Erasmus Journal for Philosophy and Economics,
Volume 9, Issue 2,
Autumn 2016, pp. 90-122.
http://ejpe.org/pdf/9-2-art-7.pdf

Axiomatic and ecological rationality: choosing costs and benefits

PATRICIA RICH
University of Bristol

Abstract: One important purpose of rationality research is to help individuals improve. There are two main approaches to the task of rendering evaluations of rationality that support guidance: the *axiomatic approach* evaluates the coherence of behavior according to axiomatic criteria, while *ecological rationality* evaluates processes according to their expected performance. The first part of the paper considers arguments against the axiomatic and ecological approaches and concludes that neither approach is unserviceable; in particular, each has the flexibility to accept important insights from the other. The second part of the paper characterizes each approach according to the profile of costs and benefits that it accepts, and shows that combining the two approaches in a particular way yields a new approach with a superior cost-benefit profile. This 'hybrid approach' uses axiomatic rationality criteria to evaluate processes that agents might use.

Keywords: ecological rationality, axiomatic rationality, normativity, heuristics, methodology

JEL Classification: A12, B40, D81

1 INTRODUCTION

This paper argues that the best approach to normative rationality for the specific purpose of fostering improvements is a strategic mixture of the two strongest contenders. The methods of each exploit particular kinds of information and attain some desiderata at the cost of others; combining them in a particular way results in a 'hybrid approach' that offers a better cost-benefit package than either alone.

The battles of the so-called 'rationality wars' are as multifarious as they are passionate, with the diversity of goals, methods, and backgrounds

AUTHOR'S NOTE: The author thanks two anonymous referees, Nathan Berg, David Danks, and Kevin Zollman for extremely helpful written comments on the content herein. The paper also benefited from discussions with audiences at Carnegie Mellon University, the University of Bristol, the London School of Economics, and the Zentrum für Allgemeine Sprachwissenschaft.

of rationality researchers leaving every position open to attack from some alternative perspective. At least among well-developed research programs, we should not expect any of the competing paradigms to be unserviceable tout court, but rather better or worse for particular purposes and given particular tools. Each broad program will also have stronger and weaker subprograms. For these reasons, this paper considers competing general approaches to normative rationality, and evaluates them with respect to the specific purpose of guiding individual agents towards greater rationality.

Specifically, I compare the *axiomatic approach* (hereafter 'AA') with *ecological rationality* (hereafter 'ER') (see, e.g., Gigerenzer and Selten 1999; Gigerenzer et al. 2011). Put simply, an AA practitioner defines rational behavior in terms of conformity to abstract axioms, while ER defines rationality in terms of the "match between mind and environment" (Gigerenzer et al. 2011, xix). More precisely, "[w]hen a decision procedure is well matched to an environment, where 'well matched' is defined as achieving good-enough levels on the performance metrics relevant to that environment, then the pair (*decision procedure, environment*) is classified as ecologically rational" (Berg 2014a, 378). AA is the methodology abstracted from existing axiomatic theories (ATs), such as AGM belief revision[1] (Alchourrón et al. 1985) and axiomatizations of rational choice; the use of axioms is the most salient characteristic of these theories, but they have additional common features that will prove important. In contrast, ER is a top-down research program that applies its abstract conception of rationality to multiple types of problems: Gigerenzer and Hertwig run interdisciplinary research centers at the Max Planck Institute with the basic purpose of developing ER (though not necessarily with a normative focus) and applying it to choice, inference, and other tasks.

Important related research programs–such as Kahneman and Tversky's work on heuristics and biases (Kahneman and Tversky 1973; Tversky and Kahneman 1974; Kahneman 2011)–are not considered as independent candidate approaches because their role is to provide descriptive insights rather than new normative criteria. I also evaluate what I take to be the strongest versions of AA and ER. For one, this means de-emphasizing those sub-approaches that are most easily and successfully attacked and remaining agnostic among the defensible ones wherever possible. For this reason, AA is construed as seeking relatively simple behavioral tests of co-

[1] AGM is the dominant theory of belief revision, named for its developers: Carlos Alchourrón, Peter Gärdenfors, and David Makinson. It axiomatizes the concept of belief revision as the minimal change to a theory or belief set in response to learned, perhaps incompatible information.

herence, with emphasis on observability and minimality. In contrast, uses of axioms that appeal to literal process statements (e.g., "the agent maximizes expected utility") make much stronger claims that invite additional criticism (see, e.g., Simon 1956), and even if behavioral economics can be given a normative interpretation (cf., Grüne-Yanoff 2010; Berg 2003) its complexity and proliferation of parameters compared to orthodox rational choice theory makes its normative usefulness highly questionable. Secondly, I consider the theoretical commitments of each approach and the best practices that could be carried out within them, rather than the details of existing practice. This means that I take most seriously ER's basic definitions of rationality–which focus on the degree of success to be expected from using a process, the need to evaluate actual choice and inference mechanisms, and the importance of speed, efficiency, and accuracy–rather than the accompanying biological and anti-orthodox rhetoric (see Hands [2014] for a critical discussion of ER's normative prospects, including the work done by evolutionary and anti-orthodox arguments). While many examples come from decision theory, the arguments pertain to the rationality of inferences, probabilistic belief and belief revision, as well as to the various types of choice problem. Considering general approaches to normative rationality–rather than individual theories–allows me to draw conclusions with significance beyond localized disputes (e.g., whether it is irrational to violate the conjunction rule in the Linda problem [Tversky and Kahneman 1983]; cf., Hertwig and Gigerenzer [1999] and Sturm [2012]), with relevance for the rationality wars as a whole.

The resultant analysis upholds the view that–contra Berg and Gigerenzer (2006) and Berg (2014a)–both the AA and ER approaches have value and a role to play, even with respect to the specific goal of guiding individuals. The first step in this argument is to examine arguments that one of the approaches is completely unserviceable, and rebut them (the arguments against AA are discussed in section 2, while criticisms of ER are described and addressed in section 3). Simply put, neither approach is unserviceable because both are flexible enough to accommodate all of the features that reasonable people might think matter to rationality. For one, neither requires that rational individuals use a particular, controversial type of mechanism (either simple or complex); AA involves checking for the coherence of outcomes and not how they were produced, while ER's standard is the performance of the process, so a process that performs poorly will not be judged as rational. Second, both permit agents to have diverse values matching the diversity actually observed; axioms are abstract and make no reference to the content of preferences or premises,

and ER's criterion of 'accuracy' is likewise neutral regarding the content of the agent's goals. Finally, both are able to incorporate sensitivity to features of the environment that legitimately influence people; the AA proponent is free to endorse different axiom systems for different contexts while ER is not committed to a particular way of carving up the space of contexts with respect to which processes are evaluated. At the same time, each approach has value because each has a way to produce concrete evaluations of rationality without dogmatically imposing goals and values onto people.

Despite this shared agnosticism, practitioners of AA and ER have historically had quite different goals in addressing rationality, and these differences are reflected in the methods each uses and the distinct package of costs and benefits each has chosen (see section 4). Section 4.2 addresses the requirement of clear rationality tests to be used in evaluating rationality; in checking the coherence of observed behavior patterns, AA has a straightforward and relatively objective way to render rationality judgments, while a major obstacle to the application of ER is that its rationality criteria are unclear in many cases of interest. Coherence tests are rather weak, though, and do not lend themselves to suggestions for improvement as easily as do ER's comparisons of concrete processes; Section 4.3 argues that ER has an advantage in that its rationality judgments function as actionable recommendations. Section 4.4 discusses how each approach delivers the desideratum of generalizability, but in different ways: AA judgments generalize across all cases in which agents exhibit the same behavior, but do not necessarily tell us much about how a particular agent will perform in other cases. ER judgments only pertain to the specific context-process pairs that are evaluated, but tell us precisely what performance we should expect when a process is used in a context. Broadly, then, AA's use of axioms enables precise, rather scientific judgments on the basis of readily observable behavior at the cost of those judgments being circumscribed and relatively weak. In contrast, ER gets to the root cause of an agent's performance and the most natural target of recommendations for improvement–the process they use–at the cost of having to accurately identify such processes and provide defensible criteria for evaluating them.

