Behavioural Public Policy (2023), 1–8
doi:10.1017/bpp.2023.1

CAMBRIDGE UNIVERSITY PRESS

BOOK REVIEW ESSAY

Escaping Paternalism: Rationality, Behavioral Economics and Public Policy

Mario J. Rizzo and Glen Whitman. Cambridge University Press, 2020, xii+496 pages.

Malte Dold

Economics Department, Pomona College, Claremont, CA, USA
E-mail: malte.dold@pomona.edu

(Received 10 January 2023; accepted 10 January 2023)

Nudges as paradigmatic i-frame policies

To date, many behavioral economists have located the essential policy problem at the level of individual choice. Chater and Loewenstein (in press) call this i-frame policies. Individuals are assumed to be prone to a myriad of cognitive and behavioral biases, such as vulnerability to framing, myopia or a lack of self-control. These biases are identified as the target of policymaking since they are said to prevent people from choosing what is best for them. Typical i-frame policies are *nudges*, i.e., subtle changes in the choice context that are supposed to help people choose what they "truly" prefer without taking any choice options off the menu. The list of proposed nudges is long and includes polices such as default enrollments in saving plans, cooling-off periods to prevent impulsive purchases or graphic warnings on cigarette packages. What unites those policies is the claim that they improve individual welfare – not by some exogenous standard, but by people's own lights (see, e.g., Halpern, 2015; Le Grand & New, 2015; Thaler & Sunstein, 2021).

Despite its success, prominent scholars have argued in recent years that the insights of behavioral economics do not warrant sweeping policy implications, particularly when it comes to the use of behavioral insights to justify i-frame policies such as nudges. Outspoken critics in this context are, for instance, Sugden (2018) and Oliver (2023). Those critical voices have gotten additional argumentative ammunition in the form of Mario Rizzo and Glen Whitman's book *Escaping paternalism: Rationality, behavioral economics and public policy* (Rizzo & Whitman, 2020). The book is a 500-page tour de force that questions the conceptual, empirical and practical foundations of behavioral paternalism, i.e., paternalism that uses insights from behavioral economics to justify governmental interventions. In Rizzo and Whitman's (RW) own words, the book presents "a series of challenges – in effect, hurdles that behavioral paternalist proposals must clear in order to be justified as a

matter of policy" (p. 16). RW make clear that they "don't expect that every reader will find all of our challenges to paternalism equally compelling. … But our hope is that, taking the gauntlet of challenges *as a whole*, readers will recognize just how tenuous the entire new-paternalist enterprise is" (p. 20).

To be clear, RW welcome the insights behavioral economics provides into choice patterns that neoclassical economics cannot easily explain, such as when people ex-ante "choose not to choose" in the form of self-commitment strategies in situations of temptation (e.g., putting cookies on high shelves, flushing unsmoked cigarettes down the toilet, etc.). Neoclassical economics has a hard time explaining such behavior since self-commitment is costly and the rational choice assumption implies that people choose what is feasible and what they most prefer irrespective of the availability of inferior alternatives. In making sense of such behavior, RW acknowledge that behavioral economics has allowed economists to better understand and model the ways in which HUMANS (as opposed to hyper-rational ECONS) make choices. RW state that "[to] the extent that behavioral economics has exposed the genuine failings of the old rational-choice models, it has been a boon to the economics profession" (p. 3). Yet, RW also point out that while behavioral economists have compellingly criticized the way choice is modeled in neoclassical economics, many have been too quick to jump from their descriptive analyses to normative conclusions about paternalistic i-frame policies such as nudges.

Target I: Narrow understanding of rationality

The book consists of two main parts: Chapters 1–5 present a critical assessment of the conceptual foundations of behavioral paternalism. Even scholars who are familiar with behavioral economics will find this part to be a valuable review of the main concepts underlying contemporary discussions of behavioral paternalism. Chapters 6–9 present a dive into a myriad of practical challenges of behavioral paternalism in the political process. This second part is particularly relevant for all those scholars and think tankers actively engaged in behavioral public policy.

