Behavioural Public Policy (2025), **9**, 324–330
doi:10.1017/bpp.2024.37

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Kahneman's tryst with reasonableness: a tease unfulfilled?

Sanchayan Banerjee[1] and Malte Dold[2]

[1]Vrije Universiteit Amsterdam, The Netherlands and [2]Pomona College, USA
Corresponding author: Sanchayan Banerjee; Email: s.banerjee@vu.nl.

(Received 31 July 2024; accepted 2 August 2024)

Abstract
Kahneman's criticism of neoclassical rationality was central to his research programme. He argued that rationality understood as temporal consistency among preferences and beliefs is inapt as a descriptive and prescriptive standard of decision-making. Descriptively, consistency ignores high decision costs and biases, such as framing effects. Prescriptively, it is problematic since it neglects the processual nature of choice and the crucial role of regret. Instead, Kahneman argued in favour of using *reasonableness* as a standard, though he did not fully develop the concept in his work.

Since the 1950s, the idea that human beings are rational has been a cornerstone of neoclassical economics. Rationality in neoclassical economics typically means that agents behave according to consistent preferences, follow the axioms of probability theory, and update their beliefs systematically when confronted with new evidence. As Daniel Kahneman (2003, 163) wrote 'the standard of rationality in economics was, and remains, the maximisation of subjective expected utility – a combination of von Neumann-Morgenstern preferences and a Bayesian belief structure.' However, beginning in the early 1970s and gaining traction in the 1990s, a psychological approach to economics emerged that challenged the assumption of neoclassical rationality. Pioneering work by Kahneman and his collaborator Amos Tversky demonstrated that human preferences and beliefs are subject to biases and can be influenced by how choices are framed, challenging the consistency axiom of rationality (Tversky and Kahneman, 1986).

Kahneman's influence has been profound, leading to the development of new fields such as behavioural economics and, more recently, *behavioural public policy* (*BPP*). While many of his theoretical insights in the context of prospect theory have been extensively discussed and applied to explain economic phenomena – such as loss aversion explaining the endowment effect (Thaler, 1980) – some of Kahneman's ideas remain underexplored but ripe for further development.

One such idea is *reasonableness*, which Kahneman contrasted with rationality, proposing the former as a more convincing or intuitive construct (Kahneman, 2011; Herfeld, 2014).[1] Although reasonableness has resurfaced in behavioural sciences recently (see, e.g., Madsen *et al.*, 2024), some initial explorations can be traced back to Kahneman. In memory of Kahneman, this article outlines his view of reasonableness and explores how this notion can align with current discussions about fostering and respecting individual agency in *BPP* (Dold and Lewis, 2023; Banerjee *et al.*, 2024). Our title, 'Kahneman's Tryst with Reasonableness: A Tease Unfulfilled?' is a fond tribute to his contributions in this direction.

Be reasonable, not rational

In several places of his oeuvre, Kahneman expressed his dissatisfaction with the construct of rationality in neoclassical economics (see, e.g., Kahneman, 2003, 2011). He criticised the theory as simple and elegant yet obviously false, questioning the premise of consistency underlying rationality. Kahneman highlighted the concept of coherence in the context of neoclassical rationality, i.e., the degree of consistency in people's choices over time. The notion of coherence is key to substantiate Kahneman's criticisms of neoclassical rationality – both when it is understood *descriptively* in the sense of how people actually behave and when it is used *prescriptively* in the sense of how people should behave.

Kahneman's critique of neoclassical rationality was based on two fundamental problems. First, he described rational choice theory in a 2014 interview as 'too demanding' (Herfeld, 2014, 3), arguing that it expects people to meet standards, such as coherence, that are unrealistic for finite minds.[2] This made neoclassical rationality untenable for a descriptively accurate theory and its predictions questionable, as people cannot always fully weigh the costs and benefits of their many opportunities and order them in a consistent manner, especially considering their potentially different future preferences. In addition, framing and menu effects contribute to people holding time-inconsistent preferences.