Each of these choices could be the more reasonable depending on the particular setting in which rationality is being evaluated, and especially on the accessibility of elusive information about agents' processes, goals, and so forth. Nonetheless, the costs of each are real and are borne only because it has seemed necessary to secure the benefits. I submit that we

can (in many cases) do better: in section 5 I propose a 'hybrid approach' that combines AA and ER so as to pay lower costs and reap more benefits than is possible with either alone. This hybrid approach evaluates processes according to their expected conformity to the relevant axioms, thus addressing the root cause of agents' performance while avoiding ER's problem of identifying objective standards. While not a panacea, I argue that the hybrid approach–when applicable–is the best we can do given the kinds and amounts of information currently available to us. In promoting a specific method of harnessing the benefits of both AA and ER, I hope to help advance meta-rationality research beyond the long-standing arguments over the degree to which different views can be reconciled and show how the approaches' differences can actually be exploited to our advantage.

2 OBJECTIONS TO THE AXIOMATIC APPROACH

2.1 Mechanisms

AA is often criticized (e.g., Simon 1956; van Rooij et al. 2012; Berg and Gigerenzer 2010), for an alleged commitment to complex mechanisms such as the maximization of a utility function or the execution of a difficult computation, but the brief descriptions above suggest that it is ER that defines rationality according to the process applied to a task, while AA evaluates the rationality of outcomes. To understand this line of criticism, we must examine some applications of AA and how they have been interpreted.

As stated, when we apply AA to evaluate an agent's rationality, evaluation proceeds by checking the agent's behavior (construed loosely to include choices, inferences, etc.) for conformity to a set of abstract axioms, which serve as rules of rationality. Which axioms should be used? Applying AA involves (at least conditionally) endorsing some set of particular axiomatic *theories* (ATs), but an AA proponent is under no obligation to accept any particular AT as normative simply because it purports to provide axiomatic requirements of rationality. Each AT must be evaluated according to its merits, and the way ATs are typically presented and defended can give rise to mechanism-based critiques.

The first step in providing an AT is enumerating the axioms themselves; these should be abstractly stated, simple, relatively few in number, and have strong intuitive force (von Neumann and Morgenstern 1953, ch. 1). Yet listing the axioms is not enough; how do we know each of those axioms should be included, that nothing needs to be added, and that the system captures (axiomatizes) the intended concept? As von Neumann

and Morgenstern (1953, ch. 1) write, a system "is usually expected to achieve some definite aim–some specific theorem or theorems are to be derivable from the axioms". In the case of Expected Utility Theory (hereafter 'EUT'), which originated with those authors, the aim is to provide axioms for choice that are jointly equivalent to the choices being representable as maximizing a numerical utility function (and indeed there are many ways of doing this; see Gilboa (2009) for discussion of several axiomatizations). Similarly, discussing the axiomatic AGM theory of belief revision, co-creator Makinson (1985, 350) says that:

> When tackling a problem like this–the logical or mathematical understanding of an intuitive concept or process–there are two general strategies that tend to present themselves: postulation on the one hand, and explicit construction on the other. On the former approach, we seek to formulate a number of postulates, preferably of a more or less equational nature, that seem plausible for the process, and then investigate their consequences and interrelations. On the latter approach, one seeks to formulate explicit definitions or constructions of the central concepts, and then investigate how far the concepts thus constructed satisfy various conditions, including in particular those which on the former approach may have been suggested as postulates.

In the case of belief revision, Grove's (1988) sphere-based modeling provides support for the AGM axioms by showing that revising in accordance with those axioms is equivalent to applying a particular minimal change function to one's beliefs, given a representation of one's prior beliefs and revision commitments as a (mathematically defined) system of nested 'spheres'.

These supporting theorems are known as representation theorems, and while not all representation theorems introduce process language,[2] it seems to arise naturally. For EUT especially, this language is often interpreted literally and the theory is criticized for assuming or requiring that rational agents make choices via the process of calculating the expected utility of each of their options, and selecting the option with the greatest expected utility (see, e.g., Simon 1956, 1957; Klein 2001).

Since EUT *is* sometimes interpreted mechanistically, there is confusion about its theoretical commitments even though the version of EUT now standard in economics (Mas-Colell et al. 1995, ch. 6) is 'only as-if'–i.e., the theory says that people choose *as if* they maximized a utility

[2] ATs for which this is not the case are still subject to mechanism-based critiques, and can still evade those critiques.

function, not that they actually or should do so. ATs more generally follow this pattern, and awareness of this is reflected in the fact that those who criticize ATs' purported mechanisms often criticize them for being 'only as-if' as well (Simon 1957, 1956; Gigerenzer et al. 2011). As noted in the introduction, I aim to evaluate the strongest version of AA, and this is the version that renders many potential criticisms inapplicable by only endorsing the weaker 'only as-if' claims.[3]

Nonetheless, even an 'only as-if' theory is vulnerable to certain mechanism-based objections. Suppose an expected utility theorist (or an AGM theorist) suspects that an agent makes an axiomatically-rational choice (or belief revision) because that agent uses a simple heuristic; the theorist explains or justifies that choice (or revision) by saying that the heuristic *approximates* the ideally-rational mechanism of maximizing expected utility (or applying the minimal change function to some set of spheres). This position does not escape the danger of appealing to implausible mechanisms: van Rooij et al. (2012) argue that the computation that an agent behaves 'as-if' they performed cannot be explanatory if there is no way that the agent could have computed it, even approximately. They essentially prove that if a computation is intractable (i.e., NP hard), we cannot simply find tractable functions–either one or several–to approximate it, for example explaining behavior via a complex calculation and a simple heuristic that we use in its stead. Since there cannot be fundamental inconsistencies between different levels on which we explain behavior, then, "the only way to ensure consistency between algorithmic- and computational-level theories in cognitive science [. . .] is that computational-level theories posit functions that are tractably computable" (van Rooij et al. 2012, 482).

The lesson is that the proponent of AA should be quite careful to avoid appeals to mechanisms that go beyond their conceptual role in representation theorems and the mathematical power and convenience that they often provide to a theory. The strongest and least vulnerable version of AA focuses all normative claims on the axioms as criteria of rationality; beyond the defensive value of this move, the operationalization of rationality is one of the chief benefits of AA over ER, as discussed in 4.2. A more constructive critique of AA takes the importance of mechanisms as its starting point; such arguments are discussed next, as well as in 4.3 and 4.4.

[3] See Okasha (2016) for an excellent, clear argument that the behaviorist interpretation is the only tenable *normative* interpretation of EUT.

2.2 Descriptive adequacy

A challenge to any AT must respect the distinction between descriptive and normative facts–not object to normative claims simply because they are at odds with what people actually do–and a general challenge to AA must reflect the theoretical commitments of ATs rather than perhaps problematic but non-essential views of particular practitioners. The AA advocate can insist on 'only as-if' ATs to evade the mechanism-based critiques referred to above, but a different variety survives this move. A compelling argument against 'only as-if' theories appeals to naturalism.

Simon and proponents of ER argue that understanding the processes people actually use to choose, decide, and so forth is critical to understanding human rationality. Why would this be true, and why would it cast doubt on an 'only as-if' normative approach? Kitcher (1992) provides an answer by defending the naturalist position that "prescriptions must be grounded in facts about how systems like us could attain our epistemic goals in a world like ours" (63); what people *should* do depends on their epistemic goals, cognitive capacities and constraints, and the nature of their environment. Descriptive facts about human choice and inference are naturally viewed as important data points in understanding the nature of human rationality, not only because they can help us to determine what is possible for people (and hence ensure that we respect the *ought implies can* principle in the appropriate way) but also because we want to understand why we perform well, or poorly, when we do. Pursuant to the "meliorative project" that Kitcher advocates and I take as a starting point, understanding *why* someone makes the choices that they do is invaluable to guiding them towards greater rationality. Kitcher thus emphasizes the investigation of *strategies* that may be implemented (see 4.3 for more on this point). I will even argue in section 4 that such process information is more valuable than the actual content of agents' choices, in a way; hence naturalism provides strong support for ER, which evaluates processes.

