wise choice* (1995, 1997). If resoluteness entails that it is rational to adopt a plan if and only if full implementation of a plan yields prospects at least as good as any alternative, then only perfectly resolute agents who give lexical primacy to rational preferences to comply with plans could be rational. Sophisticated choice, by contrast, requires an agent to predict their final choices and backward induce the best course of actions, *including the likelihood of akratic behaviors*. It presumes that agents may not be resolute, and so, it is necessary to include lapses of self-control in the rational plan. However, the core of any argument in favor of sophisticated choice (even without the caveat of backward induction) is the assumption that agents form probability distributions over choices while also deliberating about those choices. This leads Peterson & Vallentyne to speculate whether sophisticated choice needs to satisfy the criterion of historical separability. They conclude that it doesn’t. Weirich (Ch. 3) breaks with Thoma and Peterson & Vallentyne, both thematically and rhetorically. Weirich indicates that in order to determine whether cases of rational self-control possibly overturn decision theory’s standard of utility maximization, it must first be decided whether rationality is a mental capacity or an evaluative benchmark. The impetus for this consideration is to question whether the instrumental perspective in decision theory ultimately aims at a descriptive or normative justification for the rationality of self-control. With this in mind, Weirich painstakingly reconstructs a conception of the ideal sophisticated decision agent faced with a paradigm case of temptation. The resulting argument demonstrates that exercising self-control *always* entails the maximization of utility (likewise for expected utility in the case of dynamic choice under risk and uncertainty). In this regard, Weirich provides reason for thinking there is indeed no normative paradox within the classical decision-theoretic perspective: the rationality of individual acts is to be prioritized over the rationality of sequences of acts, full stop.

Chapters (4)–(7) correspond to the second theme of this collection. Here emphasis is placed on describing and qualifying the formal and psychological effects of impulsivity. However, one thing should be noted about these four chapters: The transition away from pure instrumentalism means that one must contend with wider array of jargon and desiderata for characterizing temptation and impulsivity. Bermúdez cautions as much, stating: ‘The opposite of self-control goes by a range of different names. It is variously called *weakness of will*, *impulsivity*, *incontinence*, *akrasia*, or *succumbing to temptation*, with different terminology often highlighting different aspects of a complex phenomenon’ (13). While conceptual diversity is to be expected and even welcomed in an essay collection, the dissimilarities that mark each chapter run deep and cut across disciplinary bounds. In some instances, this makes for a fresh and effective interdisciplinary discussion (such as in Green & Myerson and Rachlin); in other instances, it is decidedly less clear what is the exact contribution to this particular theme.

Ahmed (Ch. 4) ventures a model of incontinence that is said to capture three distinct forms of conflict, viz. preference reversals, prioritization of ephemeral desires, and synchronic conflict. According to his model, incontinence is problematic because it gives rise to real-world versions of Newcomb’s problem. This can be illustrated by considering the differences between causal decision theory (CDT) and evidential decision theory (EDT). Where CDT prescribes that one should maximize *utility*, EDT prescribes that one maximizes *news value*. EDT is sensitive to the agent’s counterfactual beliefs in ways the CDT isn’t: one’s choice (now) can be construed as evidence of what may occur (in the future), even if it does not causally induce what actually occurs. Ahmed indicates that the Newcomb-likeness of dynamic choice problems depends on the evidence having an effect on the probability of future occurrences separate from any causal interpretation of that probability. Hence, EDT can prescribe self-control, where CDT can’t. Ahmed concludes by suggesting that this model can be used to correct incontinence, specifically in the context of liberal paternalistic cases.