In the first part of the book, the main target of RW's conceptual criticism is the narrow understanding of rationality in behavioral paternalism. RW point out that "despite having rejected rationality as a model of how people *do* behave, the behavioral paternalists still accept rationality as a model of how people *ought* to behave" (p. 16). In this sense, the "correct" way of making choices is defined as maximizing the satisfaction of well-defined preferences, i.e., preferences that are complete and consistent. In doing so, RW argue that behavioral paternalists "have made the mistake of conflating their models with reality – and, when reality fails to conform to the model, judging it deficient" (p. 180). RW rightly deem this a peculiar development in the history of ideas since the restrictive assumptions of neoclassical rationality have originally been adopted not because they were especially plausible from a normative or welfare perspective, but because they were analytically convenient and allowed for mathematical tractability.

A common and convincing thread in RW's discussion of rationality is that it is difficult for a theorist-economist to externally define what it means to behave rationally and what it means to make a mistake. RW argue that "[in] the rush to

https://doi.org/10.1017/bpp.2023.1 Published online by Cambridge University Press

characterize certain 'anomalies of choice' as violations of rationality, behavioral paternalists have been insufficiently subjectivist" (p. 17). By this, RW mean that behavioral paternalists do not follow their mantra and seek policies on the basis of people's own values and preferences. Instead, they apply an external set of values by taking the neoclassical definition of rationality as a normative standard of "good" decision-making. What makes it worse is the fact that behaving according to neoclassical rationality does not necessarily translate into welfare-improving choices. According to RW, experimentation, making mistakes and self-discovery are crucial aspects of individual welfare. Hence, inconsistencies between one's preferences and values can be seen as integral part of people's pursuit of welfare. RW note that violations of consistency are only occasionally welfare-relevant, as in the rare case of a money pump scenario.

An astute observation of RW is that in their prescriptive understanding of neoclassical rationality, behavioral paternalists do not only subscribe to the normative idea of preference consistency, but also to a set of restrictive assumptions about people's beliefs. More specifically, people's beliefs ought to be logically coherent, reflect the facts of the world (i.e., they must be truth-tracking) and be responsive to new evidence (i.e., belief updating must follow Bayes' Rule). RW question whether these restrictive assumptions about beliefs are necessarily welfare relevant. First, they argue convincingly that it is in many cases unfeasible for people to hold logically coherent beliefs. Belief coherence is cognitively costly and hence unlikely an optimal welfare strategy. Second, RW challenge the idea that "any divergence between one's beliefs and the truth has the potential to generate suboptimal decisions" (p. 121). "Biased" beliefs can be a direct source of pleasure or comfort, enhance and motivate people's performance or be a source of learning. Third, RW question whether following Bayes' Rule constitute the uniquely reasonable way to process information and form beliefs; for instance, when updating the probability of events considering new evidence, people may reasonably weight their local experience more strongly than formal base rates. Or, when receiving "factual information", agents may reasonably include tacit knowledge from the context that would not be captured in a formal account of evidence. RW point out that behavioral paternalists neglect a key – and one may want to add obvious – issue in their normative understanding of neoclassical rationality, viz., "whether that standpoint is important and relevant to the agents themselves" (p. 31). RW give good reasons why behavioral economists should refrain from taking neoclassical rationality as a general benchmark for welfare-improving choices.

Target II: The complexities of applying behavioral insights in the political process

In the second part of the book, RW discuss a series of practical problems of implementation of behavioral paternalism. RW start off by pointing out that it is not self-evident how analysts ought to translate the myriad of findings about people's biases gained in the artificial and controlled context of lab experiments into real-world policy settings. It remains unclear what the magnitude and prevalence of the identified bias outside the lab is if one doesn't gather field data in the real-world context in

https://doi.org/10.1017/bpp.2023.1 Published online by Cambridge University Press

which a proposed paternalistic policy is to be implemented. Moreover, RW argue that in many real-world contexts people are aware of their cognitive biases and often come up with ingenious self-debiasing strategies, including asking for advice or reasoning in small groups. Consequently, the magnitude of biases in real-world settings might be less pronounced than under the artificial setting of lab experiments. This raises an intricate knowledge problem for policymakers. It is hard for external parties to judge the degree of a person's bias or whether that person has developed the "right" amount of self-control. These arguments are convincing, particularly RW's worry that behavioral paternalistic policies might crowd out self-regulatory behavior by greater external control. Yet, some readers might have reasonable doubts informed by introspection that people are as effective in overcoming their self-control issues as RW seem to insinuate in this part of the book. Others might point out that there is actually a robust demand for governmental paternalism (or "parentalism") since people want to escape, evade and even deny personal responsibilities for difficult choices (Buchanan, 2005). This reviewer would have liked RW to engage more with the demand side of behavioral paternalism in this part of the book.