This positive argument for ER does not entail that we should reject AA, though, for several reasons. For one, the value of process information and the normative status of abstract rules are not incompatible, and in fact I will argue in section 5 that ATs are well-poised to provide the standards of rationality against which processes are evaluated. If we do not grant them this role, alternative normative standards must be found. I will argue that–especially in the case of choice (as opposed to inference, e.g.)–axiomatic standards are the only kind of standards that we could justify given our current knowledge and capacities for understanding descriptive choice, mainly because axiomatic standards are coherence standards and corre-

spondence standards for preferential choice are problematic (see section 5 for the elaboration of this point).

The legitimacy of evaluating processes–including simple heuristics–using axiomatic standards depends on both the 'only as-if' interpretation singled out in 2.1 and the basic coherence I claim for the endeavor. One might worry that many proponents of ATs will not in fact be happy to accept that heuristics could be rational and that ER proponents will likewise resist any attempt to grant normative status to abstract axioms, the rhetorical enemy. Again, though, the issue is not whether everyone working within one approach will be immediately willing to incorporate ideas from the other, or even to grant the other's value, but instead to ask whether the approaches in fact have value and what role each is best suited to play in achieving our goals.

2.3 Relevance of context

A more focused naturalist worry, hinted at above, is that ATs are too detached from context and agential capabilities because they provide universal requirements for ideally rational agents (Kitcher 1992). Let's grant that the context in which a decision, inference, or belief revision is made could legitimately impact its outcome, i.e., that it can be rational to be influenced by contextual features, broadly construed. To give some intuitive Sen-style examples (Sen 1997), one might choose an apple over a pear in a context in which both are plentiful, but choose the pear in a context of scarcity in order to leave others their preferred apples; one might eagerly take the last piece of pizza from the dinner table at home but move on to celery at an office gathering (when different social norms are in effect). The context might also determine how carefully I approach an inference problem and whether I revise my beliefs in response to another's incompatible assertion or simply dismiss their claim. What is AA's best response to the charge that abstract axioms formulated without regard to content or context might fail to account for normatively-relevant differences between situations?

The standard response in defense of particular ATs is a good start. The rational choice theorist will note that–since the axioms are abstract–relevant contextual features should be included as part of the content of the choice problem. In the example above, the choice is not just between *pizza* and *celery*, or even *the last slice of pizza* and *celery*; rather, there is a choice between *the last slice of pizza when this violates no social norms* and *celery*, and another choice between *the last slice of pizza when this may appear quite rude* and *celery.* It is not inconsistent to choose pizza over celery in the first but not the second case. The rational choice theorist will

point out that their theory can accommodate anything that an agent cares about in this way (Binmore 2009).[4]

This defense derives from the positive argument in favor of abstract rules: rules should be formulated and evaluated independently of particular applications because our intuitions about particular cases can distract and mislead us. Thus Stalnaker (1998) advocates determining the beliefs rational agents may have at various stages of a game by applying the AGM axioms–which provide abstract, general requirements of rational belief change–rather than reasoning on the basis of the content of the beliefs that need to be revised (pertaining, e.g., to opponents' rationality) and leaving room for our prejudices to influence our verdicts. Similarly, Dutilh Novaes (2015) argues that abstract rules of inference were originally formulated to provide neutral, public standards by which to judge argumentative moves, helping people to distinguish between valid arguments and arguments with intuitive or appealing conclusions.

A further defense of AA exploits divergence between its rhetoric, its reality, and its principled commitments. Rhetorically, ATs apply without regard to context; their generality is emphasized because it is taken to be a source of potency, for the reasons just given. In reality, a set of axioms is never intended to apply to all situations; their domains of application are simply very broad. A version of EUT may apply only to situations of objectively-quantified 'risk' (von Neumann and Morgenstern 1953; Mas-Colell et al. 1995) or to 'small worlds' (Savage 1954) (or situations profitably analyzed as such, which is of course a judgment call). Classical logic is almost universally assumed within mathematics, but taken to have less relevance to argumentation or conversation (Grice 1975). Objections to AGM disappear when the requirements of a *revision* context are enforced (Stalnaker 2008). Endorsing an axiom system as providing normative requirements of rationality does involve a commitment to context-neutrality in one sense–axioms give abstract forms of good reasoning independent of the content–but this neutrality only 'kicks in' once it has been determined that those axioms apply in the first place. A misapplication of a particular AT poses no problem for that AT itself, let alone to the axiomatic project as a whole.

One may still object that existing ATs do not apply as broadly is as claimed, that they have taken abstraction and generality too far, or that

[4] This is also the rational choice theorist's best response to complaints that rational choice assumes agents to be selfish, care only for money, etc. (see Džbánková and Sirůček [2015] for a recent example, directed towards economics). Generally, the responses to this kind of complaint are similar to the responses to the context-based criticism, and the complaint itself is much less relevant to AA as a whole.

the existing set of ATs leaves too many contexts of interest unserved by a normative theory (while some contexts, e.g., of choice, have a confusing surplus of ATs). Note, however, that these are not theoretical problems with AA, and future development of axiom systems could (and should) remedy them.

A small number of ATs have historically been taken to cover most contexts; yet there is no reason why this must remain the case, as it becomes clear that better coverage of the space of potential rationality judgments can be achieved with more ATs. In fact, we can interpret recent developments within logic in this way: where 'logic' once implied 'classical logic', there are now logics to meet every need (an early, well-known alternative is *intuitionistic* logic; see Van Dalen [1994] for an overview). This trend has spread from within the mathematical study of logic to the use of logic in normative theories of everyday human activities, with Achourioti et al. (2014) recently arguing for the description of a plurality of logics where the context of the agent's goals determines the relevant norms. Stenning and van Lambalgen's (2008) work on the interpretation of conditionals is one example; another is the development of argumentation theory, which seeks to replace classical logic–as a normative theory of arguments–with alternative systems tailored to common real-life contexts (see Zenker [2012] for a survey connecting informal argumentation theory with developments in formal logic, and Beall and Restall [2006] for a defense of logical pluralism). For rationality research more generally, a mapping project–in which ATs are mapped to contexts, and additional, more specialized axioms are mapped to sub-contexts–could be quite fruitful.

Lastly, AA can again be critiqued via a positive argument in favor of ER: if humans do well in the world by using heuristics that exploit regularities in the environment, then addressing this directly and understanding the causes of our success will be valuable (Gigerenzer et al. 2011, intro.). Again, I stress that finding value in ER does not entail rejecting value in AA. Using a contextual heuristic–even a very successful one–does not imply that accepted abstract rules should be violated, and indeed those rules can play an important role in providing the standards according to which the heuristic may be judged successful (see section 5). Once this is granted, we see that we need more axioms–not fewer–so that there are clear standards to use in evaluating more heuristics of interest. My proposed hybrid approach thus makes the project of mapping axioms to contexts much more pressing.

2.4 Critiques of specific axioms or applications

Our goal is to evaluate AA as a whole (with respect to the goal of fostering improvement), and towards this end critiques of particular ATs or aspects thereof are not necessarily relevant. Nonetheless, a large enough collection of serious problems with particular ATs could well cast doubt on AA's viability, or at least its short-term utility. One might doubt that we could develop ATs with sufficient normative pull absent compelling examples. For this reason, I highlight the best strategies for responding to some of the stronger objections to particular ATs.

