Journal of Economic Methodology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rjec20

Introduction: methodologies of bounded rationality

Till Grüne-Yanoff[a], Caterina Marchionni[b] & Ivan Moscati[c]

[a] Department of Philosophy and History of Technology, Royal Institute of Technology, Stockholm, Sweden

[b] Department of Political and Economic Studies, Academy of Finland Centre for the Philosophy of the Social Sciences, University of Helsinki, Helsinki, Finland

[c] Department of Economics, University of Insubria, Varese, Italy

Published online: 09 Dec 2014.

To cite this article: Till Grüne-Yanoff, Caterina Marchionni & Ivan Moscati (2014) Introduction: methodologies of bounded rationality, Journal of Economic Methodology, 21:4, 325-342, DOI: 10.1080/1350178X.2014.972140

To link to this article: http://dx.doi.org/10.1080/1350178X.2014.972140

Journal of Economic Methodology, 2014
Vol. 21, No. 4, 325–342, http://dx.doi.org/10.1080/1350178X.2014.972140

Introduction: methodologies of bounded rationality

Till Grüne-Yanoff[a], Caterina Marchionni[b]* and Ivan Moscati[c]

[a]*Department of Philosophy and History of Technology, Royal Institute of Technology, Stockholm, Sweden;* [b]*Department of Political and Economic Studies, Academy of Finland Centre for the Philosophy of the Social Sciences, University of Helsinki, Helsinki, Finland;* [c]*Department of Economics, University of Insubria, Varese, Italy*

(*Received 20 August 2014; accepted 30 September 2014*)

The modelling of bounded rationality is currently pursued by approaches that exhibit a wide diversity of methodologies. This special issue collects five contributions that discuss different methodological aspects of these approaches. In our introduction, we map the variety of methodological positions with respect to three questions. First, what kinds of evidence do the respective approaches consider relevant for modelling bounded rationality? Second, what kind of modelling desiderata do the respective approaches focus on? And third, how do the respective approaches justify the normative validity of bounded rationality? To broaden the picture, we not only discusss the five contributions of this issue, but also include relevant positions from the extant literature.

Keywords: bounded rationality; heuristics; revealed preferences; evidence; modelling strategies; normative justification

1. From rational choice to bounded rationality

Standard rational choice theory (RCT) assumes that the decision-maker has well-behaved preferences over a set of alternatives, and that she chooses the most preferred alternative. By 'well-behaved', it is meant that preferences satisfy certain properties such as completeness or transitivity. The decision-maker's well-behaved preferences can be represented by a utility function, and her choices can be interpreted as the result of utility maximization. The alternatives over which preferences are defined may be either risk-free, in which case RCT is typically associated with ordinal utility theory, or risky, in which case RCT is associated with expected utility theory (EUT). For risk-free contexts, RCT shows that choices resulting from well-behaved preferences display certain consistency features that are characterized by the axioms of revealed preference theory, such as the weak axiom of revealed preferences (WARP) or the strong axiom of revealed preferences (SARP). In turn, choices satisfying these axioms can be interpreted as resulting from the maximization of well-behaved preferences (for an overview see Mas-Colell, Whinston, & Green, 1995, Chaps 1–3, 6, and Ross, this issue).

RCT is widely considered to have produced 'an elegant, parsimonious, imposing, imperial structure' (Harstad & Selten, 2013, p. 497), and its defenders point both to its alleged predictive success and to its normative validity. Moreover, RCT has provided the framework for traditional welfare analysis: if rational decision-makers always choose what is best for them, the planner should select the alternative they would choose and implement their decisions. However, in the last 30 years or so, economists and

*Corresponding author. Email: caterina.marchionni@helsinki.fi

psychologists have accumulated persuasive empirical evidence that challenges RCT as a descriptively adequate theory of human decision-making.

Generally speaking, any theory of decision-making that deviates from the RCT paradigm can be labelled as a theory of boundedly rational decision (for an overview, see Grüne-Yanoff, 2007). Since there are various ways of deviating from RCT, there are also different theories of bounded rationality (BR). For instance, we may drop the assumption that preferences are well behaved and construct a theory of decision with incomplete preferences. Or we may assume that preferences and utility are not defined only over sets of alternatives but also depend on other elements, such as the decision-maker's endowment or the way the alternatives are presented to him. Alternatively, we may argue that the decision-maker does not choose his most preferred alternative but, for example, an alternative that is sufficiently satisficing according to his preferences. An instance of the latter strategy is the one adopted by Herbert Simon in his 1955 article, which can be considered as seminal for all the subsequent BR literature (Simon, 1955). After Simon, and especially in recent years, different BR theories have been formulated. Although none of them has yet displaced RCT as the canonical theory taught in undergraduate and graduate economics courses, their popularity is on the rise (for a discussion of why RCT is still canonical, see, e.g. Hands and Ross, this issue).

In the following, we will discuss methodological aspects of three approaches to BR. The first one is often associated with the field of behavioural economics and the research of individuals such as Amos Tversky, Daniel Kahneman, Richard Thaler, Colin Camerer, George Loewenstein, and Matthew Rabin. In this approach, choices are still modelled as the result of utility maximization, but the utility function that is maximized is not the standard RCT utility function because, for instance, the utility of a given alternative depends on the utility of another alternative taken as a reference point. This approach to BR is also called the 'heuristics-and-biases' (H&B) program because differences between the standard RCT utility function and the actually maximized utility function are imputed to the fact that individuals make decisions using simple decision rules, i.e. heuristics, which, however, often generate biased decisions.[1]

The second approach to BR is often labelled as the 'fast-and-frugal-heuristics' (FFH) program and is associated with the research of Gerd Gigerenzer and his associates, who include Reinhard Selten, Peter Todd, Daniel Goldstein, Ralph Hertwig, Konstantinos Katsikopoulos and Nathan Berg.[2] In this program, utility maximization disappears and decisions are modelled as the result of simple and easily applicable heuristics. For instance, in the 'Take-The-Best' heuristic (Gigerenzer & Goldstein, 1996), the individual has an ordered vector of dimensions or 'cues' that she uses in a lexicographic way to discriminate between alternatives. If alternative A is superior to alternative B with respect to the best cue, A is chosen. If the best cue cannot discriminate between A and B, the two alternatives are compared using the second-best cue, and the alternative which is superior with respect to it is chosen. If the second-best cue cannot discriminate between A and B, the alternatives are compared using the third-best cue and so on until a discriminating cue is found and a choice is made. These heuristics have a local nature, that is, they change when the decision environment changes, but in their 'native' environment, they lead to efficient rather than biased choices.[3]

