Extending Behavioral Economics' Methodological Critique of Rational Choice Theory to Include Counterfactual Reasoning

Abstract: This paper extends behavioral economics' realist methodological critique of rational choice theory to include the type of logical reasoning underlying its axiomatic foundations. A purely realist critique ignores Kahneman's emphasis on how the theory's axiomatic foundations make it normative. I extend his critique to the theory's reliance on classical logic, which excludes the concept of possibility employed in counterfactual reasoning. Nudge theory reflects this in employing counterfactual conditionals. This answers the complaint that the *Homo sapiens* agent conception ultimately reduces to a *Homo economicus* conception, and also provides grounds for treating *Homo sapiens* as an adaptive, non-optimizing, reflexive agent.

Keywords: problem of deduction, modal reasoning, counterfactual conditionals, nudge, reflexive agents

JEL codes: A12, B41, D90, J26

Behavioral economics' critique of neoclassical economics' rational choice theory concerns not only the substantive issue of whether people behave rationally, but also important methodological issues. Indeed, the criticism many behavioral economists make of rational choice theory is that it is unrealistic (e.g., Camerer and Loewenstein, 2004; Thaler 2016). Behavioral economics' success

on this view is a result of its being better supported by empirical evidence about choice behavior. At the same time, while the axiomatic foundations of rational choice theory are seen as a cause of its being unrealistic, this methodological critique *per se* leaves unexamined the nature of the deductive reasoning employed in rational choice theory as if its effects were methodologically neutral.

This paper argues that its effects are not neutral, and that it is important to see how it creates a bias in favor of that theory and against behavioral economics. To show this, I first argue that Daniel Kahneman's methodological critique of rational choice theory as normative goes beyond the familiar argument that rational choice theory is insufficiently grounded in the evidence regarding choice, and depends on how the axiomatic character of that theory restricts the kinds of information we include in our theories of choice behavior. I then use classical logic's famous problem of deduction to identify what is problematic about traditional logical reasoning in explanations of choice behavior, and go on to show that a different type of logical reasoning associated with counterfactual conditionals, modal concepts, and natural language plays a role in behavioral economics. The paper concludes that a behavioral economics' methodological critique of neoclassical rational choice theory not only emphasizes realism and the importance of observation but also the type of logical reasoning we employ in developing theories of behavior, and argues that this is central to developing an alternative conception of economic agents as adaptive and reflexive rather than optimizing.

Kahneman's methodological critique

In his Nobel lecture, Daniel Kahneman drew an important methodological distinction between prospect theory and rational choice theory as follows:

> One novelty of prospect theory was that it was explicitly presented as a formal descriptive theory of the choices that people actually make, not as a normative model. This was a departure from a long history of choice models that served double duty as normative logics and as idealized descriptive models (2003, p. 1456).

For Kahneman, rational choice theory is "normative" in the sense that it axiomatically postulates a set of assumptions, especially in regard to the nature of preferences (completeness, transitivity, and independence of irrelevant alternatives), that have the effect of defining what rational behavior involves. As a top-down methodological approach that gives priority to logical foundations, it does this by determining what sort of evidence counts in explanations of choice

behavior, and in this way produces "idealized descriptive models." In contrast, prospect theory advances what Kahneman calls a "formal descriptive theory" that instead begins in a bottom-up way with observational evidence regarding "the choices that people actually make" and then formalizes its theories on the basis of that empirical foundation.

Kahneman's critique of rational choice theory as employing a top-down methodology, then, is that it gets the fit between theory and evidence wrong because it places the weight on the former rather than the latter. But why is this specifically normative? Why not simply say that rational choice theory is too top-down and insufficiently empirical? Consider, then, what the contrast between an axiomatic, top-down type of theory and a more empirical bottom-up type of theory involves in regard to the information that each allows.

Axiomatic reasoning deduces a set of inferences from a set of axioms or assumptions, and if the reasoning involved is valid, then according to the rules of classical formal logic the information appearing in those inferences must already be contained in those axioms. Indeed, one of the reasons that a valid deductive inference is a necessary inference is that what is inferred must follow fully from what is assumed and nothing else has been introduced in the inference. A top-down axiomatic theory is normative, then, in that it limits what information that theory admits to the information its axioms contain. For example, in axiomatic choice theory preferences are assumed to be transitive, so any information from the world showing that preferences are intransitive must appear as an anomaly, and consequently can be disregarded as irrelevant to that theory.

In contrast, a bottom-up methodology that places weight on evidence operates inductively in allowing any new information that observation might produce to be worked into theory – Kahneman's idea of a "formal descriptive theory." Its information base is not limited by a pre-given set of axioms or assumptions but is open to whatever observation of the world produces. The rules of inductive inference also play a regulatory role regarding what conclusions can be drawn from evidence, namely, there must be sufficient evidence of a relationship to claim it exists, but by comparison these rules are open in that new information can always be added to an inductive inference.