First, there are often objections to the normativity of individual axioms. This is especially true of EUT (see Mas-Colell et al. [1995, ch. 6] for the standard modern version), and its independence axiom in particular (which is implicated in the well-known Allais Paradox [Allais 1953]). There are many replies to such objections, and collectively they provide a strong defense. Most basically, to debate the normative status of a particular axiom (such as independence) is to play the game in a sense, to accept that there may be abstract rules of rationality and seek the right ones. AA requires that defensible rules can be found, not that every proposed axiom is in fact normative, or normative with respect to any given context (recall 2.3). While Allais (1953) objected to the "mathematical" approach to rationality itself, others have accepted AA and instead sought ways to weaken or do without problematic axioms. In the case of independence, for example, Machina (1982; 1983) shows that expected utility analysis can proceed with a much weaker requirement (that preferences be 'smooth'), and indeed that this change enables a unified representation of an array of decision "anomalies".

So-called 'technical' axioms (such as EUT's Archimedean axiom)–which are included for mathematical reasons, to enable the proof of representation theorems–can also be controversial. In this case, there is no real cause to worry that including such axioms will result in misguiding the agents we wish to advise; it is basically impossible to observe a violation of the Archimedean axiom (or the non-technical completeness axiom, for that matter) (Gilboa 2009, ch. 6.3.2). So the general concern that we will not be able to find axiomatizations of rational choice that are both compelling and powerful enough to place substantial restrictions on rational behavior has not been borne out. (See also Gilboa [2009, ch. 6.3.3] for a nice articulation of the argument for the reasonableness of utility maximization.)

Another important critique applies to the normative interpretation of work in behavioral economics (hereafter 'BE') as part of AA. It is question-

able whether BE should be taken to be normative at all, as researchers within the program have typically focused on capturing descriptive facts (cf., Kahneman 2003; Camerer and Loewenstein 2004) and they have been criticized for retaining the traditional axiomatic normative standard while putting forth new descriptive models (Berg and Gigerenzer 2010). If BE is interpreted normatively, though, there are legitimate worries. BE theories such as prospect theory are not provided with strong defenses qua normative theories (as orthodox theories are) and they are much more complicated than the orthodox theories, with many more adjustable parameters. This additional complexity makes it difficult even to predict behavior for a new sample (see Berg and Gigerenzer [2010] and Brandstätter et al. [2006] for critiques), let alone to impose substantial restrictions on rational behavior that would enable mistakes to be identified for individual agents. For this reason, it is important that the AA proponent need not (and should not) endorse any AT that, like BE theories, invites a lot of criticism without offering greater ability to help people improve.[5]

3 OBJECTIONS TO ECOLOGICAL RATIONALITY

3.1 Making excuses for inferior reasoning

ER is heavily based in descriptive work, and although proponents endorse descriptive, normative, and 'engineering' (i.e., loosely, performance-increasing) goals, descriptive questions have received the most attention. The thrust of ER's descriptive research is that humans (and other animals) rely on the unconscious application of "fast, frugal, and accurate" heuristics to make decisions and inferences (see, e.g., Gigerenzer and Selten 1999; Gigerenzer et al. 2011). This emphasis gives rise to the complaint that ER loses all claim to normativity in endorsing psychologically plausible, yet overly-simplistic heuristic mechanisms. In other words, ER makes (poor) excuses for humans' bad reasoning.

According to ER, there is a misconception that they endorse heuristics because they believe people can (and do) use them, even though people would be better off using more traditionally rational methods (Gigerenzer et al. 2011, intro.). By definition, heuristics ignore or forget some informa-

[5] This is not to say that BE research is useless from a guidance perspective; many of their findings are quite informative and the greatly improved understanding of descriptive choice that has come from their empirical work should surely be incorporated into the meliorative project. For instance, Pope and Schweitzer (2011) show that loss aversion has led to significant financial losses for professional golfers, and Tversky and Kahneman (1981) show that people may be prone to errors on the basis of framing (errors which can be detected by EUT). The point, however, is that orthodox ATs provide a better standard by which to evaluate behavior than does prospect theory, for example.

tion and make relatively few and simple calculations. It was thus natural to assume that their outcomes would be inferior to those of more complex algorithms, even if such losses were worthwhile due to time and energy savings; ER refers to this phenomenon as the "accuracy-effort trade-off", "believed to be one of the few general laws of the mind" (Gigerenzer et al. 2011, xviii).

Yet this critique of ER is faulty for two reasons. First, ER's most intriguing discoveries have to do with the potential for heuristics to outperform more complex procedures by uncontroversial standards, mainly by exploiting environmental regularities and avoiding overfitting (see, e.g., Gigerenzer and Goldstein 1996; Berg and Hoffrage 2008). Such findings are very valuable from a guidance perspective because they prove that successful strategies need not be difficult for people to learn or implement. ER does not forsake performance in emphasizing heuristics, although it is more explicitly permissive of trading off some performance for speed and efficiency (hence the list of three criteria–fast, frugal, accurate–for heuristics).

Furthermore, this position no more commits ER to requiring simple mechanisms than AA's position commits it to requiring complex mechanisms. Gigerenzer et al. (2011, xxi) say as much:

> In which environments is a heuristic better than, say, a logistic regression or a Bayesian model, and in which is it not? [...] Once it is understood that heuristics can be more accurate than more complex strategies, they are normative in the same sense that optimization methods [...] can be normative–in one class of environments, but not in all.

In other words, whether a simple or a complex mechanism is more rational in a context is contingent on which will produce better results in that context, a claim that AA proponents–who care about outcomes–should find agreeable. ER studies the processes that we are interested in (for the reasons described in 2.2) without thwarting our normative endeavors by holding us to lower standards than the AA; inferior mechanisms will be recognized as such.

3.2 Substantive standards

How could we answer Gigerenzer's rhetorical question about when a heuristic is better than a Bayesian calculation? ATs provide clear standards of rationality via their axioms, and those standards are coherence standards. It is less clear what standards ER holds people to–in lieu of coherence–and

whether these standards are substantive and appropriate. What makes a process a "good match" for an environment, or a better match than an alternative process? The short answer is captured by the slogan "fast, frugal, and accurate" (see, e.g., Gigerenzer and Selten 1999): it is rational to apply a process in an environment to the extent that it has these features. More accurately, it is the *expected* performance of the process that matters, since the actual outcome will vary (and as I note in 4.4, this feature is a virtue).

It is fair to ask if this checklist really amounts to *substantive* criteria. Precise definitions of the constituents have yet to be provided, but our concern is whether the criteria *can be* spelled out in a way that is meaningful, useful, and compatible with the goals we think rational agents might have. The meanings of speed and frugality are fairly intuitive, and I would not expect serious difficulties with defining them precisely and sensibly (or objections to their value). In contrast, the meaning of 'accuracy' is far from clear and providing an explanation should be a top priority. I explore a few directions here.

For particular heuristics, ER researchers define the "accurate" outcome from their own perspective, and their definitions tend to be uncontroversial because there are clear right answers in most of the situations studied (in the 'German cities task', for example, the goal is to choose the city with the larger population [Gigerenzer et al. 2011, ch. 3]). However, we cannot expect this to be true in general: especially when considering situations of risky decision-making or games, as opposed to simple inference tasks, there may be significant controversy regarding the 'right' answer or the 'good' outcome; indeed, if this weren't the case, rational choice would be a far less interesting subject. But then there would seem to be two alternatives available to ER: either it can set forth and defend a particular assignment of value to the world, enabling simple and concrete assessments of accuracy, or it can remain agnostic, and define accuracy as most in accordance with the agent's own preferences.

The first option is untenable from my perspective: philosophers have argued for millennia without settling on an account of what has value and why, or even agreeing that there is anything objective about value in the first place; so it is quite unlikely that an explication of accuracy that depended on an exogenously-given assignment of value could be satisfactorily defended. It is unclear how or why rationality would consist in obedience to someone else's standards instead of one's own.

Now, ER draws inspiration from biology, using the ideas of adaptation and evolution to explain why humans rely on heuristics and why those

heuristics work well for us (Gigerenzer and Selten 1999). This suggests defining accuracy in terms of reproductive fitness, as is done by evolutionary biologists using the tools of game theory, and ER does at times identify accuracy with 'success' (e.g., in Gigerenzer et al. 2011, ch. 2). But again, this definition would require a convincing argument that rationality requires one to prioritize the success of one's genes, and such an argument is not forthcoming. (See Hands [2014] and Grüne-Yanoff [2010] for problems with biological arguments for ER.)

