CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Why Kahneman matters

Mario J. Rizzo[1] (iD) and Glen Whitman[2]

[1]New York University, USA and [2]California State University, USA
Corresponding author: Mario J. Rizzo, email: mario.rizzo@nyu.edu

Abstract

Daniel Kahneman's legacy is best understood in light of developments in economic theory in the early-mid-20th century, when economists were eager to put utility functions on a firm mathematical foundation. The axiomatic system that provided this foundation was not originally intended to be normative in a prescriptive sense but later came to be seen that way. Kahneman took the axioms seriously, tested them for descriptive accuracy, and found them wanting. He did not view the axioms as necessarily prescriptive. Nevertheless, in the research program he conceived, factual discoveries about real decision-making were stated as deviations from the axioms and thus deemed 'errors'. This was an unfortunate turn that needs to be corrected for the psychological enrichment of economics to proceed in a productive direction.

Introduction

To grasp the significance of Daniel Kahneman's contributions to economics, we must go back in time to the early 20th century, particularly the 1930s–1950s. This was a period of increasing theoretical formalism and mathematization, driven in large part by the desire to make economics as scientific as the physical sciences. It was also the period when the word 'rational' acquired its now most popular meaning within the discipline.

Before this period, the word 'rational' was not commonly used by economists and had no precise meaning. The idea that people generally tried to advance their condition as they perceived it, given their best understanding of the world, was surely present. But classical economists such as Smith and Ricardo rarely used the word 'rational', instead saying people sought 'betterment of condition' or 'greatest advantage' (McKenzie, 2010: 161). Even Marshall, writing at the beginning of the 20th century, did not use the term 'rational' (ibid.: 175). This began to change with the marginal revolution, which, in relying on the marginal valuations of individual consumers, encouraged the building of mathematical models on an *explicit* foundation of individual choice. Jevons and Walras's models both included the maximization of individual utility functions (Jevons, 1888; Walras, 2014 [1874]). But where did these utility functions come from?

The quest for the foundation of utility functions

That question motivated a quest to find assumptions that would justify such functions. In 1944, Von Neumann and Morgenstern posited a set of axioms that together guaranteed the existence of utility functions in the context of choice under uncertainty (understood as risk) (Von Neumann and Morgenstern, 1953 [1944]: 26–27). Two of those axioms were completeness and transitivity.[1] Both axioms embody the idea of internal consistency, that is, that one's preferences ought not 'contradict' one another. Together, they guarantee a top-to-bottom ranking of all possible objects of choice. Von Neumann and Morgenstern used the term 'rational' for these axioms; to our knowledge, they were the first to do so. However, they did not treat this as a *unique* definition of rationality, but simply as one that suited their analytical purposes.

Debreu brought the axiomatic approach into choice theory generally. In 1954, he proved that three axioms – completeness, transitivity, and continuity – guarantee the existence of a continuous utility function (Debreu, 1954). Debreu did not use the word 'rational' to describe these axioms. Later that same year, however, Arrow and Debreu (1954: 269) employed Debreu's conclusion as part of their general proof of the existence of a general competitive equilibrium.[2] Uzawa (1956) and Arrow (1959), in the process of merging the axiomatic approach with the theory of revealed preference, reintroduced the word 'rational' to describe choice functions that satisfy the axioms of transitivity and completeness.

And thus axiomatic rationality was born: from a desire to provide a logical foundation for utility functions in economic models. It was about mathematical tractability. It was about constructing models that (it was hoped) would describe and predict the operation of a market economy. For predictive purposes, and superficially explanatory ones, it did not matter whether anyone *really* had preferences that satisfied the axioms. All that mattered was whether the resulting theories' predictions passed empirical muster. And thus, the doctrine of 'as if'-ism was born at the same time.