In contrast with Ahmed’s formal treatment of impulsivity, Green & Myerson (Ch. 5) and Rachlin (Ch. 6) consider how the mechanisms and structures underlying models of impulsive choice bear on questions about willpower and rationality. To this end, each discussion situates itself at the intersection of psychology and economics. Green & Myerson lead with an injunction against decision theory: if preference reversals are in need of explanation, then new decision-theoretic research must also be taken into account. This cause for warning is motivated by two new waves in psychological decision research: the first wave indicates that preference reversals aren’t always due to hyperbolic discounting; the second wave indicates a different role for conscious experience in decision-making based on studies with amnesiac patients. Green & Myerson suggest that the concept of discounting can provide a descriptive account of the trade-off between *delay* and *amount* for evaluating multi-dimensional alternatives, which can be determined by identifying the point at which individuals become indifferent about the subjective value of delayed rewards. This leads to their hypothesis that a ‘hyperboloid’ discount function provides a better fit to the existing data on preference reversal phenomena. They close out by raising questions about the function of episodic memory (such as mental time travel) on an agent’s ability to exercise self-control, using Loomes & Sugden’s (1982) anticipated regret theory as a foil. Following Green & Myerson, Rachlin (Ch. 6) takes issue with current treatments of impulsive choice, focusing on the supposed irrationality of addictive behavior. He asks: ‘what is the best way to use the concept of rationality when talking about addictive behavior?’ The answer, he suggests, ought to be understood in terms of overt behavioral patterns, not in terms of a logical mechanism. The key to this answer rests in distinguishing two common approaches to the study of rationality: these are the *neurocognitive model* and the *behavioral-economic model*. Where the neurocognitive model views rationality as the product of a logical decision mechanism, the behavioral-economic model assumes that people behave so as to maximize utility under constraints. Hence, when theories of the neurocognitive model of rationality are disconfirmed, they undergo modification to account for discrepancies. Differently, the method of revealed preference theory which undergirds the behavioral-economic model is to discover the utility function that is maximized under all constraints (ideally). Drawing on previous research (Jones & Rachlin, 2006; Rachlin &

Jones, 2008), Rachlin makes the case that there is nothing inherently irrational about hyperbolic discounting (thus echoing Green & Myerson). What *may be* irrational is the failure of an addict to account for and/or predict a change of mind when it is known that they are likely to reverse their preferences.

In bringing the second theme to a close, Andreou (Ch. 7) takes the philosophical project raised by Rachlin ('what is the best way to use the concept of rationality when talking about addictive behavior?') further by inquiring why humans experience temptation at all? Her aim is to show that our susceptibility to temptation in familiar cases (such as those discussed in Ch. 1–3) are not anomalous: temptation is integral to a type of motivational system that is generally quite well-suited to the task of balancing multiple on-going goals. The human motivational system actually resolves puzzling cases of temptation by its very nature. The conception of temptation that is of interest here figures as a test of instrumental rationality with failure understood as failure in relation to an agent's goals or the importance assigned to each of them.

Chapters (8)–(11) round out the collection by considering how self-control is possible at all, focusing in particular on the mechanisms (both formal and psychological) that enable precommitment and other self-control technologies. This is by far the most compelling part of the book, and it somehow feels well deserved after navigating the previous four chapters. The final four chapters converge thematically and conceptually, dovetailing the normative considerations raised in the first third of the book, with the descriptive ones raised in the second third. The one outlier, surprisingly, is Mele's contribution (Ch. 9), which I return to below.

Bermúdez (Ch. 8) leads off with an ambitious and wide-reaching framework that attempts to resolve the normative disputes about the formal requirements of instrumental rationality. It proposes to do so by drawing upon the psychological research on framing effects. When a decision-maker consciously considers a single outcome under two different frames and evaluates it differently in each frame, this is paradigmatic of irrational decision-making – which is to say: 'clashes of frames are not supposed to happen' (179). Bermúdez's aims to prove this perspective wrong, and further, to show how the exploitation of framing effects can be used to combat temptation. In so doing, he introduces the notion of *ultraintensionality* to make the case that it can, in fact, be rational to assign different values (utilities) to the same outcome under different presentations – practical reasoning, he argues, is an ultraintensional context. Thus being the case, he rejects both McClennen's (1990) and Holton's (2009) arguments that self-control is made possible through intentions to be resolute. The remainder of the chapter is dedicated to arguing why frame-sensitive valuation can be rational and to exploring the ways that framing can engender self-control.

This brings us to Mele's contribution (Ch. 9). Mele addresses the target question – how is it possible to exercise self-control? – from the perspective of philosophy of action, introducing a number of modifications to the paradigm case. Of those modifications, two stand out as odd: one is the introduction of the concept of an *all-things-considered judgment* (agents judge what is best for them 'on the whole'); the other is called *Principle T*, which states that intentional action reflects what agents are most strongly motivated to do at a given time. Presumably, Principle T makes the decision problem all the more interesting for it essentially rules out the possibility of self-control on the original reading of the paradigm case. Yet, Mele's answer to the question is surprisingly underwhelming: he suggests that persons may hold and act upon diverse intentions simultaneously, allowing them to be strongly motivated by temptation and at the same time to intend to resist temptation. What determines whether (or not) a person sticks to their all-things-considered judgment to avoid temptation depends on whether their evaluation of the objects of desire can be brought to bear on the motivational strength of those desires. Issuing self-commands also seems to play a role.