The third and most recent approach to BR we consider here still lacks a generally accepted label. Paola Manzini and Marco Mariotti, two of its main exponents, call it the 'model-based approach' to BR (Manzini and Mariotti, this issue). We take the liberty to call it the 'heuristics-and-revealed-preference' (HRP) program, as this name seems to better highlight the program's main features. As in the first two approaches to BR,

decisions are conceived as determined by some heuristic. What is peculiar to the HRP program, however, is that the choices generated by a given heuristic are characterized in terms of axioms of revealed preference. The characterizing axiom is less restrictive than the weak or strong axioms associated with rational choices, but still rules out a significant array of choice patterns. For instance, in the 'Categorize Then Choose' model, Manzini and Mariotti (2012) conceive decisions as the result of a two-stage process. In the first stage, the decision-maker categorizes the alternatives in broad classes and focuses on one class; in the second stage he chooses an alternative from that class. For example, the decision-maker first categorizes restaurants by type of cuisine and focuses on, say, Mexican restaurants; then he chooses the preferred Mexican restaurant. Manzini and Mariotti show that if an agent makes decisions according to a Categorize-Then-Choose heuristic, then his choices will display a specific consistency feature that can be characterized in terms of a revealed preference axiom they call weak WARP. The HRP program offers a straightforward way to test experimentally the validity of a BR model: if the choices recorded in the experiment violate the consistency features characterizing the model of BR under test, then the model is 'falsified'. In our example, if the decision-maker's choices violate the weak WARP, this suggests that he does not decide according to the Categorize-Then-Choose heuristic.[4]

Advocates of the three BR programs just discussed have sometimes been critical of the other approaches. The oldest and probably most known dispute in the field is the one between Gigerenzer (1991, 1996) and other supporters of the FFH program on the one side, and Kahneman and Tversky (1996) and other advocates of H&B on the other side. At the descriptive level, Gigerenzer has argued that the heurists described by Kahneman and Tversky are too vague to have actual explanatory power. At the normative level, Gigerenzer has criticized the heuristic-and-biases program for sticking to RCT as a normative model of rationality (see the articles by Berg, Katsikopoulos and Hands in this issue, Julien and Vallois 2014, and Vranas, 2000). Berg, also an exponent of the FFH program, criticizes Manzini and Mariotti's program for focusing on the internal consistency of the decision-maker's choices, as expressed by some axiom of revealed preference, as the unique criterion to assess and characterize her bounded rationality (Berg, this issue). In turn, exponents of the HRP program criticize the decision models put forward in the other two approaches because, they argue, their models lack precise implications in terms of choice data and, therefore, cannot be falsified in laboratory experiments (see Manzini & Mariotti, 2007; Spiegler, 2008).

2. Overview of the contributions

The five contributions collected in this special issue discuss important methodological aspects of these recent theories of BR. Practitioners, historians and methodologists of economics reflect on its most recent developments. Since the literature on behavioural economics and the H&B program is already copious, we asked our contributors to pay special attention to the other two approaches to BR mentioned above, namely the FFH and the HRP programs. We were especially interested in unravelling similarities and differences with respect to their modelling choices and the role they assign to evidence of different kinds, and with respect to their stance on the normativity of rationality assumptions. We now give a brief summary of the five papers.

In their article, Manzini and Mariotti explain how the HRP program can be applied to welfare analysis. BR is problematic for welfare analysis because, if the agent's choices are generated by some boundedly rational procedure, it is no longer obvious that these choices

should be used as a basis for welfare policies. In contrast to the BR approach to welfare economics put forward by Douglas Bernheim and Antonio Rangel (Bernheim, 2009; Bernheim & Rangel, 2009), Manzini and Mariotti argue in favour of modelling explicitly the boundedly rational cognitive processes (such as the Categorize-Then-Choose heuristic mentioned above) that may have generated the choice data. In particular, Manzini and Mariotti claim that, though decision-making models should always give priority to choice data and be characterized in terms of some revealed preference axiom, also non-choice data can be used to construct an appropriate model of decision-making and discriminate between different plausible models. Among non-choice data, Manzini and Mariotti mention physiological and neurophysiological data recorded during choice tasks, or survey responses and verbal evidence on internal states related to choice tasks. Then Manzini and Mariotti call attention to three main problems associated with the HRP program: (i) two or more different BR models of decision-making may be undistinguishable in terms of choice data alone; (ii) even if only one BR decision model is compatible with choice data, the model's primitives may have multiple interpretations; (iii) even if there is only one BR decision model compatible with choice data and only one accepted interpretation of its primitives, these primitives may be specified in multiple ways. In the final part of their paper, Manzini and Mariotti discuss how these problems can be mitigated.

Both Konstantinos Katsikopoulos and Nathan Berg have contributed to the FFH research program, and in their contributions to this issue they discuss the methodological features of this program in comparison to those of the other two approaches to BR. In particular, Katsikopoulos argues that there exist two distinct cultures of research on BR, the 'idealistic culture', which can be broadly associated with the H&B program, and the 'pragmatic culture', which characterizes the FFH program. Katsikopoulos does not discuss explicitly the HRP program, although many of its features seem to be more consonant with the idealistic rather than the pragmatic spirit. For Katsikopoulos, the two cultures differ with respect to both their underlying assumptions and their methodologies. The idealistic culture builds its models from criteria of internal consistency and models choice as the result of optimization, even though it does not assume the optimization process to be real but only 'as-if'. The pragmatic culture, in contrast, builds its models from simple rules that purportedly describe actual mental processes and models choice as an actual satisficing process. Furthermore, while the 'idealists' build models with multiple free parameters, mainly seeking to explain past choices, the 'pragmatists' build models with few if any free parameters, with the aim of making out-of-sample predictions. Finally, the two cultures also offer different 'stories'. The idealistic culture purports that people systematically behave irrationally and that, consequently, they should do better. The pragmatic culture, in contrast, maintains that people do well if they learn to use the right tool in the right situation. Neither of these two 'stories' can be tested empirically. Rather, each constitutes a general perspective, an underlying model, a *Weltanschauung*. Instead of empirical tests, a critical evaluation should therefore focus on other aspects, for example the psychological reactions of the public. Simply put, according to Katsikopoulos, the public reacts to the idealistic culture with frustration, delegating choice to supposed experts. In contrast, the public reacts to the pragmatic culture with a feeling of empowerment.