RW identify another – potentially more serious – dimension of the knowledge problem. To successfully implement paternalistic policies, policymakers would need to possess a high level of knowledge about people's "true" preferences. RW give several good reasons why such knowledge is very hard to acquire. By rejecting the revealed preference theory of welfare, behavioral economists face an epistemological dilemma of how to identify people's "true" preferences. It is unclear in which choice contexts and – given the evolving nature of people's preferences – at which point in time, people's choices or verbal statements should be taken as normative input for policymaking, e.g., what should be taken as the true rate of intertemporal discounting if a person depicts different rates in different choice contexts? RW argue that there is no uncontroversial basis for that determination. And even if policymakers have identified a set of relatively stable and consistent preferences, to successfully implement paternalistic policies, they would need to have knowledge about the extent and prevalence of biases in the population, how people's biases interact with and offset one another, and how paternalistic policies may cause substitution effects (e.g., an intervention that discourages vaping might reduce vaping but increase cigarette smoking).

As a next challenge, RW discuss behavioral paternalism from a public choice perspective. RW compellingly warn against the "nirvana fallacy" (p. 310), i.e., the idea that any discrepancy between "the ideal and the real" is deemed sufficient to justify behavioral paternalistic intervention. It might be true that real people fall behind some self-identified behavioral ideal. Yet, for a host of reasons policy interventions that aim at closing the gap between the "real" and the "ideal" might be worse than what (admittedly imperfect) individuals engaged in learning and adaptation may come up with spontaneously. One reason is that policymakers act on a similar set of cognitive biases like the people they are supposed to help, such as overconfidence, confirmation or salience biases. Another reason is that behavioral paternalistic policies might be particularly vulnerable to the influence of special interest groups. Historically, questions of paternalism have often been dominated by interest groups with strong financial interests ("bootleggers") or groups with strong religious and ideological views ("baptists"). Consequently, RW caution that behavioral paternalistic

https://doi.org/10.1017/bpp.2023.1 Published online by Cambridge University Press

policies "will tend to promote some combination of [the policymaker's] preferences, socially approved preferences, or special-interest preferences – none of which are synonymous with the real preferences of people targeted by paternalistic laws" (p. 20). This reviewer fully agrees that such public choice considerations are particularly relevant for paternalistic legislation of all types, including the "softer" or allegedly "libertarian" versions.

As a final hurdle that behavioral paternalism must clear, RW discuss its inherent slippery-slope tendency, i.e., small or moderate paternalistic interventions that increase the likelihood of more intrusive and autonomy-reducing interventions in the future. Distilling the literature on slippery slope phenomena, RW identify several key factors that increase the likelihood of policy slopes. Among these are vague and ill-defined concepts as well as complex interaction and crowding out effects. RW argue that these factors apply forcefully to behavioral paternalism. Core concepts of behavioral paternalism (such as welfare, freedom or autonomy) are often left vague and used in an ad-hoc manner. While the starting point of behavioral paternalism discussion centers around people's "true" preferences, the difficulty of conceptualizing and measuring them means that there is a latent tendency to move away from the agent's perspective to the experts inserting their own values, e.g., in the case of sin taxes where the appropriate normative rate of time discounting is typically assumed to be the longer-run rate. RW conjecture that this move "reflects, no doubt, certain intellectual middle-class values – not coincidentally, the values of many experts" (p. 371). Moreover, initial soft paternalistic policies might have unintended consequences "because of the interaction of the targets' biases, the crowding out of the targets' self-regulatory behaviors, and the substitution between targets' personal inputs" (p. 365). A lack of intended results can lead policymakers to advocate more aggressive interventions to speed up the efficacy of the soft paternalistic policy they had initially promised.