The second, more substantive critique is that rational choice theory is 'far too permissive' (Herfeld, 2014, 3). Kahneman argued that neoclassical rationality fails to account for people's decisions and feelings over time, allowing individuals to behave 'rationally from their point' (Herfeld, 2014, 7). This permissiveness means that almost any behaviour can be justified as rational by any individual. To support this, Kahneman cited Amartya Sen's concept of 'rational fools,' explaining that while someone with an addiction may act coherently and thus in accordance with rational choice theory, their behaviour must still be considered foolish, exemplifying what can

[1]Kahneman distinguishes between rationality-as-coherence and reasonableness in several places. In the conclusion of his magnum opus *Thinking, Fast and Slow,* he writes: 'The only test of rationality is not whether a person's beliefs and preferences are reasonable, but whether they are internally consistent. A rational person can believe in ghosts so long as all her other beliefs are consistent with the existence of ghosts. A rational person can prefer being hated over being loved, so long as his preferences are consistent. Rationality is logical coherence – reasonable or not' (Kahneman, 2011, 411).

[2]'The definition of rationality as coherence is impossibly restrictive; it demands adherence to rules of logic that a finite mind is not able to implement. Reasonable people cannot be rational by that definition, but they should not be branded as irrational for that reason' (Kahneman, 2011, 411).

https://doi.org/10.1017/bpp.2024.37 Published online by Cambridge University Press

be called 'rational foolishness.' The same permissiveness is not true for *reasonableness* though, a point that he made later in the interview.

Kahneman pointed out that the underlying ideas of consistency or coherence might not be what people care for after all. He put the blame broadly on rational choice theory's ignorance of people's own self-interest:

> '*I think it's true that when you confront people with the fact that they're not consistent, they're not horrified. Ok so we're inconsistent. People, in effect, know that they're not consistent. So the achievement of rationality, the achievement of coherence and consistency is not the highest value that people have …. I think rationality as defined in terms of coherence is very largely irrelevant to human affairs, because it doesn't incorporate any conception of human interests, or of what is in people's best self-interests.*' (Herfeld, 2014, 7)

Rational choice theory's inability to account for the future consequences of present actions limits its effectiveness in predicting behaviour. It also limits its relevance as a prescriptive standard for 'good' choice. Kahneman's critiques largely focused on rational choice theory's assumption that people make decisions solely with the present in mind, ignoring the possibility of future regret.[3] He illustrated this point with Gary Becker's theory of addiction, highlighting how rational choice theory overlooks long-term implications:

> '*Rational choice theory cannot take seriously the idea that an addict later will regret his choice. In [Becker's] theory, somebody who gets addicted and later regrets that he consumed drugs is compared to somebody who goes to a restaurant and has a large meal and then it turns out that he does not have money to pay for the meal. One has very little sympathy for the person who went to the restaurant but one might have sympathy for the addict, just because we consider that he or she might, in the future, regret his decision to consume drugs. I think that the sort of hyper rational choice theories cannot acknowledge that we might have sympathies for people, such as drug addicts, who make a decision that they can regret later.*' (Herfeld, 2014, 3)[4]

In view of these limitations, Kahneman proposed an alternative to rational choice theory: the construct of *reasonableness*. This concept includes people's reasoned assessment of societal norms and their own evolving interests but Kahneman did not fully define the notion. For Kahneman, being reasonable is an open-ended concept that means something like *broadening the bracketing of a decision* and *being responsive to plausible reasons* (Kahneman, 2011). Being reasonable is not the same as following slow, effortful System 2 thinking, as the latter's 'abilities are limited and so is the knowledge to which it has access. We do not always think straight when we reason, and the errors are not always due to intrusive and incorrect intuitions' (Kahneman,

[3]For a longer discussion of the role of regret in decision-making, see Kahneman (2011, 342–353).

[4]On Kahneman's view of Becker's rational theory of addiction, see also Kahneman (2003, 165; 2011, 412).

https://doi.org/10.1017/bpp.2024.37 Published online by Cambridge University Press

2011, 415). Unlike rationality, reasonableness allows for time-inconsistent behaviour and includes factors not addressed by coherence, such as respect for regret and the possibility that the preferences of one's future self are different from the ones of the currently acting self. In practice, people see a reasonable person as someone who acts in a way that their future self will approve of, rather than merely accepting the consequences of their current actions.

Reasonableness is key to human agency

Kahneman's exploration of reasonableness and his proposal to move beyond rationality in social and behavioural sciences carry profound implications for advancing the notion of *individual agency* in *BPP*. To date, there is no unified notion of agency in *BPP* (Dold, 2023; Banerjee *et al*., 2024). Yet, the observation that humans exhibit reasonableness by considering their future selves in decision-making and the impact of their actions on others aligns with many ongoing discussions on agency in *BPP*. Agency-centric approaches (e.g., Hargreaves Heap, 2017; Dold and Stanton, 2021; Dold and Lewis, 2023; Hargreaves Heap, 2023; Banerjee *et al*., 2024) are united in their critique of approaches that treat behavioural outcomes (e.g., eat less fatty food or go more often to the gym) as target variables of policy interventions and exploit people's cognitive biases (e.g., menu dependence or status quo bias) to achieve those outcomes. In contrast, agency-centric approaches focus on improving the *quality of the reasoning process* that precedes choice.