Berg puts forward a methodological taxonomy between different normative approaches to BR and distinguishes between the 'ecological-rationality approach', which can be identified with the FFH program, and the 'consistency approach', which characterizes RCT in the first place but has been imported into both the H&B and the HRP programs. In particular, for Berg the consistency approach explicitly adopts a single

normative criterion to evaluate a decision procedure, namely its internal consistency as characterized by some set of axioms on preferences (such as transitivity) or choices (such as the WARP). In this tradition, what distinguishes a rational from a boundedly rational decision is that the internal consistency conditions characterizing rational choice are more stringent than those characterizing BR. However, Berg argues, there is no empirical evidence that would confirm a close connection between the capacity of a decision procedure to generate well-being and its internal consistency. Moreover, the consistency approach implicitly relies also on normative criteria that go beyond the procedure's internal consistency, such as money pump arguments, and this creates a tacit methodological tension within it. In contrast to the consistency school, the ecological-rationality approach makes explicit use of a plurality of normative criteria to evaluate a decision procedure. These criteria focus on the procedure's performance in terms of well-being, as measured, for instance, by the average or cumulative payoffs it generates in the specific context in which the procedure is used. These criteria are external to the procedure itself and have a context-dependent nature. Therefore, they contrast with the consistency criterion, which is internal to the procedure itself and aims at abstract and de-contextualized validity. In the last section of his paper, Berg discusses Manzini and Mariotti's HRP program as an instance of the consistency approach to BR.

In his article, Wade Hands compares the normative theories of rationality associated with RCT and the FFH programs. He first identifies four main naturalistic arguments, that is, arguments based on purported empirical evidence, which have been used to defend the normative validity of RCT: (i) the purported willingness of people and possibly experts to adjust their *actual* choices in the light of RCT, (ii) the purported willingness of people (or experts) to adjust their *hypothetical* choices in the light of RCT, (iii) the purported material disadvantages that arise from violating RCT (e.g. money pump, Dutch books) and (iv) the purported ability of people to comply with the requirements of RCT. Then he discusses whether these arguments support the superiority of the FFH program over RCT as a normative theory of decision-making. Hands makes two general points about the relationship between RCT and the FFH program: first, both approaches share the same instrumental view of rationality and second, the defence of FFH in terms of generic evolutionary stories is fragile since similar evolutionary stories can be employed to defend RCT. Hands argues that there is little evidence showing either actual or hypothetical adoption of fast-and-frugal heuristics over RCT rules, and hence neither (i) nor (ii) supports the normative validity of the FFH program over RCT. Unlike Berg (this issue), Hands claims that not even (iii) is supported, as little evidence exists for sustained advantages in terms of well-being deriving from the adaptation of fast-and-frugal heuristics. Only with respect to (iv) the FFH program seems to do better than RCT, but Hands claims that that the empirical evidence on this point is still weak. Thus, Hands concludes that, at least from the viewpoint of the four naturalistic arguments considered, the advantages of the fast-and-frugal heuristics program over RCT as a normative theory of rational decision-making seem rather weak.

If Hands' article may be seen as a temperate defence of RCT, in the last article of this special issue Don Ross offers a bold epistemological justification for the persisting dominance of RCT in economics. Ross opens his piece by acknowledging that 'people are *obviously* not boundedlessly rational agents, since nothing is or could be such an agent', and then asks 'why, if economics is an empirical science, do economists introduce bounds on the rationality of agents in their models only cautiously and partially?' To answer this question, Ross constructs an elaborated argument whose main building blocks might be summarized as follows: (i) unlike psychology, economic theory is primarily about

aggregate markets phenomena rather than individual decisions, (ii) the economic theory of markets is not significantly modified if individual decisions are modelled using some BR theory rather than RCT, (iii) BR individual decision processes are very much influenced by institutional and informational properties specific to the market in question, and these properties change from market to market and (iv) there is not and there cannot be a *general* model explaining how BR decision processes depend on the institutional and informational properties of market structures. Assuming that Ross' four claims are correct (though arguably a number of BR theorists and economic methodologists would not agree with many of them), the fact that mainstream economists still stick to RCT appears epistemologically justified. The perspective adopted by Ross allows him not only to rationalize the current state of mainstream economics, but also to predict that 'no *general* model of bounded rationality should ever be expected to feature in the economist's toolkit, regardless of the extent to which psychologists successfully identify specific human cognitive limitations'.

3. Descriptive issues in BR theories

The five contributions in this special issue give a clear sense that at the moment there exists a plurality of approaches to BR that diverge not only from a theoretical viewpoint but also at the methodological level. This is why we prefer to talk of 'methodologies of bounded rationality', rather than of a single methodology. The main differences between the various BR methodologies that we single out as emerging from the five papers concern: (i) the kind of evidence deemed relevant in the assessment of the descriptive accuracy of BR models, (ii) the modelling desiderata that a BR model should display and (iii) the normative validity of BR models, that is, the acceptability of their 'ought' statements. In this section we examine the first two differences, related to the descriptive validity of the models, while in the next section we focus on issues of normative validity.

3.1 Criteria of evidential relevance

A theme that resurfaces across many of the articles in this special issue concerns the relevance, or lack thereof, of certain kinds of evidence in adjudicating between competing decision models.

Choice versus non-choice data. According to an important tradition within mainstream economics, usually associated with Samuelson (1938), and whose latest vocal exponents are Gul and Pesendorfer (2008), the economic theory of individual decision-making is not concerned with psychological mechanisms but with individuals' observable choices. This behaviourist view of decision analysis is by no means shared by all mainstream economists and has been criticized by a number of economic methodologists, beginning with Robbins (1935).[5] The issue of whether only choice data should be relevant for the confirmation of economic models concerns not only mainstream models but also BR models, and, as we have seen, there is an array of different methodological stances on this point.

Some BR theorists, such as Bernheim and Rangel (2009), maintain that only choice data are relevant for decision analysis. Others occupy less radical positions. For example, Manzini and Mariotti (this issue) argue that choice data have a priority over other types of data, and that BR models should be characterized in terms of revealed preference axioms. However, they also make room for non-choice data, such as neurophysiological data or survey responses, in order to understand the psychological processes behind choices and to model them appropriately. In particular, non-choice data play a role in addressing two

problems associated with the HRP program. First, when two or more BR models of decision-making are undistinguishable in terms of choice data, non-choice data can play an important role in discriminating between them. Second, when only one BR decision model is compatible with choice data but its primitives have multiple interpretations, non-choice data may help in selecting the most plausible interpretation of the primitives.

In the FFH program, choice data do not seem to have any special status vis-à-vis other types of data. Data from psychological introspection, physiological or neurophysiological research, interviews and responses to surveys seem to be as important as choice data to assess whether a given BR model is descriptively accurate. The divide between choice and non-choice data is also relevant to Ross' claim (this issue) that mainstream economists are epistemologically justified in sticking to RCT. His methodological position – which he labels 'Neo-Samuelsonian' – is a variation of the behaviourist view of decision analysis described above. Ross (2014 and this issue) agrees with Gul and Pesendorfer that economics is about observable choices, but disagrees with them in so far as for him the domain of economic theory is not individual decision-making but aggregate market phenomena. In Ross's perspective, it is only aggregate choice data that count, and therefore there is not much to gain in testing economic models against psychological data, let alone neurophysiological data.