RW give convincing arguments for why slippery slopes might be particularly relevant for behavioral paternalist policies. While RW discuss anti-smoking campaigns of the last decades as an illustrative example, their arguments are mainly theoretical in nature. Some readers might like to see a more extensive discussion of how prevalent such slippery slopes really are in the day-to-day practice of behavioral public policy. They might wonder whether RW's worry is justified that "if withholding information can be the correct choice, it might also be appropriate to lie – if such lies do a better job of pushing people toward what policymakers think are their best interests" (p. 377). Such a policy would be incompatible with liberal democratic principles, and therefore, one can hopefully doubt whether lying as a means in behavioral public policy would ever gain real traction and support by a majority of legislators. Admittedly, this doubt might be fueled by a youthful optimism in institutional checks-and-balances in liberal democracies.

A paternalism-resisting framework

In the final chapters, RW turn the page from deconstructing the foundations of behavioral paternalism to discussing an alternative approach to policymaking which they call *paternalism-resisting framework*. RW correctly observe that the

https://doi.org/10.1017/bpp.2023.1 Published online by Cambridge University Press

focus on i-frame policies in behavioral economics tends to formulate the policy problem "not as 'whether or not paternalism is desirable,' but as 'what form of paternalism shall we have?'" (p. 392). It is not necessary to look at behavioral problems in this way. A core idea of RW's alternative framework is the notion of *inclusive rationality*. Inclusive rationality "means purposeful behavior based on subjective preferences and beliefs, in the presence of both environmental and cognitive constraints … inclusive rationality does not dictate the normative structure of preferences and beliefs a priori. Instead, it allows a wide range of possibilities in terms of how real people select their goals, form and revise their beliefs, structure their decisions, and conceptualize the world" (p. 26). Taking inclusive rationality as a normative standard for "good" choosing is very different from neoclassical and behavioral economics' endorsement of the axioms of neoclassical rationality. While the former takes humans "as they are" the latter "carries the distinct danger of measuring real humans relative to the model – and judging them deficient when they depart from it" (p. 433).

An interesting question is how RW's paternalism-resisting framework would look like in practice. The authors give only a partial answer by suggesting a three-step procedure to behavioral public policy (p. 393): (1) analysts should first search for mechanisms that show ways in which people are in control of their choice sets; (2) if no such mechanisms are in operation, analysts need to inquire as to whether people's preference to behave differently is true or merely cheap talk; and (3) when the analyst is satisfied that (2) is actually the case and people cannot develop spontaneous debias strategies, the discussion of government paternalism begins. While this three-pronged procedure is laudable, RW do not spell out the conceptual details for its implementation. RW's emphasis on the notion of inclusive rationality makes it very hard to succeed with (1) and pinpoint when people are actually making mistakes and aren't in control of their choice sets. Inclusive rationality is such a broad concept that it lacks analytical clarity when it comes to concrete questions of choice evaluation. Moreover, to fulfill (2), analysts would need to know more about the conditions that define when to take people's preferences seriously, e.g., whether ex post assessments of decision-makers (such as feelings of regret or relief) should be taken as a characteristic of "genuine" preferences that rule out cheap talk. Since it is unclear how analysts should go about in (1) and (2), they might indeed never reach (3). In this sense, RW truly provide a paternalism-resisting framework – not necessarily because of argumentative superiority, but because of the conceptual broadness of inclusive rationality and the analytical vagueness of their proposed procedure. Of course, RW could possibly reply that conceptual broadness and analytical vagueness are features and not bugs of their framework.

While the book is already very long, this reviewer would have liked RW to discuss in more detail the notion of inclusive rationality as a normative standard. In particular, since RW accept context-dependence as a behavioral force, it would have been illuminating to see them discuss contexts that are actually conducive to people's inclusively rational thinking. RW state that "inclusive rationality … encompasses all manner of strategies people use to shape their own behavior and interpret the world around them" (p. 433). This perspective puts normative emphasis on the idea that people "are capable of stepping outside the model, reconceptualizing it, and framing their own decisions in new ways. They are thus able to see their own

https://doi.org/10.1017/bpp.2023.1 Published online by Cambridge University Press

behavior, judge it, and potentially act to change it" (p. 433). These statements reveal that RW highlight the significance of *agentic capabilities* as core ingredients of "good" choosing, i.e., people's capacity to imagine and evaluate choice options, to form preferences in a self-reflective way, and, thereby to make choices they can identify with and take responsibility for. Following this line of thought, there are unexplored links between RW's notion of inclusive rationality and recent discussions of the normative significance of agency in behavioral public policy (Dold & Lewis, 2022b).

