

Please cite as: Veit, W. (2024). What to make of intransitive preferences?. Preprint. [Add Link]

Check [www.walterveit.com](http://www.walterveit.com) for citation details once published

# What to make of intransitive preferences?

Walter Veit

Department of Philosophy, University of Reading, UK  
Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität  
München, Germany

## Abstract

The axiom of transitivity has been challenged in economic theorizing for over seventy years. Yet, there does not seem to be any movement in economics towards removing classical rational choice models from introductory microeconomics books. The concept of rationality has similarly been employed in the cognitive sciences and biology, and yet, transitivity has here not only been shown to be violated, but also rationally so. Some economists have thus responded with attempts to develop alternative theories that give up on the axiom of transitivity. In this paper, I argue that there is a conceptual confusion in this debate that rests on the mistaken idea that there is something like the ‘one true theory of rationality’ that can determine axioms like transitivity to be true or false. Instead, I defend a shift towards a pluralism of concepts of rationality as well as models in which transitivity should play a role depending on the purposes of the model at hand.

*Keywords:* idealization, rationality, transitivity, preference, choice, evolution, models

“Shall I say, ‘a rational animal’? No, for then I should have to examine what exactly an animal is, and what ‘rational’ is, and hence, starting with one question, I should stumble into more and more difficult ones.”

Meditation II of *Meditations on First Philosophy*

– René Descartes (2008, p. 25)

## 1 Introduction

When Descartes set out to provide a new metaphysical system for philosophy, he rejected the Aristotelian answer or rather definition of man as the ‘rational animal’ as methodologically flawed. While I share little agreement with Descartes’ metaphilosophy, he rightfully recognized that the question of what it means to be rational is a highly complex one. Aristotle’s motivation behind classifying humans as the ‘rational animal’ was to distinguish humans from other animals. This definition, of course, runs into a number of conceptual and empirical problems – even being mocked by Bertrand Russell:

“Man is a rational animal — so at least I have been told. Throughout a long life I have looked diligently for evidence in favour of this statement, but so far I have not had the good fortune to come across it”

– Bertrand Russell (2009, p. 45)

Naturally, the concept of rationality has been the subject of one of the longest conceptual debates in the history of philosophy. When is an agent rational? Is there a difference between the rationality of human and non-human animals (henceforth animals)? Do rational agent models accurately represent these targets in the real world? If not, can they nevertheless be explanatory? Despite the attention ‘rationality’ has received, only little consensus has emerged. The debate is so vast indeed that no single *Stanford Encyclopedia of Philosophy* article on rationality has even been attempted. There is, however, a large number of articles on preferences, decision-making, utility, practical reason, and instrumental rationality.<sup>1</sup>

---

<sup>1</sup> See Rysiew (2015) for an elegant and brief overview of the conceptual debate.

In this paper, I argue that this scattered picture should be taken serious as a reflection of the *disunified* nature of the cluster of ideas relating to rationality, rather than a mere reflection of the philosophical complexity of the term ‘rationality’. I will argue that a lot of confusion in this debate rests on the mistaken idea common among philosophers (though also economists, psychologists, and biologists) that there is something like the *one true theory* of rationality that we only have to uncover and formalize. Instead, I defend a pluralist view of the concepts of rationality, as well as a pluralist view of rational choice models, where different assumptions can be more or less appropriate depending on the purpose of the model at hand. I will do so by focusing on one of the most controversial subjects in debates on rationality, i.e. whether our choices must be transitive to be rational, i.e. the *axiom of transitivity*. But before I explain this notion in more detail and outline the structure of this paper, let me briefly introduce a distinction due to Alex Kacelnik (2006) that will be useful throughout the rest of this article.

While philosophers qua philosophers can often be overly ambitious in trying to offer accounts that are as general as possible, scientists routinely lament that such attempts can often neither be successful nor useful, due to the particular conceptual and methodological challenges of their disciplines. So perhaps it shouldn’t be surprising it was a behavioural ecologist, who has been incredibly influential for his interdisciplinary work on rational choice in animals combining methods from economics, biology, and psychology, to cast significant doubts on the idea that we can have a single cross-discipline definition of rationality. In an inter-disciplinary edited volume on the question whether animals can be rational, Kacelnik (2006) lamented that there could not be a definite answer to this question because different fields use the term rationality in very distinct ways. To make this clear, he introduced a distinction between what he called PP-Rationality, E-Rationality, and B-Rationality.

Beginning with the first, the PP in PP-Rationality stands for the concept of rationality as used in philosophy and psychology. Here, Kacelnik (2006) argues that philosophers, psychologists, and cognitive scientists are largely interested in the process of reasoning, and whether beliefs are formed in response to appropriate reasons.<sup>2</sup> In opposition, Kacelnik calls E-Rationality the concept of rationality employed in economics. The target here are actions rather than beliefs, and the outcome, rather than the process of deliberation. For economists, actions are rational

---

<sup>2</sup> This concept may require introspective capacities, and may thus surprisingly be applied to non-human animals and AIs (Browning & Veit 2023).

if the maximize expected utility. Furthermore, Kacelnik argues that economists not only emphasize - but built their theory of rationality - on the consistency of choice. While this is perhaps an unfairly simple picture of economic concepts of rationality it will serve us well for the purposes of the present paper. As I mentioned above and indicated with the title of this paper, my concern is the axiom of transitivity, which we can simply define as follows: If a rational agent prefers A over B and B over C, they should prefer A over C. To put it more formally, while making room for indifference:

**(Weak) Transitivity:**  $\text{If } A \succsim B \ \& \ B \succsim C \rightarrow A \succsim C$

Intuitively, this perhaps most fundamental idealizations in economic theorizing might seem like a common-sense criterion for rationality – not only in economics, but also in psychology, philosophy, and biology.<sup>3</sup> Yet, this seemingly innocent assumption has caused a lot of controversy. Many psychologists and behavioural economists have rejected it as an accurate idealization to describe human behaviour. But there has also been opposition to transitivity as a normative standard for behaviour to meet to be considered rational. Indeed, one immediate objection one could raise to Kacelnik’s PP-Rationality, is that philosophers as well as psychologists are very much interested in the rationality of actions, rather than just beliefs. Nevertheless, we could simply expand this concept here to include the process of rational belief formation as well as decision-making. This, however, is already quite the substantial commitment about the nature of rationality and does not reflect the entire spectrum of philosophers. Let me therefore follow Okasha (2018) and abbreviate PP-Rationality as P-Rationality. Unlike Okasha, however, I do not intend this merely as an abbreviation, but a reflection of the narrower conception of rationality within the psychological sciences to focus on a descriptive rather than normative account of rational belief formation and decision-making. Economists, as we shall see, are much closer philosophers than psychologists in their motivation to offer a concept of rationality that is also normative. Finally, B-Rationality describes the rationality concept used in biology as a place-holder for fitness-maximization. Just like for E-Rationality, behaviour is considered rational if it maximizes a quantity, but instead of utility it is fitness (i.e. reproductive value). Indeed, fields like evolutionary game theory make clear how these conceptions can influence each other (Veit 2023c).