Experimental versus non-experimental evidence. Binmore (1999) has famously claimed that laboratory evidence of failures of rationality cannot falsify game theory models when the tasks and environments in which individuals are placed are too artificial, leave no room for learning and incentives are too low. Similarly, Gigerenzer and colleagues sometimes complain that many of the behavioural laboratory experiments the H&B program uses cannot unambiguously test theories of BR. They claim that if the heuristics people use are adapted to natural environments, then, unless the laboratory succeeds in capturing the features of such environments, it cannot tell us much about the heuristics people actually use (e.g. Gigerenzer, 1991; see Kahneman & Tversky's response, 1996). In contrast, Manzini, Mariotti and other contributors to the HRP program focus almost exclusively on experimental data and regard non-experimental evidence as problematic because, they claim, it lacks the kind of control needed for theory testing (see, e.g. Manzini, Mariotti, & Mittone, 2010).

Within sample-fit versus out-of-sample prediction. Berg and Gigerenzer (2010) contend that H&B models fair well in terms of within-sample fit, but are in no way superior to other approaches (including RCT) when it comes to out-of-sample predictions. According to Berg and Gigerenzer (2010), however, out-of-sample prediction is a more severe test for a decision model (see also Berg, this issue; Katsikopoulos, this issue). In particular, the exponents of the FFH program typically claim that their models are superior to RCT and other BR models with regard to out-of-sample predictions.

Technological success. By technological success we mean the implementation of a theory or model into a successful intervention, namely an intervention that brings about the desired outcome. Technological success speaks in favour of a theory or model in that a necessary (though not sufficient) condition for the success of an intervention is that the theory gets the causal relationship on which the intervention is based right (see, e.g. Cartwright, 2009). The use of insights from the H&B program for the design of apparently successful policy interventions such as those introduced by the Behavioural Insight Team in the UK[6] can be regarded as testifying in favour of the program's empirical performance. Fast-and-frugal heuristics have been less popular among policy-makers than nudges. However, the observation that decision tools based on fast-and-frugal heuristics lead to better performance in terms of well-being (see Berg, this issue) has been taken as speaking

in favour of FFH models. However, it is unclear whether successful interventions do speak in favour of one or another approach in any general sense because the connection between theory and policy is often tenuous at best.

3.2 Modelling desiderata

Proponents of different BR models place emphasis on different modelling desiderata. From the articles in this special issue we identify three such desiderata: mechanistic detail, generality and parsimony. Whether or not these desiderata, as they are conceived by BR scholars, justifiably count as virtues, and what kind of virtues they are, is not an issue we will deal with here. However, we are inclined towards a pluralist stance according to which models can have several desiderata, some of which may trade off. This view implies that no single model is likely to display all desiderata at once (see, e.g. Gabaix & Laibson, 2008; Levins, 1966; Matthewson & Weisberg, 2009).

Mechanistic detail. The inclusion of mechanistic details is a desideratum to which all three programs appear to be committed. This is perhaps not surprising considering that building more psychologically realistic models than those of RCT has been one of the main motivations behind theories of BR. Even so, there seem to be different interpretations of what including mechanistic details into a model of decision-making entails. For example, Katsikopoulos' notion of *process model* can be read as an appeal to the idea that the more details about underlying psychological processes models include, the better are the models (Katsikopoulos this issue, see also Berg & Gigerenzer, 2010). Katsikopoulos claims that models of the idealistic culture, including many in the H&B program, do not represent such processes. Manzini and Mariotti's model is a special case. They explicitly claim that their models seek to represent 'the decision process that underlies choice' (Manzini and Mariotti, this issue). However, it appears that they do not give direct evidence for this process, and therefore we suspect that FFH representatives would not consider it a 'process model'.

Generality. An often-mentioned advantage of RCT over BR models is that the former applies to all decision situations, while the latter identifies a whole range of different decision rules tailor-made to cater for particular contexts but which are not always generalizable to other sets of circumstances. As we have seen, Ross (this issue) puts forward a version of this argument: since BR decision rules depend on the institutional and informational properties of markets, a general model of BR cannot be had and this justifies economists' use of RCT. In the FFH program, however, generality does not appear so fundamental as a modelling desideratum. As stressed, for example by Berg (this issue), FFH models are intrinsically *local*, in the sense that they state that in decision situations of a certain kind, a certain type of decision rule is adopted. It must be noted, however, that Gigerenzer and associates often appeal to generality to criticize H&B models: they argue that the latter kind of models are tailored to fit data obtained in artificial laboratory situations and cannot be generalized to actual environments outside the laboratory. Generality as a modelling desideratum appears to be more important within the HRP program. An important part of this program, in fact, is to check whether decision rules that capture very different BR psychological processes can be characterized in terms of the same revealed preference axioms. For instance, Manzini and Mariotti show that both their Categorize-Then-Choose heuristic and a different heuristic called Rationalization (Cherepanov, Feddersen, & Sandroni, 2013) generate choices that satisfy the weak WARP.[7]

Parsimony. Manzini and Mariotti (this issue) suggest that parsimony is a criterion for discriminating between observationally equivalent BR models. The principle of parsimony

has a venerable history in science. In its formulation as Ockham's razor, it recommends ontological parsimony. As a criterion of theory choice and a synonym of simplicity, it states that among competing theories the one employing the fewer assumptions should be chosen, *ceteris paribus*. Parsimony is often taken as a sign of truth, but why it is so remains a matter of philosophical debate. Alternatively, it is interpreted as a sign of predictive accuracy. To illustrate the idea behind parsimony as a criterion of selection, Manzini and Mariotti (this issue) take Selten's (1991) measure of predictive success as an example. This measure ranks models according to their descriptive power or 'hit rate' (i.e. the proportion of observed outcomes consistent with the model) minus their 'relative area' (i.e. the proportion of theoretically possible outcomes that are consistent with the model). The adoption of Selten's measure suggests that Manzini and Mariotti interpret parsimony as a measure of predictive accuracy of a model rather than of its truth. So interpreted, Manzini and Mariotti's position comes close to Gigerenzer and colleagues' complaint about the fact that H&B models have multiple free parameters (see Katsikopoulos, this issue). The idea is that the presence of many free parameters runs the risk of overfitting one's model in terms of a particular data-set and hence to make it a poor predictive tool. As Forster and Sober (1994) point out, parsimony can mitigate the risk of overfitting (for discussion in the context of BR models, see, e.g. Gigerenzer & Brighton, 2009).