The second option avoids these problems by rendering it unnecessary, even inappropriate, to put forth a specific conception of value. An outcome would be accurate insofar as it was in line with the agent's own goals or preferences–but this is likely to mean adopting the coherence standard of AA when it comes to preferential choice. (Similarly, ER would surely endorse deductive validity as the standard of inference, in the event that the agent seeks classically valid beliefs.) The argument for this hybridization is explicated in section 5.

As for ER's other main criteria–speed and frugality–these are defensible sources of value precisely because they are inescapably connected to real human preferences; it is obvious that people have a preference for making their decisions in a timely and efficient fashion (and indeed *need* to much of the time). How much these criteria should be weighted relative to accuracy has not been explained, and there might be room for the worry that ER will overvalue them; this would re-invite the accusation of making excuses for humans' bad reasoning. There is also a worry that–although ER proponents criticize EUT's implicit exchange rates between different sources of value–ER must itself specify some kind of exchange rate between speed, frugality, and accuracy in order to evaluate the rationality of agents' (real and potential) trade-offs between them. Again, these problems are best addressed by deferring to agents' own preferences, rather than exogenously imposing value judgments.

3.3 The generality problem

ER can be viewed as a generalization of *reliabilism* in epistemology, the view that the reliability of the processes or methods used is an important criterion (or even *the* criterion) for a person's belief to be knowledge (or to be justified, etc.) (Goldman and Beddor 2016). Considering objections to reliabilism is therefore instructive, and one particular objection–the *generality problem* (Conee and Feldman 1998; BonJour 2002)–is indeed quite pertinent.

The problem is this: suppose I look across my friend's office and form the belief that there is a copy of *Crime and Punishment* on her desk. 'Sim-

ple reliabilism' says that this belief is justified iff it is formed by a process that reliably generates true beliefs, but whether this is the case will depend on the level of generality at which the situation is described; my visual processes may be only moderately reliable over the full range of cases they are used, quite reliable for identifying objects around 10 feet distant, very unreliable for identifying small objects such as books at this distance, but perfectly reliable with respect to the singleton 'identifying a Russian novel exactly 10.5 feet away in bright June morning sunshine'. Reliability varies according to the description of the process and context, and there seems to be no non-arbitrary way to single out the 'correct' level of generality at which to describe them.

This problem would seem to affect ER as well, and especially as it is applied to guide agents towards greater rationality. We can understand the ecological position as maintaining that rationality claims are *conditional* claims, i.e., claims of the form "*if* the environment is like this, *then* such and such process is ecologically rational". An alternative phrasing would be "in the class of environments that share such and such features, this process will generally be more successful than alternatives". Yet without a prior reason for studying a particular context or a process that could be used to derive a context, how should the boundaries of the context and process be drawn? What is the best level of generalization? (As with reliabilism, there are problems with both too little and too much generality. See, for example, Lee's [2007] argument that a heuristic as studied by ER cannot even be explanatory until *both* the heuristic and its range of application have been fully specified; Lee suggests that ER proponents address this problem by addressing the methodological questions it raises for them. Similarly, Kitcher [1992, 66] points out the need to characterize a process' target contexts.)

We might try to avoid the problem by acknowledging the validity of the rationality claims at all levels, and leaving judgment to determine the appropriate level of generality for any given case in practice. Still, the existence of multiple, perhaps incompatible evaluations of the same situation is disconcerting. This is an interesting and highly important issue, and it will take substantial work to make headway into an answer. Here, suffice it to say that the problem is bigger than ER, but that the approach should be able to accommodate a solution since it seems to have no principled commitments that would prevent this.

This observation also addresses a related concern, that since ER makes all judgments relative to an environment, there is a danger of losing the bigger picture. ER may focus on the details of a particular situation to

such an extent that no general rules emerge that the agent can use to succeed in a new environment. The ER proponent must acknowledge that rational agents need to abstract *to some extent*because they cannot use (or, especially, learn) the unique ideal process for every situation they encounter (and indeed this "process" would be trivial, simply stipulating the best-response action for the situation). The question is the extent to which agents, and theorists, should abstract. ER proponents should be willing to generalize contexts as far as is sensible, and they have not said anything to indicate that they would do otherwise. Furthermore, the hybrid approach suggests a potentially-useful rule for fixing evaluation contexts: evaluate processes with respect to the context of application of the relevant axioms. If this context seems too broad, the mapping project advocated in 2.3 can be used to refine it.

4 THE COSTS AND BENEFITS OF AA AND ER

4.1 Chief differences

The foregoing shows that both AA and ER escape fatal flaws that would render them unserviceable. Notably, both approaches survive scrutiny because their flexibility allows them to incorporate each other's insights into their defensive strategies. Although both approaches aim at the same end–rationality judgments–the distinct methods their proponents employ do entail the acceptance of different packages of costs and benefits, and so their judgments differ in particular ways. The chief difference between AA and ER (for present purposes) is that ATs judge outcomes, while ER judges processes. An attendant, less fundamental difference is that ATs' judgments tend to be all-or-nothing (though AA researchers have sought ways to render more fine-grained judgments; see, e.g., Schervish et al. [2000]; Echenique et al. [2011]), while ER judgments are naturally comparative ("this process is *more ecologically rational* than that one in this context"). The remainder of this section shows how these methodological differences cause each approach to perform well with respect to some desiderata, and less well with respect to others. Again, the purpose with respect to which the approaches are evaluated is that of producing evaluations that can be used to guide agents towards greater rationality.

4.2 Clear rationality tests

A high-priority desideratum for the meliorative project is clear, relatively straightforward and objective tests for rationality. A prerequisite to telling people what to change in order to be more rational is to identify the problematic aspects of what they are already doing. In order to do this with proper authority and legitimacy, the criteria according to which the per-

son is judged should be as clear and objective as possible; compare giving an agent a short, standardized list of simple requirements that they have failed to satisfy with giving them a long written argument to the effect that you think they have made certain mistakes for various subjective reasons.

With respect to this desideratum, ATs are well served by their axioms and ER pays the cost of judging the rather vague "match between mind and environment" (Gigerenzer et al. 2011, xix), leaving 'accuracy' unspecified, and not defining an exchange rate between speed, frugality and accuracy.[6] For AA, once an AT is endorsed for a certain domain, the axioms provide straightforward requirements of rationality and it is fairly simple to check whether readily-observed behavior conforms to those axioms or instead violates any axioms or implications thereof. For example, if an agent makes a series of choices between lotteries, their choices can be represented on a simplex and there is a very easy geometric test for conformity to the EUT axioms (namely that the indifference curves must be straight, parallel lines, or planes, etc.) (Mas-Colell et al. 1995, ch. 6). Especially when the number of possible outcomes does not exceed three, a simple diagram suffices to show whether the choice pattern may be deemed rational. To use a system of logic to judge argumentative moves in a multi-agent debate, one might use natural deduction to show that the claims each agent makes can be derived from agreed-upon premises using the applicable inference rules; indeed, there are now computer programs that will quickly check the validity of an argument whose premises and conclusions are entered in abstract form.

While interpretational problems are not entirely eliminated–we need to identify the correct objects of choice, the implicit premises, and so on–axioms thus provide the most straightforward and objective criteria that could be asked for, given the subject matter. They make the criteria for rationality explicit, and the theorist need know nothing of an agent's inner psychology to determine whether those criteria have been met. This is not an accidental feature of ATs; rather, concepts of interest (e.g., belief revision, utility maximization) are axiomatized in order to specify and operationalize their meaning. This motivation is manifest in von Neumann and Morgenstern's presentation of EUT (1953, ch. 3.5.2); the authors expound on the need to make economic problems amenable to scientific treatment by appealing only to observables (for them, choices) and note the desirability of keeping the set of axioms small and uncomplicated.