---

<sup>3</sup> Unsurprisingly, philosophers have been among those who have criticized the rational choice axiom of transitive preferences early on (Schumm 1987).

As I shall argue in this paper, the conflicts about the status of transitivity for rationality not only reflect different disciplinary goals, but also within-discipline disagreements about the goals of our concepts and models. There is no one correct way of evaluating intransitive preferences and choices. There are parts of economic, and other sciences, where the assumption of transitivity is unproblematic and yields both predictive and explanatory insights, while there are others in which it is misleading. There is no a-priori answer that could help us determine in advance whether this idealization is a good or bad one. Sometimes, the use of this idealization functions as a deliberate misrepresentation of reality for some other purpose, explanatory or otherwise, such as the need to assign utilities to alternative options or to explain an agents choices across a narrow set of options. Worse, economists, cognitive scientists, biologists, and philosophers differ substantially in the reasons and goals for ‘rationality-talk’ even within their own disciplines. I will thus argue here, that we should surrender the idea that a term as polysemous as ‘rationality’ has anything like a one true account that could unify all its different usages. With this throat-clearing out of the way, let me provide a brief outline of the structure of this paper.

## **Outline**

In Section 2, I offer a brief history behind the adoption of transitivity as an axiom of rationality in economics and discuss why transitivity has been so controversial. In Section 3, I will discuss intransitivity observed in animal experiments and debates on the evolution of rational behaviour that cast doubt on the idea that there is a simple answer to the question of whether transitivity should be part of our concept of rationality or not. In Section 4, I draw on the philosophy of science literature on modeling and idealization to argue that the transitivity axiom of rationality cannot simply assessed as being either correct or false. Rather, we should adopt a pragmatic and pluralist stance in which we employ different concepts and models of rationality depending on the goal we are using them for. Lastly, Section 5 summarizes and concludes the discussion.

## **2 Transitive preferences and rationality**

Leaving aside the question of group-rationality and how intransitive group choices can emerge from individually ‘rational’ behavior or vice-versa, I shall offer here a

brief overview of the roles *transitivity* plays in economic theorizing and how it has been defended. I should note, however, that collective entities such as companies can be often usefully treated as individuals that conform to a rational agent model. A similar point applies to much work in contemporary political science that treats nations as individual rational agents, an assumption that has not gone without criticism (Green and Shapiro 1996).<sup>4</sup> What began with Adam Smith (2010) as the study of wealth, quickly became the science of rational choice theory. Many decision and game theorists, especially those working in philosophy, and arguably even the founders of decision theory itself, von Neumann and Morgenstern (1944), argued that it is a normative, rather than descriptive theory of how humans should act.

In one of the most influential monographs on economic methodology, Lionel Robbins (1935) detached economic thinking from psychological welfare considerations and material exchanges. He redefined the discipline more abstractly as “the science which studies human behavior as a relationship between ends and scarce means which have alternative uses” (Robbins 1935, p. 16). This could be considered the birth of microeconomics in its modern sense, i.e. the study of individual choice behavior of economic agents. Others, i.e. many behavioral economists (Camerer 1999; Ashraf et al. 2005; Thaler 2016) and philosophers (Rosenberg 1992, 1994, 1995, 2009; Angner and Loewenstein 2007), see this as an unfortunate mathematization and loss of realism of the discipline. But as economists following Robbins argued: economics is not necessarily about humans or the human domain traditionally seen as markets<sup>5</sup> – it is about the optimization of choice behaviour.

Naturally, this conception of economics has led to an expansion of the proper domain of economics and invited the charge of economics imperialism, i.e. the extension and application of economic methods and models to explain and predict phenomena traditionally viewed beyond the scope of economics (Becker 1976; Stigler 1984; Tullock 1972; Levitt and Dubner 2005; Mäki 2009a). Rational agent models have been used to explain criminal behaviour (Becker 1973, 1974), marriage (Becker 1968), politics (Tullock 1972), and science itself (Diamond 2008). For his work on expanding the bounds of economics and rational choice theory, Chicago economist

---

<sup>4</sup> These models, after all, are fundamentally based on the original one of individual human agents in economics. There are, however, interesting parallels here between such collective human organizations and collective multi-cellular organisms (Okasha 2018; Veit 2019a, 2021a).

<sup>5</sup> In addition, biologists have extended market thinking to develop what they call *biological market theory*. See Noë and Hammerstein (1994, 1995) and Noë et al. (2001).

Becker was eventually awarded the Nobel Memorial Prize in Economic Sciences. In his Nobel lecture, he stated:

“I have intentionally chosen certain topics for my research—such as addiction—to probe the boundaries of rational choice theory. [...] My work may have sometimes assumed too much rationality, but I believe it has been an antidote to the extensive research that does not credit people with enough rationality”

– Gary S. Becker (1993, p. 402)

The charge of economics imperialism against the likes of Becker can be seen in two ways, one of which is to be condemned, the other appreciated. When Becker (1993) argues that social scientists have not taken rationality of humans seriously enough, it would be a stretch to defend the thesis that all human choice behavior corresponds to a demanding set of axioms satisfying both *completeness* and *transitivity*. Behavioral economics is an antidote to this way of doing economics, not as a grand unifying theory of human rationality, but as an alternative methodology that provides a variety of models that explain the anomalies of rational choice theory. If economics is conceived of as a more pluralist discipline with a variety of alternative and complementary models for the same phenomena, there wouldn't be a problem of economic imperialism, since all that is imported is a variety of new tools to formerly distinct disciplines.<sup>6</sup> Perhaps though, the label imperialism is misplaced for the latter approach. Instead, one should see the application of economic theories and models to phenomena in other fields as *economics borrowing*, and only the additional goal of replacing theories of ‘irrationality’ with rational choice models as economics imperialism. With this lesson in mind, let us turn to actual economic modeling practice and how the *axiom of transitivity* is defended.