4. Normative issues in BR theories

BR, although bounded, is still a kind of rationality. Consequently, it still makes normative claims – how one ought to reason, what one ought to do – only *within* its bounds. Some of the contributors to this issue, for example Hands and Berg, explicitly endorse this position, while others, for example Ross, strenuously seek to avoid addressing any normative implications altogether. We leave it open whether the concept of BR necessarily has normative implications. Instead, we discuss only those models that make normative claims. The questions that arise for these models include the following:

(1) What are the normative claims of BR?
(2) Is BR normatively less valid than unbounded rationality?
(3) How is the normative validity of BR justified?

In this section, we survey some possible answers to these three questions, found in the literature on rational decision-making. Not all of them were intended for models of BR, but rather for models of expected utility maximization. Yet, as we will argue, these answers can be applied, *mutatis mutandis*, to BR as well. So here go our three questions.

4.1 What are the normative claims of BR?

The normative claims of the various models of rationality (including BR) differ both in their target and their content. Regarding the differences in targets, some models impose normative constraints on mental states and reasoning rules or on revealed preferences. We call these *internal consistency accounts of rationality*. Manzini and Mariotti (this issue) defend such an internal consistency account of BR, as when they recommend their theory as being able to deal with 'very mild forms of inconsistency', while still satisfying sufficiently weakened internal consistency requirements. In contrast to these, other models impose normative constraints on actions, often conditional on particular environments, derived from various considerations of pragmatic success. We call these *external performance accounts of rationality* (see Berg, this issue). These two types of accounts are

further differentiated by what constraints they impose. Internal consistency accounts impose for example consistency criteria such as the law of excluded middle (already proposed by Aristotle), preference transitivity (von Neumann & Morgenstern, 1944) or the Kolmogorov probability axioms (Kolmogorov, 1933/1950). External performance accounts impose performance criteria such as robustly meeting a minimal payoff threshold (Simon, 1955), frugality of the decision rule (Gigerenzer & Todd, 1999) or sufficient approximation to the environment's cue structure (Martignon & Hoffrage, 1999).

4.2 Is BR normatively less valid than unbounded rationality?

By normative validity we mean the validity of the 'ought' statements of the model, in contrast to descriptive validity, which refers to the model's 'is' statements. The concept of BR is typically contrasted with a notion of unbounded rationality. This is not a conceptual necessity: perhaps the concept of rationality only makes sense under some kinds of constraints (Bayesian rationality under uncertainty, for example, assumes constraints on information; instrumental rationality, broadly understood, assumes scarcity of resources). Yet *de facto*, all BR models have been proposed against the foil of some model of rationality that is not similarly bounded (the paradigmatic example of this kind of model being the rational-choice model). The contrastive presentation does not imply that the BR model is necessarily less normatively valid than the unbounded model, however.

Based on the above dual distinction between internal and external criteria for both BR and unbounded rationality, we distinguish three cases (Table 1). If both unbounded and BR models are based on internal consistency (case A), then additional constraints of the BR model (for example, computational limitations) necessarily reduce the stringency of the criteria of BR. Consequently, the criteria of the BR model are easier to satisfy than that of the unbounded model. For example, Manzini and Mariotti's weak WARP is less stringent than WARP. Now, in terms of validity, there are two scenarios. If the internal consistency criteria of unbounded rationality are valid, then the BR model is normatively less valid than the unbounded model in that it satisfies fewer such criteria (case A_1). If, however, some of the internal consistency criteria of unbounded rationality are not valid, then the BR model, by relaxing those criteria, might be more valid than the unbounded model (case A_2).

If the unbounded rationality model is based exclusively on internal consistency criteria, but the BR model is not (case B), then the BR model, although it reduces the stringency of internal criteria, might satisfy external criteria better than the unbounded model. In this case, the BR model might be normatively more valid than the unbounded

Table 1. Internal consistency and external performance criteria.

Account of full rationality \ Account of BR	Internal consistency criteria	External performance criteria
Internal consistency criteria	(A1) BR less valid, because BR is subject to lesser degree of internal consistency (A2) BR more valid if more stringent criteria are not themselves valid	(B) BR sometimes more valid, when BR gives better performance results than full internal consistency

model. For example, the FFH normative model is based on external criteria, and it may well be that on such criteria it fares better than the unbounded model even though the internal criteria of the FHH model are less stringent. This is precisely the kind of claim Berg makes in his contribution to this issue.

In sum, there is one case (A_1) in which the unbounded model is normatively more valid than the BR model of rationality, given that both models are based on the same kind of validity criteria *and* all the criteria by which the unbounded model is characterized are indeed valid. Conversely, there are two kinds of cases (cases A_2 and B) in which the BR model might be more valid than the unbounded model: either when the two models are based on different kinds of criteria, *or* when they are based on the same kind of validity criteria, but the criteria by which the unbounded model is characterized are not valid.

4.3 How is the validity of BR justified?

So far, we characterized the nature of normative validity criteria and the consequences for both models of BR and unbounded rationality. In this section we instead turn to the methods by which criteria of normative validity are sought to be justified (see Hands, this issue).[8] We distinguish five methods of validation found in the literature:

(a) derived from conceptual analysis,
(b) derived from universal loss-avoidance considerations,
(c) based on unambiguously normatively exemplary empirical cases,
(d) derived from narrow (d_1) and wide (d_2) reflective equilibrium between intuitive judgements and purported principles and
(e) inferential coherence meta-criteria.

We briefly describe each of these methods in turn. Note that we do not endorse any particular method of validation. The point is to map the conceptual terrain in which to place various arguments about the comparative validity of competing models of rationality.

Normative validation criteria derived from conceptual analysis (a). This argument for the normative appropriateness of certain internal consistency criteria suggests that these criteria are constitutive of the meaning of preference, belief or intention. Take for example preference transitivity. Drawing an analogy to length measurement, Davidson (1976/1980, p. 273) asks:

> If length is not transitive, what does it mean to use a number to 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'.

Violating transitivity, Davidson claims, thus undermines the very meaning of preferring one option over others. Consequently, to the extent that one prefers anything at all to something else, one must satisfy the transitivity criterion. In a similar vein, Robbins had earlier argued that minimal criteria of preference ordering are logical prerequisites for economic action. 'The main postulate of the theory of value – Robbins (1935, pp. 78 and 79) wrote – is the fact that individuals can arrange their preferences in an order, and in fact do so.' This fundamental postulate is 'an essential constituent of our conception of conduct with an economic aspect' (p. 75). According to Robbins, '[w]e do not need controlled experiments to establish their validity: they are so much stuff of our everyday experience

that they only have to be stated to be recognised as obvious' (p. 79). Hence, such fundamental postulates do not need any further justification.

In sum, according to this method, one's representational tools must satisfy certain criteria (e.g. Davidson's transitivity criterion, or Robbins' preference ordering) if one aims to represent economic behaviour at all.