Critically, axiomatic rationality was not a normative (i.e., prescriptive) project at this point. Before explaining further, we must clarify a terminological confusion. The word 'normative' has multiple meanings. Usually it refers to judgments of value: good/bad, right/wrong, should/should not. In this sense, normative is distinguished from descriptive or positive. However, 'normative' is also sometimes used in a quasi-descriptive sense to mean a standard or criterion for the behavior of an idealized agent within a well-specified model. In this sense, it is not directly applicable to changing or correcting the behavior of real-world, non-idealized individuals. For example, Luce and Raiffa (1957: 63) clarify that the axiomatic form of game theory is:

> … not *descriptive*, but rather (conditionally) *normative*. It states neither how people do behave nor how they should behave in an absolute sense, but how they should behave if they wish to achieve certain ends. It prescribes for given

---

[1] Von Neumann and Morgenstern did not use these exact terms, but the properties we have described here appear as (3:A:a) and (3:A:b), respectively.

[2] For a more thorough version of this history, see Rizzo and Whitman (2020: 52–55.)

assumptions courses of action for the attainment of outcomes having certain formal 'optimum' properties. These properties may or may not be deemed pertinent in any given real world conflict of interest.

Hereafter, we will use 'normative' in its broader prescriptive sense. Our point is that economists of this era had not introduced axiomatic rationality for normative (prescriptive) purposes.[3]

Prominent economists of the period, including Von Neumann and Morgenstern themselves, questioned axiomatic rationality's descriptive accuracy as well as its normative import and offered persuasive reasons why reasonable people might not satisfy its requirements (Rizzo and Whitman, 2020: 55). Chief among these was that normal people with limited time and cognitive resources would not find it worthwhile to conduct a comprehensive audit of all their preferences for perfect consistency, as doing so would fail a reasonable cost-benefit test.

Nevertheless, the word 'rational' itself has a tone of approval. Its opposite, 'irrational', is typically an insult. So we should not be surprised that something labeled rational would eventually start to seem like 'a good thing'.

Furthermore, while the principal goal was positive modeling, it was widely known that general competitive equilibrium had potential normative implications. The famous first and second welfare theorems were already understood by this time: any competitive equilibrium was also Pareto-optimal, and any Pareto-optimal allocation could be the outcome of a competitive equilibrium (Arrow and Debreu, 1954: 265). To the extent that we normatively endorse Pareto optimality, therefore, competitive equilibrium would seem like an attractive arrangement. However, and crucially, Pareto optimality's attractiveness as a norm is contingent on defining 'betterness' in terms of the actual well-being of the people involved. Yet we could imagine a world that is Pareto-optimal in terms of as-if utility functions that accurately describe *behavior* but do not describe what people *actually* want in a subjective sense. But for the welfare theorems to have normative weight, the utility functions involved also needed to express the actual satisfaction of subjective preferences. What economists did not fully realize is that this required the axioms to be both descriptive and normative. This is the inexorable logic of the theorems. And so, economists who wanted the welfare theorems to matter ended up assuming exactly that, even if they never (or rarely) said so overtly.

Unlike their initial treatment of preferences, economists' treatment of beliefs was never clearly limited to positive purposes. Classical logic, the rules of probability, and Bayes' Theorem were assigned normative status as well. As with preferences, this was an axiomatic system embodying the presumption that consistency is necessary and worthwhile, not merely in theory but in practical life. The possibility that rational people with limited time and cognitive resources would *not* wish to root out all possible inconsistencies in an entire system of beliefs – or that beliefs might serve purposes other than truth-tracking, or that other forms of belief formation and information processing might be viable or even preferable – received little attention.

---

[3]In 1988, the distinction between normative and prescriptive was clarified by Raiffa and coauthors. See Bell *et al.* (1988: 16–18).

Thus, in the domain of both preferences and beliefs, the mid-20th century witnessed the elevation of axiomatic rationality from mathematical construct to normative ideal. The same models were used to describe both how people *do* behave and how they *should*. And that is where things stood when Kahneman arrived.