In this regard, the recent literature on "boosts" might help address the largely unanswered question of what kinds of policies and institutions increase individuals' agency. Boosts are intended to "foster competences through changes in skills, knowledge, decision tools, or external environment" (Hertwig & Grüne-Yanoff, 2017: p. 974). Paradigmatic boosts include teaching people tools for improving motivation and self-control or training them to translate relative probabilities into natural frequencies (Hertwig & Grüne-Yanoff, 2017: p. 979). Boosts do not require knowledge about people's "true" preferences since they do not aim at steering people toward concrete choice outcomes. Instead, their goal is to facilitate the choice process by fostering agency-enhancing competences. Advocates of this approach emphasize that boosts "require the individual's active cooperation" and that "[i]ndividuals choose to engage or not to engage with a boost" (Hertwig & Grüne-Yanoff, 2017: p. 982). In this sense, boosts take a core concern of RW seriously that behavioral economists as analysts and policy advisors should "approach [humanity] as fellow human beings doing the best they can, trying to improve their own choices, and offering friendly advice on how others might do the same" (Rizzo & Whitman, 2020: p. 439).

To date, RW's book is the most comprehensive discussion of the conceptual foundations of behavioral paternalism and its potential epistemic and practical problems. The book does an excellent job in critically discussing the often taken-for-granted foundations of behavioral paternalism (particularly the normative understanding of neoclassical rationality) and the practical problems of its implementation in the political process. Any scholar or policy analyst interested in behavioral paternalism – especially those interested in nudging – should engage with "the gauntlet of challenges" RW present the reader with in their formidable book. Since its publication, the book has already provoked a series of scholarly reactions, from the publication of a special issue in the *Review of Behavioral Economics* (Cowen & Dold, 2021) to research on the implications of dynamic preferences for tax policy (Delmotte & Dold, 2022) to discussions of the possibility of a Hayekian behavioral economics (Dold & Lewis, 2022a). One can therefore hope that in the years to come RW's book will help advance the methodological debate about key concepts in behavioral economics and the normative debate about the implications of behavioral insights for policymaking.

References

Buchanan, J. M. (2005), 'Afraid to be free: dependency as desideratum', *Public Choice*, **124**: 19–31.

Chater, N. and G. Loewenstein (in press), 'The i-frame and the s-frame: how focusing on individual-level solutions has led behavioral public policy astray', *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X22002023

https://doi.org/10.1017/bpp.2023.1 Published online by Cambridge University Press

Cowen, N. and M. Dold (2021), 'Introduction: symposium on escaping paternalism: rationality, behavioral economics and public Policy by Mario J. Rizzo and Glen Whitman', *Review of Behavioral Economics*, **8**(3–4): 213–220.

Delmotte, C. and M. Dold (2022), 'Dynamic preferences and the behavioral case against sin taxes', *Constitutional Political Economy*, **33**(1): 80–99.

Dold, M. and P. Lewis (2022a), 'FA Hayek on the political economy of endogenous preferences: an historical overview and contemporary assessment', *Journal of Economic Behavior & Organization*', **196**: 104–119.

Dold, M. and P. Lewis (2022b), 'A neglected topos in behavioral normative economics: the opportunity and process aspect of freedom', doi: 10.13140/RG.2.2.33165.00486.

Halpern, D. (2015), *Inside the Nudge Unit: How Small Changes Can Make a Big Difference*. London: Random House.

Hertwig, R. and T. Grüne-Yanoff (2017), 'Nudging and boosting: steering or empowering good decisions', *Perspectives on Psychological Science*, **12**(6): 973–986.

Le Grand, J. and B. New (2015), *Government Paternalism*. Princeton: Princeton University Press.

Oliver, A. (2023), *A Political Economy of Behavioral Public Policy*. Cambridge: Cambridge University Press.

Sugden, R. (2018), *The Community of Advantage: A Behavioral Economist's Defense of the Market*. Oxford: Oxford University Press.

Thaler, R. H. and C. R. Sunstein (2021), *Nudge: The Final Edition*. New Haven and London: Yale University Press.

https://doi.org/10.1017/bpp.2023.1 Published online by Cambridge University Press