Normative validation criteria derived from universal loss-avoidance considerations (b). A different argument for the validity of consistency criteria is derived from considerations of the negative consequences incurred when these criteria are violated. Such considerations show that violating consistency of beliefs or preferences can lead to situations where the violator incurs a sure loss. The most famous argument of this kind is the so-called 'Dutch Book' argument. The basic idea was developed by Ramsey (1928/1950, p. 182), who pointed out that if a subject's behaviour violates the axioms of probability, then '[h]e could have a book made against him by a cunning better and would then stand to lose in any event'.[9] If an agent violates the transitivity condition on preferences, then that individual can be 'money pumped': all wealth can be taken from her, simply by trading goods with her in a way that exploits her preference intransitivity (for discussion of these arguments, see Berg, this issue and Hands, this issue). Consequently, to the extent that any one wants to avoid such sure losses, one must satisfy the corresponding internal consistency criteria.

Normative validation criteria based on unambiguously normatively exemplary cases (c). A very different argument for normative validity claims that rationality models should capture the reasoning and decision-making of those who are most competent and successful in the relevant domain. In the history of decision theory, such arguments have been employed – with rather different purposes – by Daniel Bernoulli, Condorcet and Maurice Allais. Each of these authors argued for or against the normative validity of EUT by showing 'that it captures [or does not capture] the decision rule actually applied by those considered wisest in making choices under uncertainty' (Jallais, Pradier, & Teira, 2008). Specifically, Bernoulli cited the reasoning and decision-making of businessmen, gamblers and insurers to support his claim that EUT was the normatively most valid theory of decision-making under uncertainty. In contrast, Condorcet cited the reasoning and decision-making of the same kind of experts to show that EUT was *not* normatively valid. Allais follows Condorcet in citing such empirical evidence against EUT, but for him, the relevant experts come from a different domain: namely, scientists in high standing who are well versed in probability theory (Jallais et al., 2008). Consequently, for any model of rationality – whether bounded or unbounded – to be normatively valid, it needs to be shown that it represents the most competent and successful decision-making in the relevant domain.

External performance accounts of BR (cf. Berg, this issue) need to address the following question about normatively exemplary cases: is the average or cumulative payoffs we observe really generated by the application of certain heuristics, or is it a matter of chance or other factors? One can answer this question only by explaining why the observed behaviour should be considered normatively exemplary, and the employed measure not merely accidentally inflated. Hand's criticism (this issue) sets in exactly here. He argues that there is little evidence showing that *relevant actors* either actually or hypothetically adopt fast-and-frugal heuristics over RCT rules, and therefore he concludes

that the normative validity of the FFH program over RCT has not been established on these grounds.

Normative validation criteria derived from narrow reflective equilibrium (d_1). The above accounts of a method of normative justification rely on the existence of such reference points as conceptual analysis, pragmatic consequences or exemplary cases. It is not obvious that any of these methods is available or sufficiently reliable to base the normative justification of a particular model of rationality. By contrast, reflective equilibrium approaches take two (or more) reference points into account, allowing for the fallibility of either reference point or of both. The assumption is that despite their fallibility, these reference points, when allowed to correct each other, will contribute to an overall coherent account of normatively valid rationality criteria. We distinguish a *narrow* and a *wide* notion of reflective equilibrium, depending on the number and kind of reference points taken into account.

On the *narrow* notion of reflective equilibrium, (i) intuitively plausible normative criteria are exposed to (ii) actually observed, common inferential and decision-making practices. Typical proponents of such a narrow notion are Marschak (1951), Savage (1954/1972) and Cohen (1981). Savage, in response to Allais's empirical result that the majority of experimental subjects violated the sure-thing principle (an internal consistency requirement on preferences over lotteries), argued that:

> If, after thorough deliberation, anyone maintains a pair of distinct preferences that are in conflict with the sure-thing principle, he must abandon, or modify, the principle; for that kind of discrepancy seems intolerable in a normative theory. (Savage, 1954/1972, p. 102)

Savage recognized that the consistency requirements of EUT were in principle revisable in the light of counterexamples. Then, he identified normativity with the *convincing power* of a theory, and proposed his test. Notably, Savage's method was not founded on any formal (however minimal) definition of rationality (see Guala, 2000, p. 72).

This account of reflective equilibrium is *narrow*, because it only aims to bring to equilibrium intuitive normative principles on the one hand, and empirical counter-examples on the other. In this way, Cohen (1981) claims, the procedure is similar to devising a grammar that fits a population's linguistic practice: it commences from principles that are considered normatively valid ('how one should speak'), but these principles are revised in the light of certain counterexamples of how people really do speak.

Normative validation criteria derived from wide reflective equilibrium (d_2). Wide reflective equilibrium, in addition to the normative principles and relevant examples of actual reasoning and choice practices, also takes into account a number of background information and theories, specifically about cognitive capacities and limitations, goals of inferential behaviour and philosophical theories. One important motivation for this inclusion is the aspiration to provide something more than merely a grammar of reasoning: the wide equilibrium should yield inferential principles that are *optimal* given goals and philosophical theories, but nevertheless feasible within the given limitations (Thagard, 1982, p. 35). Simply matching normative principles and counterexamples does not suffice for this aspiration.

Such a theory of wide equilibrium has been proposed mainly in ethics (see Daniels, 1979; Rawls, 1971), but some people (e.g. Goldman, 1978) have argued that

considerations of this sort are relevant also in epistemology and logic. Amongst the contributors to this special issue, Manzini and Mariotti as well as Hands pursue such a wide equilibrium. Manzini and Mariotti (this issue, our emphasis) insist that '*the reason for individual choice is important to us*, besides choice itself [. . .] we seek to understand the mechanism generating individual choice'. They thus seek to incorporate into their model not only normative principles such as weak WARP and counterexamples of choice violating SARP, but also reasons and purposes pursued with the choice. Hands (this issue) cautiously concedes that considerations of 'ought implies can' might give FFH models an advantage in normative validity over RCT models.

However, the problem with wide reflective equilibrium is the possibility of multiple equilibria. Depending on their own sense of diligence and exactitude, some people might be too readily willing to settle on principles that are clearly fallible. This raises the question, whose equilibrium should count? There are two possible answers, a populist and an elitist one. The populist strategy, favoured by Cohen (1981), is to emphasize the reflective equilibrium of the *average person*. This strategy founders, because education in sophisticated inferential techniques can be expected to provide the individual with a much more efficacious system. The elitist strategy, favoured by Stich and Nisbett (1980), is to emphasize the reflective equilibrium of *experts*. This too is inadequate, for it leaves us no way of saying why the experts should be in equilibrium, or of mediating disputes among experts (see Thagard, 1982).