Putting axioms to the test

Kahneman and his coauthors took the axioms seriously – not as undeniable truths, but as testable hypotheses about human behavior. Kahneman was not content to test the outputs of theoretical models (e.g., does a higher price lead to a lower quantity demanded?); instead, he tested the inputs (e.g., do people consistently rank two objects of choice in the same way or correctly apply Bayes' Theorem?). By this method, Kahneman and his coauthors amassed considerable evidence that the axioms were descriptively wrong. Kahneman's experiments threw a bucket of cold water on the 'as-if' approach, because people *did not* behave as if their preference rankings were complete and transitive, and their inferences *did not* follow the strict dictates of logic and probability.

Kahneman's work put the profession at a crossroads. The first path would have involved expanding the notion of rationality beyond the axiomatic straitjacket, remembering that those axioms had never had a strong normative basis to begin with. The second path involved rejecting axiomatic rationality for descriptive purposes but maintaining it for normative ones. The profession largely took the latter (wrong) path.[4] As a result, wide swaths of reasonable and understandable human behaviors were tarred as irrational, problematic, and even pathological.

Among other problems, the maintenance of axiomatic rationality as a normative standard set the stage for the emergence of behavioral paternalism: the use of behavioral findings to justify policy interventions 'for people's own good', even when there is no traditional (i.e., interpersonal) market failure involved. There is a litany of objections here, foremost among them the *non sequitur* of resolving inconsistent preference rankings (i.e., someone seems to prefer *both* A to B *and* B to A, depending on the framing or other supposedly irrelevant factors) in favor of one such ranking, without much basis other than the observer's judgment for choosing which ranking to treat as someone's 'true' preference (Rizzo and Whitman, 2020: 75–78).

But is Kahneman to blame for this wrong turn in the discipline? Unlike the majority of behavioral economists, Kahneman adopted a broader view of normative rationality than the technical rationality axioms would allow. First, he argued the axioms are too restrictive (Herfeld, 2014: 3). Since they cannot be satisfied by real human beings, they cannot be normative. *Ought* implies *can*, after all. Secondly,

---

[4]There is also a third path. Because the axiomatic approach does not specify the objects of choice, it is possible to use a strategy of *redescription* to force all preferences into the axiomatic frame. For example, intransitive preferences (A is preferred to B, B to C, and C to A) can be made transitive by defining A, B, and C more specifically: there is A-when-compared-to-B, which is different from A-when-compared-to-C, and these can be treated as separate objects instead of as a single object A. This approach is analytically possible, albeit mathematically awkward, but it raises questions of when and why it should be allowed. In any case, our concerns about it are beyond the scope of this article (see Rizzo and Whitman, 2020: 69–75).

Kahneman said the axiomatic requirements are too loose because the pursuit of any ends whatsoever can be rational in the technical, instrumental sense without being genuinely rational. For example, he said addiction to drugs cannot be rational in part because people will come to regret such a choice (Ibid: 3). Instead, Kahneman thought it was better to adopt an evaluative standard of 'reasonableness'. Although admittedly imprecise, this reasonableness criterion would embody those ends which the individual is unlikely to regret and those means which are most adequate to the task.[5]

On the question of the axioms' restrictiveness, Kahneman was right. On regret, we think not. The problem with a *no-regret* standard of reasonableness is that regret has no unique meaning. A person may 'regret' that a risky choice did not turn out well. A person may willingly accept that he will 'regret' some action tomorrow, but think it is worth doing anyway; he would do it again under the same circumstances. A person may 'regret' that she did not pay enough attention to unfavorable information that was available about her action. The first two of these are fully consistent with reasonable behavior. Only the last seems to be what Kahneman had in mind. We suspect, however, that even the last type of regret could play an important role in learning processes.[6]

Setting aside these concerns about regret, Kahneman evidently had little reverence for the axiomatic approach, rightly regarding it as deficient. Despite his more recent explicit observations on normativity – and his flirtation with a hedonic welfare standard ('happiness') in earlier work (Kahneman *et al.*, 1997) – Kahneman's main interest was always the discovery of novel *facts* about decision-making (Herfeld, 2014: 6). This was a worthy project that he pursued with alacrity. But what theory did he use to guide his discoveries? The standard neoclassical theory of rational choice. His *facts* are all identified relative to its axioms; they are essentially amendments to the standard theory.