For the purposes of this paper, Kacelnik's definition of economic rationality as consistency will do well enough. Here, he is not so much drawing his own distinction, but rather using the notion of rationality that rational-choice theorists have defended for decades. This way of thinking about rationality goes beyond Robbins' definition of economics as the study of the optimal achievement of goals under scarcity, i.e. instrumental rationality. With the introduction of expected utility theory (von Neumann and Morgenstern 1944), consistency as transitive orderings

---

<sup>6</sup> Thaler (2016) and Rodrik (2015) offer similar conciliatory words.

among preferences became a necessary axiom to calculate utility. Von Neumann and Morgenstern's theorem assumes probability distributions to be given over the outcomes of actions. Their theorem shows that we can only assign utilities if an agent's preferences conform to the axioms of rational choice theory. Because we often do not know the objective probabilities over outcomes, Savage (1954) developed a highly influential theory of 'subjective' probability that was subsequently adopted and used to calculate subjective expected utility. The axiom of transitivity plays therefore a necessary role in much of economic theorizing and has been defended as a necessary idealization. Critics on the economics side have attempted to develop more realistic alternatives such as bounded rationality (see Herbert Simon 1955, 1972, 1991, 1997) that is in line with research in behavioral economics. Despite the development of alternatives, however, most of contemporary rational-choice models, whether normative or descriptive continue to rely on the transitivity of preferences. But as already pointed out, it is not my goal here to defend one account over another. Indeed, as the following discussion will illustrate – I will argue these methodological discussions to rest on outdated views in the philosophy of science.

Transitivity of preferences is at the very center of methodological debates about rational choice theory. Much empirical evidence, however, has accumulated showing that the assumption of transitive preference orderings lacks real-world evidence.<sup>7</sup> Economic models that make use of transitive preference orderings frequently fail to make accurate predictions about the choice behavior of humans. Unfortunately, however, many of these economic models are reliant on this assumption, without which it would not be possible to move from preferences to utility. Due to considerations of space, I leave the question open here of *what* preferences are. It would be a mistake, however, to think that psychological approaches to economics are all in support of a mentalistic interpretation. The phenomenon of rationalization in psychology, i.e. the retrospective attribution of hidden beliefs and desires to oneself, could support a behaviorist interpretation of preferences (see Veit et al. 2020). If the 'behaviorist' interpretation of preferences is correct, E-Rationality and B-Rationality would move closer together. If unification is the goal, however, there is strong case to be made for a preference account based on Daniel Dennett's (1989) *intentional stance*, which attributes beliefs and desires to systems to predict and explain their behaviour as those of a rational agent. This idea

---

<sup>7</sup> See Sen 1969, 1970, 1971, 1977; Grether and Plott 1979; Suzumura 1983; Korhonen et al. 1990; Bradbury and Ross 1990; Fishburn 1991 for a number for important criticisms and proposed alternatives.



has subsequently been developed by Don Ross (2005, 2014) for the purposes of economics. I have sympathies for this ambitious account, as unlike anything offered in the literature so far, it has at least some potential to unify all three accounts of rationality. In a recent work with others, Don Ross has attempted to develop the idea of a “quantitative intentional stance”, as a truly economic, rather than merely philosophical, account of preferences as constructions (see Alekseev et al. 2019). Intransitive preferences could then (at least to some extent) be explained away as mere ‘noise’.

Some economists have proposed alternatives that seek to maintain something close to ‘quasi-transitivity’ (Sen 1969; Panda 2018) in order to improve the realism in their models. Others have defended the transitivity assumption as a normative principle, rather than an empirical one – but even this assumption has been challenged by many philosophers and economists. These debates are notably absent from most economic textbooks (with the exception of behavioral economics). Anand (1993) while very critical of transitivity assumptions in economics, considers the basic idea of “considerable pedagogical value” (p. 345). This is an idea that has been picked up by several economists and philosophers to argue that introductory books and lectures to economics give a misleadingly narrow picture of the field at large.<sup>8</sup> This, however, need not be a problem. The subject matter of economics is complex and it might be best to start with highly idealized models that include the axiom of transitivity, even when its role is merely heuristic.

Nevertheless, the literature has provided three primary arguments for transitive preference-orderings that Anand (1993) in his influential essay sought to dispel. Firstly, Anand argues that transitivity has been defended as logical consistency. Here, intransitivity is simply a logical mistake – analogous to a mistake in logical reasoning – defended for instance by Broome (1991). This, Anand argues does not work, for it locates the mistake not in the logical preference relation, but the assumption that preferences cannot change if options are added or removed (an assumption that has been challenged in the literature, see Sugden 1985).

Secondly, Anand points to the defense of transitivity as something embedded in the concept of rationality itself. Here Anand (1993, p. 340) quotes a passage Davidson (1980),<sup>9</sup> who argues that:

---

<sup>8</sup> See Rodrik (2015); Ylikoski and Aydinonat (2014); Aydinonat (2018); Veit (2019b, 2021b).

<sup>9</sup> Anand (1993, p. 340) accidentally cites page 237 of Davidson. The actual page number is 273.

“theory [...] is so powerful and simple, and so constitutive of concepts assumed by further satisfactory theory [...] that we must strain to fit our findings, or interpretations, to fit the theory. If length is not transitive, what does it mean to use a number of measure length at all? We could find or invent an answer, but unless or until we do, we must strive to interpret ‘longer than’ so that it comes out transitive. Similarly ‘for preferred to’.”

– Donald Davidson (1980, p. 273)

Anand argues that we should not overestimate this metaphor. In order to do so, he introduces an alternative metaphor, i.e. idea of pair-wise competitions of sport teams. While the highest ranked team frequently beats the second ranked team, a lower-ranked team might have the perfect composition to beat the first ranked team. There is nothing surprising about such reversals in sports, indeed, it would be ludicrous and boring if the highest ranked team beats all others, the second highest ranked team beats all except for the first – and so on for the entire ranking list.

Anand (1993) does not so much as argue that this is the right interpretation of preferences, but rather to make the point that these are mere metaphors and there is no a-priori reason or empirical evidence as of yet to think that one of them is *the way* of seeing preferences. Instead, we might be well-advised to see these different suggestions as mere metaphors. Interestingly, Nancy Cartwright (2019) makes a similar argumentative move when she criticizes the metaphorical idea of “laws of nature” and “nature doing it by the book”, instead introducing her own metaphor of “nature as an artful modeler”. While I find the metaphor misplaced, one can see how easy it is to be tempted by metaphors. If one disagrees with the metaphors of a particular theory, whether in philosophy or science, it will often be necessary to come up with alternative metaphors. Dennett (1991) vaguely alludes to this possibility as “war of metaphors“ (p. 455), when he defends the use of metaphors as tools of thought. When there are two sides of a debate, and one has metaphors in their arsenal while the other doesn’t, the latter will be put into a disadvantaged position. Defenders of the transitivity axiom unfortunately had this unrecognized advantage for the majority of the debate.