[6] To be clear, performance criteria are specified for particular problems; it is not impossible to do so. The issue is rather that a general, theoretical explication is lacking and, importantly, there are situations of interest for which the right criteria are far from clear.

The same motivation pervades economics more generally and can even be found in the origins of logic (as noted earlier; see Dutilh Novaes [2015]).

4.3 Actionable results

A second desideratum for an approach to rationality with meliorative goals is that it produce directly actionable judgments, i.e., that the agent whose rationality is judged be in a position to make real improvements on the basis of the judgment. AA's benefit of clear behavioral rationality tests comes at an actionability cost, while ER's judgments (if prudently formulated for the purpose) are directly actionable.

ER evaluates processes, and its evaluations can therefore be read as recipes for improvement. Processes are *ways to* complete tasks, and these can be taught, learned, and implemented, especially when they are simple (as heuristics are). ER evaluations are also comparative: process A is more ecologically rational than process B, which is more ecologically rational than process C (with respect to a particular task and context). An agent who is told this, and told that they have been using process C, knows immediately that they can make a rational improvement by switching to process B, and an even greater improvement by switching to A. (Of course accounting for the costs–including opportunity costs–of such switches may be difficult, but the point here is that the recommendations for improvement follow directly from the evaluations because the focus is on processes.) One caveat is in order: ER's judgments may be sensitive to whether each process is employed unconsciously or deliberately, because the deliberate application of a process by an agent may take time and energy that would not be needed if their brain implemented it automatically; it is important not to recommend that an agent switch to a process that would no longer be judged superior once the costs of deliberately implementing it were taken into account.

In contrast, there is a gap between AA judgments and implementable recommendations for improvement. If an agent's behavior is judged rational according to the relevant AT, then no action is needed as far as we can tell. If the behavior is judged irrational, though, the question of how the agent should respond is left open. Take the Allais Paradox, for instance: the paradox is that a commonly-displayed pattern of choices between lotteries violates EUT (Allais 1953). As noted, the AA proponent is under no obligation to endorse EUT and may well prefer a weakened theory that permits the Allais choice pattern, but let us suppose that we endorse EUT and aim to help the agents who made the problematic choice to avoid the error in the future. It is easy to see that the axioms are violated, but this does not tell us why the agent made an error, how serious

the error is (though see Zynda [1996]; Schervish et al. [2000]; Staffel [2015] for attempts to address this), or how frequently we can expect the agent to make similar errors (see 4.4 for more on this point). The agent has no recipe for improvement; perhaps they can (and in the Allais case may well) avoid making the error if they face exactly the same situation again, but they cannot hope to avoid similar errors in the future unless it is explained to them why their original choices are problematic and what to do differently (i.e., *how* to choose better). As Kitcher (1992, 68) writes,

> The philosophical dichotomies rational / irrational and justified / unjustified may stand in need of replacement rather than analysis. When we note that a student falls short of the external ideal (as we conceive of it), debate about whether the failure to undergo the epistemically optimal process is excusable or not can profitably be sidestepped in favor of a psychologically richer explanation of what occurred. Cognitively inferior performances can be based on laziness, methodological ignorance or misinformation, failure to perceive relevant similarities, lack of imagination, and numerous other kinds of factors.

I do not suggest that the AA proponent would be unable to generate sensible recommendations on the basis of axiom violations, but it is important to recognize that doing so requires going beyond ATs' basic judgments. The appropriate recommendation is also likely to depend on the reasons for the error (as Kitcher suggests), i.e., the process that ER but not AA is explicitly interested in. Furthermore, to the extent that these recommendations remain informal, subjective supplements to the basic axiomatic tests, they are unlikely to inherit the full authority of the tests themselves.

4.4 Generalizable evaluations

A final desideratum is that evaluations of rationality be generalizable; this is in part an efficiency consideration and in part a corollary of the second desideratum (in that it is useful to be able to generalize from an agent's performance in one case to their expected performance in other cases). Both ATs and ER produce evaluations that are generalizable, but in different respects.

Since ATs evaluate observational records–and features internal to the agent are not considered–their evaluations apply equally to all agents who display the same pattern of choices, inferences, and so forth. For example, all agents who display the Allais choice pattern choose irrationally according to EUT, while all agents who rate the proposition 'Linda is a bank

teller and an active feminist' more probable than 'Linda is a bank teller' (Tversky and Kahneman 1983) make an error according to a straightforward application of probability theory. Axiomatic evaluations generalize across agents when we hold the problem fixed. They do not generalize, however, to other problems faced by the same agent: violating axiomatic requirements in one case does not imply that an agent will violate them in other cases, nor does conformity in one case imply conformity in general. A brilliant reasoner may have a bad day and fall prey to a fallacy, while an agent who makes perfect logical inferences on an exam may (for all the AT user knows) have flipped a coin on the tougher problems and simply got lucky.

ER avoids this problem because processes are judged according to their *expected* performance in a context; the theorist considers (perhaps simulates) the track record of results that the process would yield, and evaluates this bigger picture. The difference between AA and ER in this respect is analogous to the difference between making a prediction on the basis of a statistical distribution and doing so on the basis of a single point sampled from that distribution. While a sample is not uninformative, it can mislead; knowing the full distribution is far preferable.

ER judgments will also generalize to all cases where the same process is used in the same context, just as AA judgments generalize to cases with the same outcomes. But while the possible outcomes are often limited, the space of processes that agents could use is surely infinite and it may be difficult or impossible to determine which an agent uses. In practice, therefore, a particular ER judgment is likely to lack the broad applicability of an axiomatic judgment of a behavioral pattern.

5 IMPLICATIONS

The implication of these distinct packages of costs and benefits is that neither AA nor ER should be abandoned in favor of the other, especially once we restrict attention to the meliorative project. The better approach will be a function of the information available in any given situation: ER will be advantageous given insight into the processes agents actually use for a task or given the opportunity to teach agents new strategies, while AA allows us to evaluate rationality in the (exceedingly common) situation in which only outcome data is readily available.

Furthermore, the complementarity of the approaches' costs and benefits suggests a stronger conclusion, namely that we would often be better served by a strategic combination of AA and ER than by either of them alone. ER has a significant strength in using process information when it is available, but the problem of clear rationality tests for those pro-

cesses looms large (recall 3.2). Berg argues that since the true markers of well-being for agents are health, wealth, and the like, our normative projects should assess strategies for achieving those goods (as opposed to behavioral consistency). Criticizing the use of money-pump arguments to defend theories of coherent choice, he writes, "[i]f the compelling normative principle is, for example, wealth, then why not simply study the correlates of high-wealth-producing decision procedures and rank those procedures according to the wealth they produce" (Berg 2014a, 382)? Unfortunately, such an inquiry will not be entirely sufficient. While Berg and colleagues are entirely correct to ask whether and to what extent the traditional rationality (often, coherence) of agents' choices and beliefs coincides with the success those agents achieve according to independent metrics such as health and wealth (see, e.g., Berg and Lien 2003, 2005; Berg 2014a; Berg et al. 2016), the claim that coherence metrics are essentially useless (Berg and Gigerenzer 2006; Berg 2014a) is far too strong. I side instead with Sturm (2012, e.g., 77-78), who suggests that traditional (often, axiomatic) rationality requirements provide the background standards against which performance (e.g., the ecological rationality of heuristics) can be evaluated. In spelling out a specific way in which axioms can provide background standards, I construct a stronger response to Berg's challenge than has previously been offered.

Let's begin by granting the claim (by naturalists, ER proponents, and others) that in order to make useful normative judgments, we should evaluate choice strategies in terms of their expected performance. Perhaps we want to know which investment strategies to endorse.[7] First, we must note that we could not discover the most rational strategies simply by identifying the wealthiest people and studying the strategies they have used, because actual wealth is a product of luck and circumstance as well as one's own strategy. Instead, we would need to ask which strategies lead to the most *expected* wealth.