Both narrow and wide equilibrium give a considerable role to intuitive judgements. However, in contrast to the conceptual analysis approach of method (a), narrow and wide equilibrium approaches adopt an explicitly fallibilist position towards normative intuitions: informal notions of rationality play a regulative role in assessing formal rationality criteria, but they are corrigible in the light of *some* counterexamples. In contrast to the exemplary cases approach (c), formal rationality criteria are not exclusively built on empirical case studies. The intuitive notions retain a guiding and regulating role, while the counterexamples are useful for theory improvement, but not every example is allowed to falsify a theory (see Guala, 2000, pp. 70 and 71).

Normative validation criteria derived from inferential coherence meta-criteria (e). Thagard (1982) explicitly criticizes the exemplary cases approach (c) and the equilibrium approaches (d_1) and (d_2) as inadequate methods to get from the psychology of reasoning and decision-making to normatively valid principles of deductive, inductive and practical reasoning. Instead of relying on finding some equilibrium between the various reference points as in (d_1) and (d_2), he suggests that the process of arriving at normatively valid criteria of reasoning and decision-making should be governed by a number of inferential meta-criteria. Specifically, he proposes three such criteria:

(1) Robustness: to what extent do the normative principles account for inductive practice?
(2) Accommodation: to what extent do background theories account for deviations of inductive practice from the normative principles?
(3) Efficacy: given background theories, to what extent does following the normative principles promote the satisfaction of the inferential goals?

The third meta-criterion, in particular, distinguishes Thagard's account from internalist coherence accounts such as those derived from conceptual analysis (a) and from loss-avoidance consideration (b). Furthermore, this account differs from (d_1) and (d_2)

in that equilibrium might be a consequence of satisfying these meta-criteria, but it need not be. In the case that a non-equilibrium set of beliefs better satisfies these criteria than an equilibrated set, the former should be chosen:

> Coherence is to be evaluated according to criteria to which the achievement of reflective equilibrium is irrelevant. What we are really after is not equilibrium, but progress: the development of better and better inferential systems. (Thagard, 1982, pp. 39 and 40)

This of course raises the question how these meta-principles themselves are justified. Thagard answers with a strong naturalist position on normativity: 'What in turn justifies these? They seem to be the ones actually used when we set out to evaluate inferential practices' (Thagard, 1982, p. 40). It is noteworthy that none of the authors contributing to this issue seeks to justify the normative validity of their respective models with such a set of meta-criteria of coherence.

5. Concluding remarks

In this introduction, we examined some of the questions concerning the descriptive accuracy and normative validity of alternative approaches to BR, which we thought were raised by the articles collected in this special issue. In particular, we focused on ideas about the relevance of different kinds of evidence, preferences for different modelling desiderata and alternative methods for the justification of the normativity of theories of rationality. We deliberately refrained from taking a stance in favour of one or the other conception, preference or justificatory method. Our hope is that the special issue as a whole will contribute to stimulate further reflections on these important issues among philosophers of economics and practitioners.

Funding

Till Grüne-Yanoff thanks Vetenskapsrådet for financial support. Caterina Marchionni thanks the Academy of Finland for financial support. Ivan Moscati thanks Fondazione Cariplo for financial support.

Notes

1. Some of the main contributions within the behavioural economics approach to BR are collected in Kahneman and Tversky (2000), Camerer, Loewenstein, and Rabin (2004), and Loewenstein (2007).
2. An alternative label for the FFH program is the 'simple heuristics' program. However, in this introduction we follow the terminology adopted by some of the contributors to the present issue and thus use the FFH label.
3. Some significant contributions to the FFH program are collected in Gigerenzer, Todd, and the ABC Research Group (1999), Gigerenzer and Selten (2001), Gigerenzer, Hertwig, and Pachur (2011), and Todd, Gigerenzer, and the ABC Research Group (2012).
4. Some important contributions to the HRP research program are Kalai, Rubinstein, and Spiegler (2002), Rubinstein and Salant (2006, 2012), Manzini and Mariotti (2007, 2012, 2014), Mandler, Manzini, and Mariotti (2012), Masatlioglu, Nakajima, and Ozbay (2012), and Cherepanov et al. (2013). Important predecessors of this approach can be found in Luce (1956, 1978). Within FFH, some authors have also proposed axiomatizations, albeit in terms of preferences, nor revealed choices (e.g. Drechsler, Katsikopoulos, and Gigerenzer, 2014).
5. One of the main reasons why the behaviourist view appears inadequate to capture what mainstream economists in fact do is that fundamental parts of decision analysis rely on psychological elements that can hardly be inferred from choice data. The single most important example is game theory, which refers to unobservable variables such as interactive beliefs or out-of-equilibrium strategies. For a methodological discussion of the behaviourist view of

decision analysis in its Gul-Pesendorfer version, see Hausman (2008), Caplin (2008), Spiegler (2008), and Moscati (2010). In historical perspective, Moscati (2013) shows that the conventional portrait of Samuelson as the early champion of the behaviourist decision analysis is misleading.

6. See http://www.gov.uk/government/organisations/behavioural-insights-team.
7. As we have seen, in Categorize Then Choose the agent first simplifies his choice task by categorizing the alternatives and focusing on one single class, and then chooses the most preferred alternative from that class. In Rationalization, the agent first selects the alternatives that he/she can justify on the basis of some rationale (e.g. alternatives that are 'morally acceptable'), and then chooses the most preferred alternative from the rationalized ones. From a psychological viewpoint, the first stages of Categorize Then Choose and Rationalization are significantly different.
8. Hands (this issue) employs naturalistic criteria of justification. Our criteria partially coincide with his. We will clarify the connection later on in this section.
9. The argument was independently developed in more detail by de Finetti (1931/1993). For decisions under certainty, a similar argument has been developed by Davidson, McKinsey, and Suppes (1955).

References

Berg, N., & Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*, *18*, 133–165.

Bernheim, B. D. (2009). Behavioral welfare economics. *Journal of the European Economic Association*, *7*, 267–319.

Bernheim, B. D., & Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics*, *124*, 51–104.

Binmore, K. (1999). Why experiment in economics? *Economic Journal*, *109*, F16–F24.

Camerer, C. F., Loewenstein, G., & Rabin, M. (Eds.). (2004). *Advances in behavioral economics*. Princeton, NJ: Princeton University Press.

Caplin, A. (2008). Economic theory and psychological data: Bridging the divide. In A. Caplin & A. Schotter (Eds.), *The Foundations of positive and normative economics: A handbook* (pp. 336–371). New York, NY: Oxford University Press.

Cartwright, N. (2009). *Hunting causes and using them*. Cambridge: Cambridge University Press.

Cherepanov, V., Feddersen, T., & Sandroni, A. (2013). Rationalization. *Theoretical Economics*, *8*, 775–800.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, *4*, 317–331.

Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy*, *76*, 256–282.

Davidson, D. (1976/1980). Hempel on explaining action. In D. Davidson (Ed.), *Essays on actions and events* (pp. 261–275). Oxford: Oxford University Press.

Davidson, D., McKinsey, J., & Suppes, P. (1955). Outlines of a formal theory of value, I. *Philosophy of Science*, *22*, 140–160.

de Finetti, B. (1931/1993). On the subjective meaning of probability. In P. Monari & D. Cocchi (Eds.), *Induction and probability* (pp. 291–321). Bologna: Clueb.

Drechsler, M., Katsikopoulos, K., & Gigerenzer, G. (2014). Axiomatizing bounded rationality: The priority heuristic. *Theory and Decision*, *77*, 183–196.

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science*, *45*, 1–35.

Gabaix, X., & Laibson, D. (2008). The seven properties of good models. In A. Caplin & A. Schotter (Eds.), *The foundations of positive and normative economics: A handbook* (pp. 292–299). New York, NY: Oxford University Press.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology*, *2*, 83–115.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*, 592–596.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased make better inferences. *Topics in Cognitive Science*, *1*, 107–143.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.
Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). *Heuristics: The foundations of adaptive behavior*. New York, NY: Oxford University Press.
Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality*. Cambridge, MA: MIT Press.
Gigerenzer, G., & Todd, P. M. (1999). *Fast and frugal heuristics*. New York, NY: Oxford University Press.
Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.
Goldman, A. I. (1978). Epistemics: The regulative theory of cognition? *Journal of Philosophy*, *75*, 509–523.
Grüne-Yanoff, T. (2007). Bounded rationality. *Philosophy Compass*, *2*, 534–563.
Guala, F. (2000). The logic of normative falsification: Rationality and experiments in decision theory. *Journal of Economic Methodology*, *7*, 59–93.
Gul, F., & Pesendorfer, W. (2008). The case for mindless economics. In A. Caplin & A. Schotter (Eds.), *The foundations of positive and normative economics: A handbook* (pp. 3–39). New York, NY: Oxford University Press.
Harstad, R. M., & Selten, R. (2013). Bounded-rationality models: Tasks to become intellectually competitive. *Journal of Economic Literature*, *51*, 496–511.
Hausman, D. (2008). Mindless or mindful economics: A methodological evaluation. In A. Caplin & A. Schotter (Eds.), *The foundations of positive and normative economics: A handbook* (pp. 125–151). New York, NY: Oxford University Press.
Jallais, S., Pradier, P. C., & Teira, D. (2008). Facts, norms and expected utility functions. *History of the Human Sciences*, *21*, 45–62.
Jullien, D., & Vallois, N. (2014). A probabilistic ghost in the experimental machine. *Journal of Economic Methodology*, *21*, 232–250.
Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582–591.
Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press.
Kalai, G., Rubinstein, A., & Spiegler, R. (2002). Rationalizing choice functions by multiple rationales. *Econometrica*, *70*, 2481–2488.
Kolmogorov, A. N. (1933/1950). *Foundations of the theory of probability*. New York, NY: Chelsea Publishing Company.
Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*, 421–431.
Loewenstein, G. (2007). *Exotic preferences*. New York, NY: University Press.
Luce, R. D. (1956). Semi-orders and a theory of utility discrimination. *Econometrica*, *24*, 178–191.
Luce, R. D. (1978). Lexicographic trade-off structures. *Theory and Decision*, *9*, 187–193.
Mandler, M., Manzini, P., & Mariotti, M. (2012). A million answers to twenty questions: Choosing by checklist. *Journal of Economic Theory*, *147*, 71–92.
Manzini, P., & Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, *97*, 1824–1839.
Manzini, P., & Mariotti, M. (2012). Categorize then choose: Boundedly rational choice and welfare. *Journal of the European Economic Association*, *10*, 1141–1165.
Manzini, P., & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, *82*, 1153–1176.
Manzini, P., Mariotti, M., & Mittone, L. (2010). Choosing monetary sequences: Theory and experimental evidence. *Theory and Decision*, *65*, 327–354.
Marschak, J. (1951). Why "should" statisticians and businessmen maximize "moral expectation"? In *Proceedings of the second Berkeley symposium on mathematical statistics and probability* (pp. 493–506). Berkeley: University of California Press.
Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). New York: Oxford University Press.
Masatlioglu, Y., Nakajima, D., & Ozbay, E. (2012). Revealed attention. *American Economic Review*, *102*, 2183–2205.

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*. New York, NY: Oxford University Press.
Matthewson, J., & Weisberg, M. (2009). The structure of tradeoffs in model building. *Synthese*, *170*, 169–190.
Moscati, I. (2010). Review of Caplin and Schotter 2008. *Economics and Philosophy*, *26*, 101–108.
Moscati, I. (2013). How cardinal utility entered economic analysis, 1909–1944. *European Journal of the History of Economic Thought*, *20*, 906–939.
Ramsey, F. P. (1928/1950). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (pp. 156–198). London: Routledge & Kegan Paul.
Rawls, J. (1971). *A theory of justice*. Cambridge: Harvard University Press.
Robbins, L. (1935). *An essay on the nature and significance of economic science*. London: Macmillan.
Ross, D. (2014). *Philosophy of economics*. New York, NY: Palgrave Macmillan.
Rubinstein, A., & Salant, Y. (2006). A model of choice from lists. *Theoretical Economics*, *1*, 3–17.
Rubinstein, A., & Salant, Y. (2012). Eliciting welfare preferences from behavioral datasets. *Review of Economic Studies*, *79*, 375–387.
Samuelson, P. (1938). A note on the pure theory of consumer's behaviour. *Economica [NS]*, *5*, 61–71.
Savage, L. J. (1954/1972). *The foundations of statistics* (2nd ed.). New York, NY: Dover.
Selten, R. (1991). Properties of a measure of predictive success. *Mathematical Social Sciences*, *21*, 153–167.
Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118.
Spiegler, R. (2008). On two points of view regarding revealed preference and behavioral economics. In A. Caplin & A. Schotter (Eds.), *The foundations of positive and normative economics: A handbook* (pp. 95–115). New York, NY: Oxford University Press.
Stich, S., & Nisbett, R. (1980). Justification and the psychology of human reasoning. *Philosophy of Science*, *47*, 188–202.
Thagard, P. (1982). From the descriptive to the normative in psychology and logic. *Philosophy of Science*, *49*, 24–42.
Todd, P. M., Gigerenzer, G., & the ABC Research Group (2012). *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press.
Vranas, P. B. M. (2000). Gigerenzer's normative critique of Kahneman and Tversky. *Cognition*, *76*, 179–193.
von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.