In addition to Anand’s criticism, it is important to note that Davidson’s defense of transitive preference orderings is based on outdated views in the philosophy of science. Davidson (1980) states that “Hempel set out to show that reason explanations do not differ in their general logical character from explanation

in physics or elsewhere” and that his own “reflections reinforce this view” (p. 274). While he avoids the conclusion that we can extrapolate to general laws about human behavior – he argues that we can find general laws about individual humans such as *Gerald Ford* that would apply under certain conditions. This idea is deflating the idea of laws to such narrow domains, that it is hardly even worth speaking of laws, and even in such a narrow domain they are unlikely to be exceptionless. More commonly, philosophers of science are now following Nancy Cartwright’s (1983) suggestion to see such generalizations as useful idealizations in models. The discovery of general laws is no longer seen as a necessary condition for successful explanation.

Lastly, and perhaps most importantly, Anand (1993) discusses a popular *reductio ad absurdum* argument against critics of the transitivity axiom, i.e. the money pump. The argument goes as follows. Suppose we have an agent who prefers A over B, B over C, and C over A. Suppose now that this agent is in possession of B. Because of the cyclical preference structure of this agent, a merchant who is in possession of A and C should be able to swap his own A for the agent’s B in addition to a tiny amount of money such that the preference relation between A and B remains intact. Since the merchant is also in possession of C he will be able to expose the agent to a continuous set of exchanges with a minor additional cost that he would be ‘rational’ to agree to given his cyclical preference ordering. These repeated exchanges, however, would eventually lead to the bankruptcy of our agent holding cyclical preferences. Hence, they are being money-pumped.

This argument is a strong and intuitive one, for it seems to suggest that unless we accept the transitivity of preferences as a necessary requirement of rationality – it would be rationally required to give away all of one’s money. The assumption has been criticized on the grounds that it seems to assume a stable preference set over an entire life, but this does not seem to be a requirement of rationality. There is a stronger counterpoint against the money pump argument, however, that draws on literature in evolutionary biology and behavioral ecology. But before we turn to the literature on intransitive choice in animals, let me briefly summarize this section.

As this section hoped to make clear, the axiom of transitivity has long played a central role in economics in order to enable meaningful attributions of utilities to alternative choices. This instrumentalist defense of transitivity, however, has been criticized by economists and psychologists who were interested in actual choice behaviour. One might describe this conflict thus as one between the normative-idealist stance of mainstream economics and the descriptive-realist stance of behavioural economists and psychologists. Some economists may object to being described as ‘normativists’, but arguments like the money pump rely upon the

normative assumption that it is bad to be exploited. Nevertheless, economists have tried to justify the normativity of the transitivity axiom through recourse on a purely descriptive kind of normativity in biology to which we shall now turn, i.e. the maximization of fitness.

### 3 Intransitivity and evolution

Unlike the ‘Rational Animal’, non-human animals are often taken to be irrational. This philosophical conception of rationality goes back to Aristotle and was intended to distinguish man from animal. For the purposes of this paper, we will discard this a priori distinction between humans and animals and show that there is much to learn from the debate on intransitive preferences in non-human animals.

While the P- and E-concepts of rationality seemed incompatible, economists frequently suggest that there is a more important form of rationality economists can rely on, even if the E-concept fails to represent and accurately explain actual human thought processes in markets, i.e. B-Rationality. This Biological Rationality concept is simply the maximization of fitness – and, hence, was often used as an analogue to justify models that assume the maximization of utility (see Okasha 2018; Okasha and Binmore 2012). E-Rationality, however, is frequently violated by both humans and animals. So it is worth exploring whether the connection to B-Rationality can actually help economists to justify their highly idealized form of E-Rationality.

In a biological context, ‘optimal’ often replaces talk of ‘rational’ (see Smith and Harper 2003; Okasha 2018). The optimal choice, in terms of maximization of fitness, then becomes the parallel to the rational choice, i.e. the choice that maximizes utility. The parallel is obvious, but it is not clear how far the analogy stretches and whether it is, indeed, a useful one.

When it comes to E-Rationality there is now an extensive literature on rational choice behavior in animals. McGonigle and Chalmers (1992) for instance argue that squirrel monkeys are capable of transitive choice behavior. For non-human animals, it is sometimes assumed that optimal behavior, i.e. fitness-maximizing behavior, would always correspond to the transitivity axiom, but as Okasha (2018) points out this need not be the case. He discusses a biological optimality model of Houston et al. (2007) in which transitivity is violated - and yet fitness maximized. The Houston et al. (2007) paper is thus aptly titled “Violations of transitivity under fitness maximization”. In their model, animals have to choose between three different foraging options. Each option is associated with a different predation risk and an

associated chance of success. The nutritional value itself is equal for all. Whether a particular option is preferred to another depends on the state the animal is in. The ‘goal’ for the animal, however, as Okasha (2018) notes is to survive the winter and avoid starvation. Houston et al. (2007) show that the best strategy (to maximize fitness) involves intransitive choices for a range of intermediate energy reserves, i.e. neither full nor starved.<sup>10</sup>

The moral here, as Okasha points out, is a similar one to an important result in the behavioral economics literature. When we analyse choices in isolation, they may violate transitivity and appear irrational. The actual strategies that underlie the choice behavior, however, might be rational because they are about repeated actions. What should be rationally evaluated then is not the individual choice but the strategy itself.

Consider the simple thought experiment of a hypothetical conference meeting with a long queue in front of the food-stand. Our human agent, let us call him Bob, is given the option between eating a salad, a plate with sliced peaches, or a steak. Bob picks the steak. However, it turns out there is more food than participants so everyone is allowed to choose again. After Bob has enjoyed his steak, he proceeds to join the queue again. This time, however, he chooses the salad. How odd you say? Let us make matters worse. Once again, there are food leftovers. Bob joins is faced with the three items once more. This time, however, he chooses the sliced peaches. Now our straw-man economists might yell: “How irrational!” Psychologists, of course, have no problem with explaining such choice behavior. But neither do contemporary economists.