[7] Berg (2014b) himself gets at this question indirectly in a paper showing that the best explanation of entrepreneurs' business location choices is a simple heuristic model rather than an optimization model; his discussion is highly suggestive of the idea that the heuristic, which ignores or fails to gather much of the available information, can be rationalized by features of the choice environment, which include uncertainty and frequent change. This paper's analysis is certainly very useful for understanding the factors behind which areas see development, the likely consequences of public policies and taxation strategies, and for providing more evidence that heuristics can lead to real-world success. Nonetheless–as explained below–I would argue that rigorously comparing the rationality of the individuals' available choice processes requires sensitivity to the individuals' risk preferences, and hence some appeal to coherence standards. If this is not possible then assessments of heuristics' *success* must fall short of assessments of their *rationality*.

Second, we must acknowledge that expected wealth cannot be the correct rationality criterion for an old, familiar reason: the conception of rationality as maximizing expected monetary value was replaced with the conception of maximizing expected *subjective utility* for the simple reason that the two differ, and the latter (by definition) expresses the preferences that we are interested in. Daniel Bernoulli discovered this now-obvious fact, and the correctness of his reasoning was immediately apparent:

> The price of the item is dependent only on the thing itself and is equal for everyone; the utility, however, is dependent on the particular circumstances of the person making the estimate. Thus there is no doubt that a gain of one thousand ducats is more significant to a pauper than to a rich man though both gain the same amount (Bernoulli 1954, 24).

Bernoulli developed a new theory of expected utility maximization on the basis of this insight, and famously used it to explain the St. Petersburg Paradox, which is set up as follows. Suppose a gamble is available with the payout to be determined by the flipping of a fair coin. Let n be the number of the flip on which the coin first lands heads; the gambler then receives $\$2^{n-1}$. In other words, the gambler gets \$1 if the coin comes up heads on the first flip, with the payout doubling each time 'tails' appears. Suppose people have the opportunity to pay in exchange for this gamble. The puzzle is that the expected value of the gamble is infinite, but most people would not pay \$20 for it, and furthermore this choice is intuitively reasonable; but if the rational choice is the expected value maximizing choice, the rational agent would choose the gamble over \$20 for certain. If subjective utility is what matters, however, and the agent values each additional dollar less than the previous one, then it may be rational to refuse the gamble even for \$10.

The upshot is that the expected monetary value of an option may be very different from its subjective value, and we are liable to be drastically misled if we assess rationality on the basis of the former. For example, a strategy that often leads to great wealth but occasionally results in penury may look quite rational from an 'objective' perspective that most real agents would reject. Diminishing marginal utility for money is likely to be particularly relevant to major choices–such as investment, insurance, or career choices–which we should be especially concerned to analyze correctly.

Furthermore, there is no canonical utility function that, once discovered, would solve this problem; individuals may legitimately differ in their

subjective valuations of money and their implicit risk preferences.[8] The third step in our argument is therefore to ask how we could determine the extent to which a choice strategy results in choices that agree with an agent's subjective evaluations. The answer provided by decision theory is that, since we cannot observe agents' 'true' personal preferences, we can instead observe their choices and determine whether the agent *could* be choosing what is best by their own lights given *some* preferences that satisfy simple, compelling properties such as transitivity. While it would arguably be better to compare choices to verifiably-true preferences, the decision theorist accepts that this information (if preferences are even taken to be real) is inaccessible to us and moves on, constructing a theory around available observations. The result, then, is that in attempting to replace coherence standards with an independent, external standard of rationality, we find we must fall back on those axiomatic standards if agents' subjective preferences are to be respected.

It is critical not to read too much into the subjective preferences that this argument refers to; adopting the full apparatus of a complete and stable preference ordering would seem to beg important questions. (Utility likewise should not be read in this section as a modern technical term, but rather as Bernoulli would have used it.) While it is quite reasonable to criticize the assumption that people have preferences in the strongest sense, my argument only requires that people have the sort of preferences that figure in folk psychology; such preferences are both harder to deny and a minimum requirement for normative judgments of choice for both AA and ER. Hence, it will not be easy for the ER-purist to reject my hybrid approach on the grounds that agents do not have preferences to which it makes sense to apply axiomatic standards.

In fact, the preferences that EUT must posit are more like folk-psychological preferences than is often realized. On the strong characterization of preferences, a person comes equipped with a complete ordering over all possible outcomes; this ordering is stable over time and reflects their true, inner self. All of the substantial aspects of this characterization can be dispensed with, however. First, regarding stability, EUT permits people's preferences to change over time and from situation to situation; as

[8] Bernoulli (1954, 32) himself defined a unique utility function, the natural logarithm of the objective value, but this function is not taken to have special normative status; it is easy to imagine that different people might care differently about different ultimate levels of wealth as a result of different personal tastes. Even with the Bernoulli logarithmic utility function, it would be necessary to know an agent's initial wealth level to determine their expected utility-maximizing choice. For example, the St. Petersburg gamble would be worth $2 to an agent with no wealth whatsoever, and $6 to an agent with $1000 of wealth.

Ross (2014) explains in his defense, it is a mistake to equate the perhaps short-lived 'agents' that EUT refers to with temporally-extended human beings. If my preferences differ from last week's, then I am now a different agent. This observation limits our ability to apply EUT over extended periods of time, but this is exactly as it should be; if my tastes and goals have changed, it makes no sense to demand coherence between past and present behavior. Second, assuming that preferences are complete is also fairly innocuous; we need not endorse the metaphysical claim that an agent's mind *contains* a full preference ordering at all times, but only that the person can form a preference when called upon to choose (Gilboa 2009, 62). It is no problem for my argument if preferences are constructed on the fly. Intriguingly, behavioral economics experiments suggest that agents do this, and that the preferences they construct are both arbitrary in an important sense (influenced by irrelevant factors such as priming numbers) and largely coherent (Ariely et al. 2003). Third, these weak interpretations of the stability and completeness requirements already suggest that the idea of a 'true inner self' is dispensable as well. Infante et al. (2016) show that this idea is both ill-founded and integral to the project of "preference purification", which seeks to align people's actual choices with the idealized preferences of their perfectly-rational true selves. While the authors are right to point out that it is often impossible to determine which particular choice is mistaken in an incoherent pattern–and that indeed there may be no fact of the matter–their attendant critique of behavioral economics' 'nudge' program does not automatically imply a critique of EUT itself. Importantly, we can deem the incoherent choice pattern to be irrational simply because it is incoherent, and not because we think this incoherence indicates a failure of the agent to express their true self in some particular way. Finally, references to risk preferences should not be read as implying that individuals have risk aversion or risk affinity as part of their true natures; as is standard in economics, these labels are merely short-hand for agents who display preferences such that a set payout is preferred or dispreferred to a risky gamble with the same expected value. To have risk preferences in this sense is simply to make choices one way or the other when called upon to do so.

It would be difficult to deny that people have preferences in this weak sense. Freed of their metaphysical baggage, they are simply an experience that people have; in von Neumann and Morgenstern's (1953, 17) words, a preference is just a "clear intuition" of how two outcomes rank. While the burden of proof is on those who posit (rather than question) preferences in the strong sense, the opposite is the case regarding preferences in the

weaker sense, as they are a basic part of our folk psychology. One might then ask whether preferences in the weak sense are too flimsy to support normative judgments at all; if they might change tomorrow, aiming to satisfy them today seems less important. But if these preferences cannot support normative judgments nothing is left to do so, and surely we do not want to abandon the normative project altogether. Furthermore, as a matter of fact, we respect people's rights to their preferences irrespective of their source or permanence; and we find this natural because our preferences are usually *relatively* stable and grounded in other aspects of our folk psychology. So a reflective agent will be troubled if their current preferences are shown to be basically inconsistent.