Paths taken and not taken

One natural approach would have been to first understand why heuristics work – and only then try to explain why and under which conditions they do not. Unfortunately, Kahneman's laudable descriptive or positive exercise was compromised by calling the new facts 'errors' relative to the axiomatic benchmark. To call them such endowed them with negative connotations based simply on their deviation from a theory he said was impossible for human beings to satisfy. The temptation is nigh irresistible; an industry of doctor economists has now emerged to cure people of their rational perversities.

The framing of deviations from axiomatic rationality as 'errors' has allowed unjustified and inadvertent prescriptivism to creep into what, in principle, could have been a purely descriptive endeavor. We will offer three brief examples.

---

[5]Kahneman says that he never doubted that the standard rationality axioms are 'normative' (Herfeld, 2014: 18). What he seems to intend here is normativity in the quasi-descriptive sense mentioned earlier. The axioms are normative in that they specify the internal logic of a system or idealized agents within the system. But they are not normative in the prescriptive sense that real people should attempt to satisfy them or that they should be somehow incentivized to behave consistently with them.

[6]We are also skeptical about the potential for analysts (much fewer policymakers) to distinguish between these different sorts of regret.

First, behavioral economists have typically studied only one bias at a time, a concern Kahneman himself pointed out (Herfeld, 2014: 9). Analyzing more than one did not seem tractable. However, biases move in various directions and to various degrees. What is the overall effect? Clearly, some biases can offset (or magnify) others (Besharov, 2004). Characterizing one bias in isolation as an 'error' is an analytical mistake when biases interact; a system-level approach is called for. Nevertheless, behavioral economists have used lone biases to justify intervention; that is the normative impulse at work. This is what comes of taking axiomatic rationality as a 'benchmark': every bias taken in isolation is *ipso facto* seen as an error.

Second, the axiomatic benchmark privileges the analyst's or experimenter's perspective over that of the people involved. The famous Linda problem is an important example of this phenomenon. In this experiment, which has been repeated with variations many times, participants are given a description of 'Linda', which suggests that she is a 'liberal' with progressive views on many subjects. And then they are asked a question: which is more probable, (1) that Linda is a bank teller or (2) that Linda is a bank teller active in the feminist movement? The original modal result (Linda is a bank teller active in the feminist movement) was deemed an error in probability theory because the probability of a subset can never be greater than the probability of the full set. There have been many criticisms of this experiment, as well as many variations that change (and typically reduce) the proportion of people making the so-called error. The most important objections are these: First, people may understand the term 'probability' differently from the experiment designer, which means they are simply answering a different question. Subjects are deemed deficient in understanding the language in a way that does not conform to the experimenter's usage. Second, the information provided in these experiments is strictly irrelevant, thereby violating standard norms of cooperative communication (Grice, 1989). The 'correct' answer would have been the same if no information at all had been provided. Subjects are thus deemed deficient for assuming that the experimenters were being cooperative rather than purposely misleading in giving that information. And just how important in daily life is this purported failure to understand probability theory, anyway?

Third, the axiomatic benchmark creates the false impression that there exists a unique normative standard for belief formation. Kahneman and Tversky, following the neoclassical economists before them, suggested that everyone should update their beliefs according to Bayes's Theorem. That theorem is, of course, entirely correct in its proper place and properly applied. However, it is strictly applicable to decision-making under risk, not uncertainty – and subsequent research has shown that non-Bayesian methods can perform about as well or even better than Bayes in situations of substantial uncertainty (Todd and Goodie, 2002; Juslin *et al.*, 2009). Furthermore, Kahneman and Tversky (among many others) simply assumed that prior probabilities must be equal to the given base rates – which is *not* required by Bayes' Theorem. Since subjects were not given full information about problem situations wherein the theorem was supposed to be applied, they were free (implicitly) to make assumptions based on reasonable guesses and their own experience. These assumptions would naturally affect prior probabilities. Lastly, and remarkably, there was no scope for learning over time in the typical experimental setup. People

were simply given data and expected to process it via a tautological relation (i.e., Bayes's Theorem). We know that Bayesian learning is far more likely to be displayed when there is feedback and revision (Gigerenzer, 2023: 64).