A top-down axiomatic theory type of theory, consequently, has a 'self-fulfilling' character in that the evidence it allows must always confirm its assumptions. This is important for understanding why a realist critique of rational choice theory is incomplete and potentially ineffective. The claim that only behavioral economics is evidentially grounded is not persuasive to defenders of rational choice theory who reject evidence inconsistent with its axiomatic foundations, and point instead

to evidence supporting rational choice theory – their own "idealized descriptive" view of the evidence.

Indeed, Kahneman's view that people should behave rationally were they not subject to the behavioral biases we observe betrays an ambivalent understanding of the role of axiomatic reasoning plays in decision theory, and opens the door to the defenders of rational choice theory. I suggest, then, that the full force of the realist critique rests not only on the theory-evidence fit issue but also on the relationship between the nature of logical reasoning and allowable information. Understanding this further requires we look more closely at the nature of deductive argument and the classical logic's paradoxical problem of deduction.

Classical logic's problem of deduction

Deductive reasoning in classical logic formally derives conclusions from premises by well-established rules of valid inference, such as the *modus ponens* 'if-then' conditional (if p implies q and p is true, then q is true). Deductive reasoning and inductive reasoning are our two main forms of explanation, and since inductive reasoning, when it is justified, provide new information about the world, it seems that deductive reasoning as the companion to inductive reasoning should also provide new information about the world. Yet that the conclusions to deductive arguments only contain information already present their premises is contrary to this. Logically speaking, deductive arguments are tautological or circular, and the truths they contain are only truths by definition. At the same time, many of the results of complicated deductive arguments, such as in mathematics, are non-trivial and accordingly appear to tell us something new, so we have a paradoxical situation.[1]

The problem of deduction is a long-standing, much discussed problem in the history of philosophy, and here I only aim to identify one general response to it on the part of philosophers that I will argue below bears on the nature of the axiomatic reasoning in rational choice theory. That general response, then, is to argue that the problem of deduction derives from how classical logic limits our representation of how people actually reason deductively and the information they possess and rely on in doing so: specifically it excludes what are called modal concepts, especially the concept of possibility, that express ideas about what might or could be the case. If p then q

---

[1] "If in an inference the conclusion is not contained in the premises, it cannot be valid; and if the conclusion is not different from the premises, it is useless; but the conclusion cannot be contained in the premises and also possess novelty; hence inferences cannot be both valid and useful" (Cohen and Nagel, 1934/62/93), pp. 173-176).

guarantees the truth of q if p is true. But we are often uncertain about what we think is true and reason in terms of contingencies.

The standard argument, then, is that every deduction has behind it a set of possibilities that have been eliminated, and that, our knowing this, allows us to confidently say that the inference could not have been otherwise. Yet these possibilities remain latent in that inference – latent because they have been eliminated – while as such the information they possess cannot be present in the inference itself. The inference thus appears not to convey any new information, but in fact does in the form of the possibilities that are ruled out. An example of this would be if you assume it is true that firms are perfectly efficient (agents behave in a specific manner), and infer by *modus ponens* what must then be true, assuming away the possibility of x-inefficiency (a point made by Leibenstein). You thus have a biased and mis-specified but logically consistent model.

What grounds might there be, then, for giving up traditional logical thinking and reasoning in terms of contingency? One good reason is that this is the way that natural languages work, which rely heavily on modal concepts and use expressions such as 'what might' or 'could be' the case, as reflected in their distinction between indicative and subjunctive moods.[2] Natural languages are thus rich in information. Classical logic, such as the *modus ponens* principle, on the surface excludes possibilities eliminated as a direct element in producing a deduction. Indeed, it might be said that its success in explaining valid inference trades off against and comes at the price of suppressing information that we use when we think in terms of possibilities. This response to the classical logic's problem of deduction, then, essentially argues that the information base people employ in deductive reasoning is implicitly broader than it appears, and more like what we observe in natural languages, so that deductions are actually information producing.[3]

An important implication of this (which bears on the idea of nudges) is that, in addition to the standard view of conditional statements associated with the *modus ponens* principle, there exists a whole set of counterfactual conditionals about which we commonly reason. The standard *modus ponens* inference involves an 'if-then' reasoning where the 'if' part (or p) refers to a statement that is true in order for that conditional inference to be true. In the case of counterfactual conditionals, however, we often infer a true 'then' statement when the 'if' in question is a possibility that could

---

[2] The indicative mood is used to say what is/was/will be the case whereas the subjunctive mood is used to say what could, would, should, may, might, etc. be the case.