Clearly, it need not be irrational if Bob chooses the steak, and is subsequently allowed to once again choose between the two after he has devoured the steak, other people have made their choices, and there are leftovers. As Okasha (2018) nicely illustrates, behavioral economists have here responded in a similar way to biologists such as Houston et al. (2007); McNamara et al. (2014) who note that the irrationality disappears once we change our perspective to look at the level of strategies, rather than just the individual choices, a view that is gaining support through recent work in the neurosciences (see Kalenscher et al. 2010). Thus, the evolutionary most ‘rational’ strategy can lead to intransitivity among individual choices.

This explanation is also able to explain the tendency of children and infants to exhibit intransitive preferences that seems to stem from a preference over novelty that is lost over time (Bradbury and Ross 1990). We could rationalize this as the

---

<sup>10</sup> Okasha (2018) discusses this example in more detail.

progressive development of ‘rationality’ into adulthood – or a beneficial exploratory phase during early years. Curiosity could be a useful exploratory strategy in rapidly changing environments, for instance. Similar patterns can be found in the foraging behavior of bees (Shafir 1994). This is a better response to the money pump argument: we often need to take the context, time, and number of repeated choices into account. This has led Gigerenzer and Todd (1999) and Smith (2003) to develop, what they call *Ecological Rationality*, as an alternative to standard Rational Choice Theory. Again, it is not my goal here to defend one ‘Rationality’ account over another, but rather to highlight the importance of idealization when the concept is used in practice.<sup>11</sup>

Having addressed the major opposition to the abolition of the transitivity axiom we shall now turn to the much more interesting philosophical questions concerning idealization and representation by drawing on the philosophy of models literature.

## 4 Rationality Redux

As the previous sections should make clear, the disagreements about how we should conceptualize rationality do not just reflect the complexity of the concept. Rather, the disagreements are indicative of deeper differences in regard to why we use the models, concepts, and other clusters of ideas related to rationality at all. Thus, my goal in this section will be to draw on the philosophy of science literature on modeling and idealization to argue that the transitivity axiom of rationality cannot simply be assessed as being either correct or false. Instead, I will defend a pragmatic and pluralist stance in which we employ different concepts and models of rationality depending on the goal we are using them for.

As is indicative of the rational choice axiom of transitivity that I have focused on in this article, the last 70 years appear to show no success in removing classical rational choice models from introductory microeconomics books despite many criticisms. Indeed, in these 70 years a huge variety of elegant alternatives have been

---

<sup>11</sup> I will note, however, that this doesn’t mean that there can be useful connection between these concepts. As I’ve argued in a recent book, the demands on animals to engage in optimizing behaviour could explain the evolution of Benthamite creatures with economic agency that have a common currency to rank/evaluate alternative actions, thus perhaps providing an evolutionary bridge between these concepts (Veit 2022, 2023a).

developed that do not rely on the axiom of transitive preference ordering, or least only a weaker version. To some extent, this literature may appear an endeavor in futility. None of the successor models have achieved sufficient prominence to replace the original status of the transitivity axiom. Here, both economists and philosophers have been misguided. It is a mistake the following quote from Fishburn's (1991) review of the literature elegantly illustrates:

“If the variety of representations is more confusing than illuminating, one would hope that further research during the next few decades will help to identify the most viable models on the basis of philosophical arguments, empirical robustness, and applications potential. General but elegant models that are capable of representing what most researchers agree are reasonable patterns of preference will likely prevail. Some of these surely await discovery.”

– Peter C. Fishburn (1991, p. 131)

Almost 30 years later, we must recognize that Fishburn's prediction failed. No general model has been ‘discovered’ that is able to represent all reasonable patterns of preference.<sup>12</sup> Is this a failure of economics? I suggest not. Indeed, we should see the extreme proliferation of rational choice models as an utter success. But we need to change our understanding of what economists have achieved. Even though many of the economists engaged in this debate had the goal of developing a general model that is able to cover a broader range of phenomena, almost all of them failed. But this does not mean that there was no progress in the last 70 years in our understanding of rational choice behavior. A consensus has emerged that there are certain circumstances under which the transitivity axiom is unproblematic, elegant, and predictively powerful.

Reasonable economists have given up on the idea that transitivity of preferences is a general feature of *all* rational choice behavior. To this end, a large number of theoretical and empirical contributions from psychology, economics, philosophy, and biology have added to our understanding of ‘rationality’ as a cluster of concepts, rather than a single one. There is no single phenomena of rationality in nature that could unify these different concepts and models. To recognize this,

---

<sup>12</sup> Let alone elegant.

however, we must shift our understanding of models away from what Veit (2019b, 2023b) has called “model monism” or “model essentialism”, and towards a more pluralist position he has dubbed “model pluralism”:

“(i) any successful analysis of models must target sets of models, their multiplicity of functions within science, and their scientific context and history and (ii) for almost any aspect  $x$  of phenomenon  $y$ , scientists require multiple models to achieve scientific goal  $z$ .”

– Veit (2019b, pp. 92–93)

While unification is certainly a worthwhile goal, there is a misguided tendency within economics to seek *the one perfect and general model*. This tendency should be avoided. But in practice, not much will have to change for economists – they can and should continue to build new models and expand our toolkit of possible explanations. Articles, such as Regenwetter et al. (2011), attempt to rationalize many of the empirical studies on intransitive choice as *actually* consistent with transitive preferences. I see this as a double-edged sword. On the one hand, I am reluctant to accept the calls to abolish traditional rational choice theory by some of its critics. On the other, I am not willing to grant that the conclusion, that because many of these studies are somewhat consistent with axiom of transitive preference orderings, we do not need alternative models. The debate, however, is often put in a very monist and competitive way. This, I hope to have successfully illustrated, is a mistake. Instead, we need to embrace a pluralism of alternative models.

Granted, for my proposed changes to succeed, there will have to be a major change in the public understanding of the core role of idealizations in economics. Philosophers are well-advised to promote this change, rather than argue against the viability of idealizations in science. Idealizations are everywhere. It is important to see them as tools for our models to perform their intended roles. Whether it is explanation, prediction, or even unification – idealizations are a must.