In the final two chapters, Gold (Ch. 10) and Easwaran & Stern (Ch. 11) present complementary theses depicting how self-control emerges from the cooperation among transient agents. Gold's contention is that standard decision theoretic models of self-control lack the concept of an intention that guides behavior over time – this is the recurring problem for resolute choice. Gold argues that, in fact,

the theory of intrapersonal team reasoning not only explains why it is rational to exert self-control, but also demonstrates how it is possible (absent an external mechanism or precommitment tool). To achieve cooperation, one has two options: If you apply Sugden's (1993, 2003) conception of mutually assured team reasoning, you have to reconcile the assumption of reciprocity and social contract with selves over time (this is unrealistic). Likewise, if you apply Bacharach's (1999, 2006) conception of circumspect team reasoning, you have to reconcile the assumption of unconditional cooperation (this is also unrealistic). Gold surmises that a Bacharach-style theory of intrapersonal team reasoning could generate a plausible theory of rational self-control if the structure of the intrapersonal problem is couched in terms of a *threshold public goods game*. Interestingly, Easwaran & Stern (Ch. 11) take the opposite approach: they posit that self-control can be achieved without adverting to a mechanism for normative coherence (which is what Gold's account of the Bacharach-style team reasoning essentially does). They consider different cases in which time-slices of an agent, which have unique first-person interests and no future-oriented intentions, can avoid temptation and achieve something like intertemporal unity. The explanation for this unity arises from the fact that each agent's utility function is sensitive to the previous agent's choices – this view assumes that each time-slice experiences greater marginal utility by continuing on its course of action rather than either switching or stopping to deliberate on the best course of action. In short, the cases that Easwaran & Stern consider demonstrate that there are multiple ways to achieve diachronic unity; though, they don't show that an agent's earlier time-slices can govern her later time slices without changing what the later time slices want to do. Put another way, they don't shed light on what is rationally required to achieve self-governance when one has prior intentions to act, but one is currently neutral about her actions.

In conclusion, the study of self-control is an intellectually demanding job. This is due partly to the fact that experts often disagree over what justifies self-control, which is to say, they disagree over what is required of agents to be rational in the face of temptation. It is also due to the fact that 'the experts' constitute a mixed bunch – philosophers, psychologists, and economists may have equal stakes in what constitutes practical rationality and self-control, yet they don't always interpret the normative and descriptive questions in the same ways or with the same sensitivities. This collection of essays is a testament to both these facts. To this end, Bermúdez has achieved something special in this book: not only has he captured the state of the art by collating a representative sample of the most pertinent issues on the topic of self-control from a broadly instrumental decision-theoretic conception of rationality, but he has done so in a way that encourages interdisciplinarity. For these reasons, I highly recommend this collection to graduate students and young scholars looking for new questions and new directions to take their work.

References

- Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics*, 53 (2), 117–147.
- Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory*. Princeton, NJ: Princeton University Press.
- Gauthier, D. (1987). *Morals by agreement*. Oxford: Oxford University Press.
- Gauthier, D. (1994). Assure and threaten. *Ethics*, 104(4), 690–721.
- Holton, R. (2009). *Willing, wanting, waiting*. New York, NY: Oxford University Press.
- Jones, B., & Rachlin, H. (2006). Social discounting. *Psychological Science*, 17(4), 283–286.
- Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368), 805–824.
- McClennen, E. F. (1990). *Rationality and dynamic choice: Foundational explorations*. Cambridge, UK: Cambridge University Press.
- Rabinowicz, W. (1995). To have one's cake and eat it, too: Sequential choice and expected-utility violations. *The Journal of Philosophy*, 92(11), 586–620.
- Rabinowicz, W. (1997). Wise choice on dynamic decision-making without independence. In E. Ejerhed & S. Lindström (Eds.), *Logic, action and cognition* (pp. 97–109). Dordrecht: Springer.

- Rachlin, H., & Jones, B. A. (2008). Social discounting and delay discounting. *Journal of Behavioral Decision Making*, 21(1), 29–43.
- Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy and Policy*, 10(1), 69–89.
- Sugden, R. (2003). The logic of team reasoning. *Philosophical Explorations*, 6(3), 165–181.

James D. Grayot

*Department of Theoretical Philosophy, University of Groningen
Tilburg Center for Logic, General Ethics, and Philosophy of Science (TiLPS), Tilburg University*

james.grayot@gmail.com

© 2020 James D. Grayot

<https://doi.org/10.1080/1350178X.2020.1729531>