[3] For example, the philosopher Hintikka argued that we might solve the problem of deduction if we distinguish between 'surface' information and 'depth' information since this would allow us to refer to information in a set of premises not explicitly appearing in them as part of the inference involved. Then valid inference would convey new information relative to that 'depth' information (Hintikka, 1973).

have been true but is not true as a matter of fact. For example, many people would say that the statement 'If Clinton had won the state of Wisconsin in the 2016 U.S. presidential, she would have won that election' is a true statement. The 'if' part of this inference is contrary to fact, since Clinton did not win the state of Wisconsin, but it was nonetheless possible that Clinton might have won Wisconsin. The truth of the inference, then, depends on our understanding that possibility. The type of non-standard logical reasoning used to evaluate counterfactual conditionals of this kind is broadly called modal logic, and expressions such as 'it is possible that' which it investigates are referred to as modal operators.[4]

However, my goal here is not to review modern developments in the philosophy of logic, but rather to simply show in the next section that behavioral economics' nudge thinking shares the ideas motivating these developments, and that this provides a further foundation to its methodological critique of rational choice theory.

Behavioral economics' analysis of nudges

In *Nudge* (2008), the justification for choice architects intervening in people's choices is that this would "make choosers better off, *as judged by themselves*" (Thaler and Sunstein, 2008, p. 5; their emphasis). The emphasis on what choosers themselves judge would make them better off provides a response to the criticism that such interventions are paternalist and limit individuals' freedom. If nudges produce outcomes that people would want were they informed about their consequences, then they constitute outcomes they would freely choose in such circumstances.

Key to this argument, then, is the language of 'would' and 'were.' In effect, the other side of a world of bounded rationality are the nudges designed to achieve unrealized possibilities by expanding the information individuals have about their choices. Nudges, accordingly, can be formulated in terms of counterfactual conditionals – 'if-then' relationships with matters contrary to fact that are nonetheless possible. For example, when individuals fail to invest in retirement plans they would prefer had they better information, investing in those preferred investments is both contrary to the facts and was yet still possible. Like in the Clinton example above, we would say that the (true) counterfactual conditional in this case is: 'if they had possessed better information regarding their retirement options and invested in retirement plans they preferred, they would have been better off.' Being more knowledgeable about their retirement options, then,

---

[4] Modal logic is a logic of modal operators such as 'it is possible that' or 'it is necessary that' influentially developed by the philosopher Saul Kripke (Kripke, 1980).

is a matter of their reasoning in terms of the space of possibilities they face. This indeed involves a different emphasis than much of the nudge literature, which is formulated in such a way as to work 'behind the backs' of those affected by nudges.

On this reading, behavioral economics' methodological critique of rational choice theory is not just that it is unrealistic because its predictions are not well supported by observations of choice behavior. It is also unrealistic because its axiomatic logical foundations prevent it from explaining the richer informational base choice behavior involves. This limitation walls off observational evidence contrary to its rational choice predictions, and thereby advances an explanation of choice behavior that fails to capture a fundamental part of how people reason about their choices. As natural languages show, people commonly reason counterfactually in terms of what could have happened but often did not in order to understand the possibilities they face in making choices. In contrast, the axiomatic foundations of rational choice theory require that choice be modeled as determinate and closed to the realm of possibility – thus justifying its optimization interpretation of behavior.

Broadening the logical and information basis of choice theory has another consequence. On many interpretations of behavioral economics, it is argued that *Homo sapiens* agents ultimately reduce to *Homo economicus* agents because, should nudges eliminate bounded rationality, then *Homo sapiens* agents would be indistinguishable from a *Homo economicus* agents. In effect, 'a *Homo economicus* wolf lurks in every *Homo sapiens* sheep's clothing.' Yet on the argument here this identification fails, since the characteristics of the *Homo economicus* depend solely on assumptions about preferences attributed to individuals, whereas the characteristics of a *Homo sapiens* depend upon individuals' grasp of the complex possibilities operating in the world they occupy, where this includes the nature of social relationships and all the possible ways that they can change and evolve. Indeed, social context matters to people's choices, but social context is also fluid and in need of constant interpretation.