The topic of idealization, however, has been one of the most longstanding debates in the philosophy of science literature, much of which we consider too critical (e.g. Cartwright 1983, 2009; Hausman 1992; de Donato Rodriguez and Bonilla 2009; Knuuttila 2009; Mäki 2009b; Reiss 2012; Northcott and Alexandrova 2015; Fumagalli 2015, 2016). Idealizations as distortions, misrepresentations, and falsehoods, have often been viewed with suspicion, if not contempt, by more traditionally inclined philosophers. These views are indicative of a more general tendency among



philosophers of science to come up with sweeping generalizations about science – a dangerous tendency that has contributed to a sometimes quite dismissive picture of philosophy of science by scientists.<sup>13</sup>

This way of thinking, however, is beginning to change. Thanks to philosophers such as Michael Weisberg,<sup>14</sup> Ronald Giere (1999, 2006), Peter Godfrey-Smith (2006), Angela Potochnik (2017), N. Emrah Aydinonat (2018), and hopefully myself (Veit 2019b), there is now a growing understanding of the necessary and diverse roles idealizations play within science. It is into this tradition that the present article squarely falls.

As I have illustrated above, the debate about rationality in economics has unfortunately suffered from a lot of bad methodological and conceptual confusions regarding the need for consensus on a single definition of rationality. Akin to debates between political parties a rift has opened between critics and proponents of economics, with both sides seeing the other as political partisans and holders of naive views about science. Economists have responded to challenges of the transitivity axiom in variety of ways. Critics, however, especially from the psychology-friendly side of economics, i.e. behavioral economics, remain unconvinced. Subsequently, economists have developed a number of alternatives for traditional expected utility maximization that do not rely on transitive preference orderings and that are more or less in line with the idea of bounded rationality (see Morrison 1962; Tversky 1969; Fishburn 1982; Bell 1982; Loomes and Sugden 1982).

How should one interpret these alternative models of rational choice? It was my goal here to dispel the perceived need for a unified account that covers all of economic (and possibly biological) choice behavior. Economic imperialism has led to the application of rational choice theory to a variety of phenomena, formerly seen as outside the domain of economics. The problem here is not the application of the models itself. We should treat them as idealized tools that can at best only partially represent the world. Yet, the use of diverse tools enables us to discover new explanatory insights, a point that has recently gained prominence through a position

---

<sup>13</sup> See Maynard Smith (1997); Godfrey-Smith (2003); Veit (2019b, 2023b).

<sup>14</sup> Weisberg has published a number of highly influential articles on models that I deem to be of special importance for the shift towards a more pluralist understanding of models in the literature: see Weisberg (2003, 2006b,a, 2007b,a, 2012), Weisberg and Reisman (2008), Matthewson and Weisberg (2009), Weisberg et al. (2011), Elliott-Graves and Weisberg (2014)

that has come to be named ‘Perspectivism’ or ‘Perspectival Realism’.<sup>15</sup> This does not entail that we should become anti-realists about ‘Rationality’, yet it does require changes in how we perceive it.

Should we, for instance, consider failure to exhibit transitive choice behavior in other animals, such as hoarding gray jays (Waite 2001), as a depiction of their ‘irrational’ behavior? I think not. The question is ill-posed and presumes that there is a general answer to questions involving the concept of ‘Rationality’, which Kacelnik (2006) early on tried to warn us off. As I hope to have convinced the reader, rationality might not be the unified phenomenon philosophers have taken it to be. Rather, it is a loose collection of metaphors, models, and idealizations – epistemic tools that help us to explain and make sense of the world. The different concepts we may associate with rationality, such as E-, P-, and B-Rationality reflect genuinely different phenomena that may have similarities, but shouldn’t be grouped together. Indeed, we should move away from attempts to provide the one true account of rationality. This is, as has hopefully become clear now, a hopeless endeavor. A more subtle and pragmatic way forwards for economics (and other disciplines making use of the concept of rationality), would be to embrace a pluralist perspective, and defend models that are not intended to replace all others but instead illuminate a novel aspect or provide a new perspective of a phenomena.

## 5 Conclusion

In this article, I have criticized the common attempts to find something like the one true theory of rationality or for that matter truth or falsity of the axiom of transitivity. One immediate response to such criticisms will naturally be what we should be doing instead. Drawing on the philosophy of modeling literature, I have therefore argued that we should reconceptualize these debates in terms of determining useful models for different purposes. This will help us to recognize the different conceptualizations of rationality in (evolutionary) biology, economics, and psychology as reflecting different interests. We should see the concepts of rationality and its axioms such as transitivity as idealized conceptual tools, rather than accurate explications of “the one true” concept of rationality.

---

<sup>15</sup> See Giere (2006) for the first articulation of the view, and Massimi (2017) for a recent overview.

There is a special explanatory force that comes from explanations invoking ‘Rationality’ and ‘Reason’ to us as cognitively limited agents that evolved to talk and think in normative terms – but it is a tempting force that might lead us into the wrong conclusions if we mistake what are useful tools for representations of reality.<sup>16</sup> The final conclusion for economists (and for that matter biologists and cognitive scientists) is a simple, but philosophically less interesting one: there is set of cases where it is reasonable and/or useful to accept the axiom of transitive preference orderings, while it is not for others. No generalized defense or rejection of this idealization can be offered. The real insight and philosophically much more interesting one is this: we may have to give up on the idea of rationality as a unified concept or phenomena, and instead think of it as a useful set of metaphors, models, and idealizations.

## References

I would like to thank Heather Browning and two anonymous reviewers for their feedback on this manuscript.

## References

- Alekseev, A., G. W. Harrison, M. Lau, and D. Ross (2019). Deciphering the Noise: The Welfare Costs of Noisy Behavior. CEAR Working Paper, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.
- Anand, P. (1993). The philosophy of intransitive preference. *The Economic Journal* 103(417), 337–346.
- Angner, E. and G. Loewenstein (2007). Behavioral economics. *Handbook of the philosophy of science: Philosophy of economic*, 641–690.
- Ashraf, N., C. F. Camerer, and G. Loewenstein (2005). Adam Smith, behavioral economist. *Journal of Economic Perspectives* 19(3), 131–145.
- Aydinonat, N.E. (2018). The diversity of models as a means to better explanations in economics. *Journal of Economic Methodology* 25(3), 237– 251.

---

<sup>16</sup> See Godfrey-Smith (2013) and Okasha (2018).

- Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pp. 13–68. Springer.
- Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political economy* 81(4), 813–846.
- Becker, G. S. (1974). A theory of marriage: Part II. *Journal of political Economy* 82(2, Part 2), S11–S26.
- Becker, G. S. (1976). *The economic approach to human behavior*. University of Chicago press.
- Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *Journal of political economy* 101(3), 385–409.
- Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations research* 30(5), 961–981.
- Bradbury, H. and K. Ross (1990). The effects of novelty and choice materials on the intransitivity of preferences of children and adults. *Annals of Operations Research* 23(1), 141–159.
- Broome, J. (1991). *Weighing goods: Equality, uncertainty and time*. Oxford: Basil Blackwell Press.
- Browning, H. and Veit, W. (2023). Studying Introspection in Animals and AIs. *Journal of Consciousness Studies*. 30 (9):63-74. <https://doi.org/10.53765/20512201.30.9.063>
- Camerer, C. (1999). Behavioral economics: Reunifying psychology and economics. *Proceedings of the National Academy of Sciences* 96(19), 10575– 10577.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press.
- Cartwright, N. (2009). If no capacities then no credible worlds. But can models reveal capacities? *Erkenntnis* 70(1), 45–58.
- Cartwright, N. (2019). *Nature, the artful modeler: Lectures on laws, science, how nature arranges the world and how we can arrange it better*, Volume 23. Open Court Publishing.
- Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.
- de Donato Rodriguez, X. and J. Z. Bonilla (2009). Credibility, idealisation, and model building: An inferential approach. *Erkenntnis* 70(1), 101–118.
- Dennett, D. (1991). *Consciousness explained*. United States: Little, Brown and Co.
- Dennett, D. C. (1989). *The intentional stance*. MIT press.
- Descartes, R. (2008). *Meditations on first philosophy: With selections from the objections and replies*. Oxford University Press.

- Diamond, A. M. (2008). Science, economics of. *The New Palgrave Dictionary of Economics* 7.
- Elliott-Graves, A. and M. Weisberg (2014). Idealization. *Philosophy Compass* 9(3), 176–185.
- Fishburn, P. C. (1982). Nontransitive measurable utility. *Journal of Mathematical Psychology* 26(1), 31–67.
- Fishburn, P. C. (1991). Nontransitive preferences in decision theory. *Journal of risk and uncertainty* 4(2), 113–134.
- Fumagalli, R. (2015). No learning from minimal models. *Philosophy of Science* 82(5), 798–809.
- Fumagalli, R. (2016). Why we cannot learn from minimal models. *Erkenntnis* 81(3), 433–455.
- Giere, R. (1999). *Science without Laws*. Chicago: University of Chicago Press.
- Giere, R. (2006). *Scientific Perspectivism*. Chicago: University of Chicago Press.
- Gigerenzer, G. and P. M. Todd (1999). Ecological rationality: the normative study of heuristics. In *Ecological rationality: Intelligence in the world*, pp. 487–497. Oxford University Press.
- Godfrey-Smith, P. (2003). *Theory and reality: An introduction to the philosophy of science*. University of Chicago Press.
- Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and philosophy* 21(5), 725–740.
- Godfrey-Smith, P. (2013). Darwinian individuals. In F. Bouchard and P. Huneman (Eds.), *From groups to individuals: Evolution and emerging individuality*, pp. 17–36. MIT Press Cambridge, MA.
- Green, D. and I. Shapiro (1996). *Pathologies of rational choice theory: A critique of applications in political science*. Yale University Press.
- Grether, D. M. and C. R. Plott (1979). Economic theory of choice and the preference reversal phenomenon. *The American Economic Review* 69(4), 623–638.
- Hausman, D. M. (1992). *The inexact and separate science of economics*. Cambridge University Press.
- Houston, A. I., J. M. McNamara, and M. D. Steer (2007). Violations of transitivity under fitness maximization. *Biology Letters* 3(4), 365–367.
- Kacelnik, A. (2006). Meanings of rationality. In S. Hurley and M. Nudds (Eds.), *Rational animals*, pp. 87–106.
- Kalenscher, T., P. N. Tobler, W. Huijbers, S. M. Daselaar, and C. Pennartz (2010). Neural signatures of intransitive preferences. *Frontiers in Human Neuroscience* 4, 49.

- Knuuttila, T. (2009). Isolating representations versus credible constructions? Economic modelling in theory and practice. *Erkenntnis* 70(1), 59–80.
- Korhonen, P., H. Moskowitz, and J. Wallenius (1990). Choice behavior in interactive multiple-criteria decision making. *Annals of Operations Research* 23(1), 161–179.
- Levitt, S. D. and S. J. Dubner (2005). *Freakonomics*. New York: William Morrow. Harper Collins.
- Loomes, G. and R. Sugden (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The economic journal* 92(368), 805–824.
- Mäki, U. (2009a). Economics imperialism: Concept and constraints. *Philosophy of the social sciences* 39(3), 351–380.
- Mäki, U. (2009b). MISSING the world. Models as isolations and credible surrogate systems. *Erkenntnis* 70(1), 29–43.
- Massimi, M. (2017). Perspectivism. In *The Routledge handbook of scientific realism*, pp. 164–175. Routledge.
- Matthewson, J. and M. Weisberg (2009). The structure of tradeoffs in model building. *Synthese* 170(1), 169–190.
- Maynard Smith, J. (1997). Interview by Richard Dawkins. *Web of Stories*. <https://www.webofstories.com/play/john.maynard.smith/1> [Online; accessed 13-September-2023].
- McGonigle, B. and M. Chalmers (1992). Monkeys are rational! *The Quarterly Journal of Experimental Psychology* 45(3), 189–228.
- McNamara, J. M., P. C. Trimmer, and A. Houston (2014). Natural selection can favour ‘irrational’ behaviour. *Biology letters* 10(1), 20130935.
- Morrison, H. (1962). Intransitivity of Paired Comparison Choices. University of Michigan. *Ann Arbor, Ph. D. Dissertation*.
- Noë, R. and P. Hammerstein (1994). Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral ecology and sociobiology* 35(1), 1–11.
- Noë, R. and P. Hammerstein (1995). Biological markets. *Trends in Ecology & Evolution* 10(8), 336–339.
- Northcott, R. and A. Alexandrova (2015). Prisoner’s Dilemma doesn’t explain much. In M. Peterson (Ed.), *The Prisoner’s Dilemma*, pp. 64–84. Cambridge University Press.
- Noë, R., J. A. R. A. M. V. Hooff, and P. Hammerstein (2001). *Economics in Nature: Social Dilemmas, Mate Choice and Biological Markets*. Cambridge University Press.
- Okasha, S. (2018). *Agents and goals in evolution*. Oxford University Press.