The debate over how much agency is appropriate and what interventions foster agency effectively remains contentious within *BPP* (Banerjee *et al*., 2024). Recently, there has been renewed interest in shifting from individual interventions to broader system-level approaches that tackle structural issues (e.g., subsidies for the sugar industry or lack of social security) that drive behavioural phenomena, such as obesity or old-age poverty (Chater and Loewenstein, 2023; Dold, 2023). This contrasts with the standard practices in modern behavioural sciences of nudging individuals towards choices that improve their welfare (Thaler and Sunstein, 2008). Yet, system-level interventions might be consistent with and complementary to newer proposals aimed at enhancing reasonableness and human agency by enabling individuals to freely form their own intentions and act on them (Dold and Lewis, 2023; Banerjee *et al*., 2024). Similarly, the idea, rooted in a liberal perspective on the political economy of *BPP* (Oliver, 2023), that behavioural interventions are justified when they address externalities rather than internalities, reflects the distinction between viewing individuals as reasonable rather than strictly rational. In the interview with Catherine Herfeld, Kahneman himself underscored three specific reasons that hint at why reasonableness aligns with an approach that enhances individual agency in *BPP* (Herfeld, 2014).

First, Kahneman highlighted how the concept of reasonableness takes into account the context and future implications of a decision. Unlike neoclassical rationality, which might justify an addiction as rational, reasonableness considers whether the decision aligns with the view of the individual as *the author of her own life* who is capable of forming intentions self-reflectively *ex ante* and fully identifying with her choices *ex post*. This account acknowledges the possibility of intention-action gaps

https://doi.org/10.1017/bpp.2024.37 Published online by Cambridge University Press

and decisions that individuals will regret in the future, such as consuming drugs that lead to addiction which undermine one's agency over time:

> 'With a theory of reasonableness, you don't have a reasonable addiction, except for people who are about to die or something, it's reasonable for them to be addicted to morphine. But otherwise, while you might be able to have rational addiction, it cannot be reasonable.' (Herfeld, 2014, 8)

Second, Kahneman emphasised the importance of the 'remembering self' in evaluating reasonableness. Once again, this account respects human agency by extending the temporal dimension of analysis and considering the individual's future perspective beyond immediate satisfaction, thus promoting decisions that are in line with the more permanent view an agent has about herself and her life:

> 'What I say is that a theory of reasonableness takes the remembering self very seriously. That's what I meant. So it's the retrospective view of behaviour, which is the perspective that defines whether behaviour is reasonable or not. It's not reasonable to do something that you will regret later.' (Herfeld, 2014, 8)[5]

Third, Kahneman suggested that evaluating decisions as reasonable involves considering their future impact and potential regret, acknowledging the uncertainty of the future and evolving preferences. As before, this respects human agency by acknowledging the *processual nature of the self*, which is a core idea of many agency-centric approaches (Dold, 2023). Unlike rational choice theory, which focuses on the immediate decision point, reasonableness considers the ongoing and changing nature of human feelings and emotions, supporting a more comprehensive view of individual agency:

> 'You can evaluate the decision now as reasonable or not, if it takes into account, in a sensible way, the perspective of the future. And by that I mean that the future is probabilistic and uncertain and so on. But the focus on the future and on regret, I think, is really quite important in defining reasonableness, because that's where the more permanent interest of the individuals comes in. That's where the future comes in and the relevance of the future selves. The moment that you abandon consistency, then you allow the self at different times to have different feelings and emotions and so on. The dominant perspective, the perspective of rational choice theory, is exclusively the point of decision.' (Herfeld, 2014, 8)

Kahneman's account for *reasonableness* places individuals' capacity to reflect and reason at the centre of decision-making. It is critical to realise that for

[5]Kahneman (2011, 352) admits that a feeling of regret should not be taken at face value and as the sole criterion for reasonableness: 'regret and hindsight bias will come together, so anything you can do to preclude hindsight is likely to be helpful. … Hindsight is worse when you think a little, just enough to tell yourself later, "I almost made a better choice. … you should not put too much weight on regret; even if you have some, it will hurt less than you now think."

https://doi.org/10.1017/bpp.2024.37 Published online by Cambridge University Press

reasonableness to prevail, it is important that individuals develop and form their intentions self-reflectively in dialogue with others (Hargreaves Heap, 2023). *BPP* has only recently begun to explore interventions that foster people's agentic capabilities to reflect and reason, and in doing so enhance their reasonableness (Dold and Lewis 2023; Banerjee *et al*., 2024). While Kahneman's tryst with reasonableness is a tease unfulfilled, we hope that future work takes up the idea and explores its descriptive validity and prescriptive relevance for the debate about the possibilities and limits of an agency-centric *BPP*.