Again, these are only examples. The larger moral is this: the treatment of axiomatic rationality as a benchmark, even for purely descriptive purposes, tempts analysts into making unjustified normative conclusions. The axiomatic benchmark further creates a blindspot to the ways that real people can be rational in a more inclusive sense (Rizzo and Whitman, 2020).

Many friends of Daniel Kahneman report that he liked to discuss his ideas with people who did not necessarily agree with him. He enjoyed re-evaluating his own ideas. So we do not think that our criticisms are amiss; we hope he would have welcomed our criticisms and suggestions. Kahneman's research program has been enormously valuable, opening questions among economists that were long suppressed by conventional wisdom. The challenge now is to purge the program he pioneered of its unjustified elements – especially the remnants of an axiomatic approach that Kahneman himself questioned – so that the enrichment of economics with psychology can move forward with its best parts intact.

References

Arrow, K. J. (1959), 'Rational choice functions and orderings', *Economica*, **26**(102): 121–127.

Arrow, K. J. and G. Debreu (1954), 'Existence of an equilibrium for a competitive economy', *Econometrica: Journal of the Econometric Society*, **22**(3): 265–290.

Bell, D. E., H. Raiffa and A. Tversky (1988), 'Descriptive, normative, and prescriptive interactions in decision making', *Decision Making: Descriptive, Normative, and Prescriptive Interactions*, **1**: 9–32.

Besharov, G. (2004), 'Second-best considerations in correcting cognitive biases', *Southern Economic Journal*, **71**(1): 12–20.

Debreu, G. (1954), 'Representation of a preference ordering by a numerical function', *Decision Processes*, **3**: 159–165.

Gigerenzer, G. (2023), *The Intelligence of Intuition*, Cambridge, UK: Cambridge University Press.

Grice, H. P. (1989), *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.

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

Jevons, W. S. (1888), *The Theory of Political Economy*, 3rd edn, London: Macmillan and Co, https://oll.libertyfund.org/titles/jevons-the-theory-of-political-economy

Juslin, P., H. Nilsson and A. Winman (2009), 'Probability theory, not the very guide of life', *Psychological Review*, **116**(4): 856–874.

Kahneman, D., P. P. Wakker and R. Sarin (1997), 'Back to Bentham? Explorations of experienced utility', *The Quarterly Journal of Economics*, **112**(2): 375–406.

Luce, R. D. and Raiffa H. (1957), *Games and Decisions*. New York: John Wiley & Sons, Inc.

McKenzie, R. B. (2010), 'Predictably Rational?', in *Search of Defenses for Rational Behavior in Economics*. Berlin Heidelberg: Springer, 1. Kindle Edition.

Rizzo, M. J. and G. Whitman (2020), *Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy*, Cambridge, UK: Cambridge University Press.

Todd, P. M. and A. S. Goodie (2002), 'Testing the Ecological Rationality of Base Rate neglect', in B. Hallam, D. Floreano, J. Hallam, G. Hayes, and J. A. Meyer (eds), *From Animals to Animals 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, Cambridge, MA: MIT Press, 215–223.

Uzawa, H. (1956), 'Note on preference and axioms of choice', *Annals of the Institute of Statistical Mathematics*, **8**(1): 35–40.

von Neumann, J. and O. Morgenstern (1953 [1944]), *Theory of Games and Economic Behavior*, 3rd edn, Princeton, NJ: Princeton University Press.

Walras, L. (2014 [1874]), *Elements of Theoretical Economics: Or, the Theory of Social Wealth, D.A. Walker and Jan van Daal*, Trans, Cambridge, UK: Cambridge University Press, https://www.google.com/books/edition/L%C3%A9on_Walras_Elements_of_Theoretical_Eco/Srq1BAAAQBAJ?hl=en&gbpv=1