- Okasha, S. and K. Binmore (Eds.) (2012). *Evolution and Rationality: Decisions, Cooperation and Strategic Behaviour*. Cambridge University Press.
- Panda, S. C. (2018). Rational Choice with Intransitive Preferences. *Studies in Microeconomics* 6(1-2), 66–83.
- Potochnik, A. (2017). *Idealization and the Aims of Science*. University of Chicago Press.
- Regenwetter, M., J. Dana, and C. P. Davis-Stober (2011). Transitivity of preferences. *Psychological review* 118(1), 42.
- Reiss, J. (2012). Idealization and the aims of economics: three cheers for instrumentalism. *Economics & Philosophy* 28(3), 363–383.
- Robbins, L. (1935). *An Essay on the Nature and Significance of Economic Science*. London: MacMillan and Co.
- Rodrik, D. (2015). *Economics rules: Why economics works, when it fails, and how to tell the difference*. OUP Oxford.
- Rosenberg, A. (1992). *Economics—Mathematical Politics or Science of Diminishing Returns?* University of Chicago Press.
- Rosenberg, A. (1994). If economics isn't science, what is it. In D. M. Hausman (Ed.), *The Philosophy of Economics: An Anthology*, pp. 376–394.
- Rosenberg, A. (1995). The metaphysics of microeconomics. *The Monist* 78(3), 352–367.
- Rosenberg, A. (2009). If economics is a science, what kind of a science is it? In Don Ross (ed.), Harold Kincaid (ed.), *The Oxford handbook of philosophy of economics*. Oxford: Oxford University Press.
- Ross, D. (2005). *Economic theory and cognitive science: Microexplanation*. MIT Press.
- Ross, D. (2014). *Philosophy of economics*. Palgrave Macmillan.
- Russell, B. (2009). *The basic writings of Bertrand Russell*. Routledge.
- Rysiew, P. (2015). *Rationality*. Oxford Bibliographies.  
<https://doi.org/10.1093/OBO/9780195396577-0175>
- Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.
- Schumm, G. F. (1987). Transitivity, preference and indifference. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 52(3), 435–437.
- Sen, A. (1969). Quasi-transitivity, rational choice and collective decisions. *The Review of Economic Studies* 36(3), 381–393.
- Sen, A. (1970). *Collective Choice and Social Welfare*. San Francisco: Holden Day.

- Sen, A. (1971). Choice functions and revealed preference. *The Review of Economic Studies* 38(3), 307–317.
- Sen, A. (1977). Social choice theory: A re-examination. *Econometrica: journal of the Econometric Society*, 53–89.
- Shafir, S. (1994). Intransitivity of preferences in honey bees: support for ‘comparative’ evaluation of foraging options. *Animal Behaviour* 48(1), 55–67.
- Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics* 69(1), 99–118.
- Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization* 1(1), 161–176.
- Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization science* 2(1), 125–134.
- Simon, H. A. (1997). *Models of bounded rationality: Empirically grounded economic reason*, Volume 3. MIT press.
- Smith, A. (2010). *The Wealth of Nations: An inquiry into the nature and causes of the Wealth of Nations*. Harriman House Limited.
- Smith, J. M. and D. Harper (2003). *Animal signals*. Oxford University Press.
- Smith, V. L. (2003). Constructivist and ecological rationality in economics. *American economic review* 93(3), 465–508.
- Stigler, G. J. (1984). Economics: The imperial science? *The Scandinavian Journal of Economics* 86(3), 301–313.
- Sugden, R. (1985). Why be consistent? A critical analysis of consistency requirements in choice theory. *Economica* 52(206), 167–183.
- Suzumura, K. (1983). *Rational choice, collective decisions, and social welfare*. Cambridge University Press.
- Thaler, R. H. (2016). Behavioral economics: Past, present, and future. *American Economic Review* 106(7), 1577–1600.
- Tullock, G. (1972). Economic imperialism. In J. M. Buchanan and R. D. Tollison (Eds.), *Theory of public choice*, pp. 317–329. University of Michigan Press Ann Arbor.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological review* 76(1), 31–48.
- Veit, W. (2023a). *A Philosophy for the Science of Animal Consciousness*. New York: Routledge.
- Veit, W. (2023b). Model Anarchism. *THEORIA. An International Journal for Theory, History and Foundations of Science*. <https://doi.org/10.1387/theoria.23849>



- Veit, W. (2023). Evolutionary Game Theory and Interdisciplinary Integration. *Croatian Journal of Philosophy*. 23(67), 33-50. <https://doi.org/10.52685/cjp.23.67.2>
- Veit, W. (2022). Complexity and the Evolution of Consciousness. *Biological Theory*. <https://doi.org/10.1007/s13752-022-00407-z>
- Veit, W. (2021a). Agential Thinking. *Synthese*. <https://doi.org/10.1007/s11229-021-03380-5>
- Veit, W. (2021b). Model Diversity and the Embarrassment of Riches. *Journal of Economic Methodology*. <https://doi.org/10.1080/1350178X.2021.1898660>
- Veit, W. (2019a). Evolution of multicellularity: cheating done right. *Biology & Philosophy* 34(3), 34. <https://doi.org/10.1007/s10539-019-9688-9>
- Veit, W. (2019b). Model pluralism. *Philosophy of the Social Sciences* 50(2), 91–114. <https://doi.org/10.1177/0048393119894897>
- Veit, W., J. Dewhurst, K. Dolega, M. Jones, S. Stanley, K. Frankish, and D. C. Dennett (2020). The Rationale of Rationalization. *Behavioral and Brain Sciences* 43, e53.
- von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*.
- Waite, T. A. (2001). Intransitive preferences in hoarding gray jays (*Perisoreus canadensis*). *Behavioral Ecology and Sociobiology* 50(2), 116–121.
- Weisberg, M. (2003). When less is more: Tradeoffs and idealization in model building. Ph.D. thesis, Stanford University.
- Weisberg, M. (2006a). Forty years of ‘The strategy’: Levins on model building and idealization. *Biology and Philosophy* 21(5), 623–645.
- Weisberg, M. (2006b). Robustness analysis. *Philosophy of Science* 73(5), 730–742.
- Weisberg, M. (2007a). Three kinds of idealization. *The journal of Philosophy* 104(12), 639–659.
- Weisberg, M. (2007b). Who is a Modeler? *The British journal for the philosophy of science* 58(2), 207–233.
- Weisberg, M. (2012). *Simulation and similarity: Using models to understand the world*. Oxford University Press.
- Weisberg, M., S. Okasha, and U. Mäki (2011). Modeling in biology and economics. *Biology & Philosophy* 26(5), 613–615.
- Weisberg, M. and K. Reisman (2008). The robust Volterra principle. *Philosophy of science* 75(1), 106–131.

Ylikoski, P. and N. E. Aydinonat (2014). Understanding with theoretical models.  
*Journal of Economic Methodology* 21(1), 19–36.