In the original emphasis on nudges as effectively occurring behind the backs of agents (Thaler and Sunstein, 2008), with the rational *Homo economicus* lurking in the wings, it is easy to retreat to the familiar world of classical logic with its highly determinate (optimal) outcomes. Cass Sunstein's subsequent *Choosing Not to Choose* book (2015) takes a different view. Moving away from nudges, he argues that people intentionally pass responsibility for many of their choices to a multitude of others (doctors, accountants, stores, schools, etc.) by adopting default rules regarding certain types of choices they wish to delegate to others ('in such-and-such circumstances, I will rely on the expertise of *Y*'). Implicitly, that is, Sunstein assumes that people

recognize the complex nature of the world they face, conclude they cannot anticipate all possibilities, and feely elect others expert in the possibilities they may encounter as their representatives. This more social vision of behavior contrasts sharply with the original nudge conception rooted ultimately in the atomistic *Homo economicus*.[5]

Reflexive non-optimizing economic agents

Here I offer brief final comments regarding what the discussion here contributes to our thinking about the nature of economic agents. As a point of entry, recall that Herbert Simon argued that "Human rational behavior … is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor" (Simon 1990, p. 7). This suggests that adaptation rather than optimization best characterizes people's choice behavior. Yet rational choice theorists can still reply that whatever environment they find themselves in, agents must still optimize, so it is not immediately clear why we should think adaptation should characterize choice behavior rather than optimization.

I have argued, then, that optimization behavior is a product of neoclassicism's axiomatic analysis of choice that makes possibility irrelevant to behavior. This means that the environment agents find themselves in must also always be determinate, which would indeed make it possible for agents to optimize. However, if we reason instead in natural language terms about choice behavior, then the environment that agents find themselves in should be seen as exhibiting a spectrum of determinable and indeterminable possibilities, rather than always determinate (if stochastic) states of affairs. In effect, there are no unique deterministic equilibrium states in a world that is open to a range of possibility. This then means that agents' choices need to be seen as adjustable or revisable, allowing them to continually adapt to how the circumstances of choice evolve. Thus, just as classical deductive reasoning obscures how human thinking operates in terms of possibilities, so rational choice theory's optimization focus obscures how agents reason in terms of a range of possible interactions between "the structure of task environments and the computational capabilities of the actor."

I characterize such agents as reflexive to capture how their choices have feedback mechanisms built into them. Every choice is framed by the range of possibilities agents can imagine, and as

---

[5] An important implication of this is that, contrary to J.S. Mill's famous *On Liberty* view (1859) that individual freedom stems from individual action alone, in a world in which people delegate responsibility for many of their choices to others, their freedom is increased beyond the reach of their own choices (Davis, 2018).

such come with a range of adjustment strategies that feed back into their subsequent choices. Choice is then explained in process terms with interacting forward and backward linkages. In contrast, optimization analysis, as framed by classical logic, essentially occurs outside of time, and thus substitutes an equilibrium account of many agents' optimizing choices for an in-time dynamics of interaction of many adaptive reflexive agents.

This latter subject is what behavioral economic theory should investigate. It moves away from the neoclassical equilibrium set-up to describe how agents evolve and interact in changeable social settings. This involves a bottom-up methodological approach that not only gives greater space to empirical research but also employs a richer logical thinking reflecting agents' counterfactual reasoning. Kahneman's critique of rational choice theory only goes part way in this regard, and its failure to examine the nature and implications of standard logical reasoning that theory employs runs the risk of leaving that theory in place. This paper has attempted to rebuild that methodological critique on stronger foundations.

References

Camerer, Colin and George Loewenstein (2004) "Behavioral Economics: Past, Present and Future," in C. Camerer, G. Loewenstein, and M. Rabin, eds., *Advances in Behavioral Economics*, Princeton: Princeton University Press: 3-51.

Cohen, Morris and Ernest Nagel (1934/62/93) *An Introduction to Logic*, edited and introduced by J. Corcoran, Indianapolis, Cambridge: Hackett Publishing Company.

Davis, John (2018) "Behavioral economics and the positive-normative distinction: Sunstein's *Choosing Not to Choose* and behavioral economics imperialism," *Revue Éthique et Économique* 15 (1): 1-15.

Hintikka, Jaakko (1973) *Logic, Language-Games, and Information*, Oxford, Clarendon Press.

Kahneman, Daniel (2003) "Maps of Bounded Rationality: A Perspective on Intuitive Judgment," *American Economic Review* 93, 1449-1475.

Kripke, Saul (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.

Mill, John Stuart (1859) *On Liberty*, London: John Parker and Son.

Simon, Herbert (1990) "Invariants of human behavior," *Annual Review of Psychology*, 41: 1-19.

Sunstein, Cass (2015) *Choosing Not to Choose: Understanding the Value of Choice*, New York: Oxford University Press.

Thaler, Richard (2016) "Behavioral Economics: Past, Present, and Future," *American Economic Review* 106 (7): 1577-1600.

Thaler, Richard and Cass Sunstein (2008) *Nudge: Improving decisions about health, wealth, and happiness*, New Haven CT: Yale University Press.