Putting people's reasonableness at the core of *BPP* debates might challenge paternalistic attempts to steer individual behaviour, but still highlight the crucial role of the situational and social environment for individuals to exercise their agency. For instance, there is encouraging evidence that *experiments in living* (Sharot and Sunstein, 2024) and *structured deliberation* (Niemeyer *et al*., 2024) can help people become aware of the context-dependent nature of their evolving preferences, thus extend the temporal dimension of their decisions, and consider their options transpositionally – all points in line with Kahneman's idea of reasonableness. In further exploring these avenues, it is clear that Kahneman's legacy cannot (and should not) be reduced to his early insights on biases and heuristics, prospect theory, or system I and II thinking, but must be seen as a set of provocative ideas that provide food for thought for generations of behavioural researchers to come.

References

Banerjee, S., T. Grüne-Yanoff, P. John and A. Moseley (2024), 'It's time we put agency into behavioural public policy', *Behavioural Public Policy*, 1–18.

Chater, N. and G. Loewenstein (2023), 'The i-frame and the s-frame: how focusing on individual-level solutions has led behavioral public policy astray', *Behavioral and Brain Sciences*, **46**(e147): 1–84.

Dold, M. (2023), Individual agency in behavioural public policy: New knowledge problems. Forthcoming in *Social Philosophy and Policy*. https://tinyurl.com/y85a5nz8.

Dold, M. and P. Lewis (2023), 'A neglected topos in behavioural normative economics: the opportunity and process aspect of freedom', *Behavioural Public Policy*, **7**(4): 943–953.

Dold, M. and A. Stanton (2021), 'I choose for myself, therefore I am: the contours of existentialist behavioural economics', *Erasmus Journal for Philosophy and Economics*, **14**(1): 1–29.

Hargreaves Heap, S. P. (2017), 'Behavioural public policy: the constitutional approach', *Behavioral Public Policy*, **1**(2): 252–265.

Hargreaves Heap, S. P. (2023), 'Mill's constitution of liberty: an alternative behavioural policy framework', *Behavioural Public Policy*, **7**(4): 933–942.

Herfeld, C. (2014), *A Conversation with Daniel Kahneman. Forthcoming in Herfeld, Conversations on Rational Choice*, Cambridge, UK: Cambridge University Press, https://philarchive.org/archive/HERACW-2.

Kahneman, D. (2003), 'A psychological perspective on economics', *American Economic Review*, **93**(2): 162–168.

Kahneman, D. (2011), *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Madsen, J. K., L. de-Wit, P. Ayton, C. Brick, L. de-Moliere and C. J. Groom (2024), 'Behavioral science should start by assuming people are reasonable', *Trends in Cognitive Sciences*, **28**(7): 583–585.

Niemeyer, S., F. Veri, J. S. Dryzek and A. Bächtiger (2024), 'How deliberation happens: enabling deliberative reason', *American Political Science Review*, **118**(1): 345–362.

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

Sharot, T. and C. R. Sunstein (2024), *Look Again: The Power of Noticing What Was Always There*. London: The Bridge Street Press.

https://doi.org/10.1017/bpp.2024.37 Published online by Cambridge University Press

Thaler, Richard (1980), 'Toward a Positive Theory of Consumer Choice', *Journal of Economic Behavior and Organization*, **1**: 39–60.
Thaler, R. and C. Sunstein (2008), *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven: Yale University Press.
Tversky, A. and D. Kahneman (1986), 'The Framing of Decisions and the Evaluation of Prospects', *Studies in Logic and the Foundations of Mathematics*, **114**: 503–520.

Cite this article: Banerjee, S. and M. Dold (2025), 'Kahneman's tryst with reasonableness: a tease unfulfilled?', *Behavioural Public Policy* **9**, 324–330. https://doi.org/10.1017/bpp.2024.37