Observe that, while the conclusion of this argument–that coherence provides our best test of choice rationality–is controversial and rejected especially by ER proponents, the steps of this chain of reasoning are *not* contested as part of such critiques, are not taken to be controversial in general, and in fact seem quite inescapable. Indeed, in imagining how we might implement the suggestion to evaluate processes or strategies rather than coherence in a concrete setting, we are essentially forced to rehearse the decision-theoretic tradition that culminates in a collection of *axiomatic* theories of rational choice. So if an AA opponent wants to reject this conclusion, it is incumbent on them to explain at which step the argument goes wrong, and how something better can be provided for situations in which the standards of success are clearly subjective.

A hybrid approach which evaluates processes as in ER, but uses AA's method of checking for conformity to axioms, solves this problem. Axiomatic criteria apply to outcomes, but by simulating the performance of a process in its intended context, an expected track record of outcomes is produced. By applying the axiomatic test to this track record instead of to a single behavior pattern, we also avoid AA's generalizability problem. Of course, this hybrid approach can only be applied when processes of interest can be identified, and it is true that the empirical task of identifying the process an agent actually applies is not easy. "Processes of interest" include many more than those that can be definitively ascribed to agents, however: a critical component of the meliorative project is teaching agents *new* strategies for choice and inference, and the theorist can construct and test candidates without knowing precisely what processes they might replace.

An example will help to illustrate the hybrid approach and its virtues. Consider again the Allais Paradox, a sequence of two choices between pairs of lotteries in which the historic modal choice violates EUT (Allais

1953). As discussed above, an AA proponent may apply EUT to determine that this choice pattern is irrational (or apply another AT to determine that it is rationally permitted); thus a clear verdict is delivered but questions about the broader significance of and appropriate response to this verdict are left unanswered. Proponents of ER provide a causal explanation for the choice pattern: Brandstätter et al.'s (2006) 'priority heuristic' is a simple decision procedure for lottery choices–constructed on the basis of the large body of descriptive findings pertaining to such choices–that predicts the Allais pattern and a host of other empirical phenomena. The authors stop short, however, of providing an explicit normative assessment of the priority heuristic, despite the heuristic's prominence in the ER literature and the avowed normativity of ER. ER proponents' emphasis on the success humans can achieve by using heuristics suggests that they view the priority heuristic favorably, but in principle its normative status should depend on the degree to which it is well-matched to its context of application. What could this mean in the case of lottery choice?

As already noted, the performance standards to be applied can neither be biological nor objective. Appeals to biology may be rhetorically useful, but we simply do not think that rationality requires us to maximize our expected number of offspring. The only acceptable performance standard for lottery choice must defer to agents' subjective preferences, and as the decision-theoretic tradition shows, the way to determine whether agents could be choosing in accordance with their subjective preferences is to apply an axiomatic test to their choices. The hybrid approach provides a straightforward way to evaluate the rationality of the priority heuristic: simulate its lottery choices over its purported context of application and calculate its expected conformity to the chosen AT. This procedure yields a numerical measure of accuracy, facilitating direct comparison with other processes. (This is the practical manifestation of the formal compatibility between AA and ER that I demonstrate in Rich [2014].) ER alone cannot deliver this.

The particular strategy for combining AA and ER into a hybrid approach is not just supported by the value of both processes and outcomes, but also by the related interplay between coherence and correspondence criteria. These values have long been seen as competing within epistemology, leading to different theories of truth, knowledge, and justification (cf., Goldman 1967; Quine and Ullian 1970). Berg et al. (2016, fn. 3) credit Hastie and Rasinski (1988) with bringing the distinction between coherence and correspondence into the psychological literature on rational choice and belief. Hammond (1990, 1996, 2007) explores the interplay

between these values in great detail, taking an interdisciplinary viewpoint and with an eye towards real-life choice and inference. As he writes in *Beyond rationality* (2007, xvi),

> [Y]ou don't turn to logic to prove that the tree you see over there is larger than the one over here [...] But sometimes there is no "tree" [...] For example, a story told by someone usually offers no empirical criterion for its truth. Then, we can evaluate it by referring to the coherence of the story.

For preferential choice especially, Hammond's "tree" is conspicuously absent–hence the development of coherence standards. Discussing ER specifically, he writes (2007, 98),

> There are some judgments–and usually very important ones–that demand justification before the action is taken. However, the justification for correspondence judgments (accuracy, speed, and frugality) can only be determined after the judgment is made. You won't know whether the judgment was accurate [...] until later. [...] Since no empirical criterion for the correctness of such judgments will be available, the justification will have to be made on the coherence of the argument for it, and on the argument's content.

The question often posed to coherentists is what reason we have for thinking that coherence–for example of an agent's beliefs–is an indicator of truth. We can see that this question can just as easily be directed towards the AA proponent–why think that an agent whose choices merely avoid manifest incoherence is in fact choosing what is best by their own lights?–and indeed this concern is an important part of the motivation for ER, which avoids the concern by getting at the source of the choices, in a sense. Yet the critical point, as Hammond argues, is that coherence is often the only criterion we have available; the ultimate goodness of a choice, inference, or belief revision is simply not accessible to us in many situations, and especially in situations of preferential choice. For this reason, even if we endorse naturalism, the meliorative project, and the core tenets of ER, AA will often be indispensable because it provides clear, operationalized coherence standards to which there exists no viable alternative.

REFERENCES

Achourioti, Theodora, Andrew J. B. Fugard, and Keith Stenning. 2014. The empirical study of norms is just what we are missing. *Frontiers in Psychology*, 5 (1159).

Alchourrón, Carlos E., Peter Gärdenfors, and David Makinson. 1985. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50 (2): 510–530.

Allais, Maurice. 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica*, 21 (4): 503–546.

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. 'Coherent arbitrariness': stable demand curves without stable preferences. *The Quarterly Journal of Economics*, 118 (1): 73–106.

Beall, Jeffrey C., and Greg Restall. 2006. *Logical pluralism*. Oxford: Clarendon Press.

Berg, Nathan. 2003. Normative behavioral economics. *Journal of Socio-Economics*, 32 (4): 411–427.

Berg, Nathan. 2014a. The consistency and ecological rationality approaches to normative bounded rationality. *Journal of Economic Methodology*, 21 (4): 375–395.

Berg, Nathan. 2014b. Success from satisficing and imitation: entrepreneurs' location choice and implications of heuristics for local economic development. *Journal of Business Research*, 67 (8): 1700–1709.

Berg, Nathan, Guido Biele, and Gerd Gigerenzer. 2016. Consistent Bayesians are no more accurate than non-Bayesians: economists surveyed about PSA. *Review of Behavioral Economics*, 3 (2): 189–219.

Berg, Nathan, and Gerd Gigerenzer. 2006. Peacemaking among inconsistent rationalities? Comment on Alex Kacelnik et al. In *Is there value in inconsistency?*, eds. C. Engel and L. Daston. Baden-Baden: Nomos, 421–433.

Berg, Nathan, and Gerd Gigerenzer. 2010. As-if behavioral economics: neoclassical economics in disguise? *History of Economic Ideas*, 18 (1): 133–165.

Berg, Nathan, and Ulrich Hoffrage. 2008. Rational ignoring with unbounded cognitive capacity. *Journal of Economic Psychology*, 29 (6): 792–809.

Berg, Nathan, and Donald Lien. 2003. Tracking error decision rules and accumulated wealth. *Applied Mathematical Finance*, 10 (2): 91–119.

Berg, Nathan, and Donald Lien. 2005. Does society benefit from investor overconfidence in the ability of financial market experts? *Journal of Economic Behavior & Organization*, 58 (1): 95–116.

Bernoulli, Daniel. 1954. Exposition of a new theory on the measurement of risk. *Econometrica*, 22 (1): 23–36. Translation of 1738 paper.

Binmore, Ken. 2009. *Rational decisions*. Princeton: Princeton University Press.

BonJour, Laurence. 2002. Internalism and externalism. In *The Oxford handbook of epistemology*, ed. Paul K. Moser. Oxford: Oxford University Press, 234–263.

Brandstätter, Eduard, Gerd Gigerenzer, and Ralph Hertwig. 2006. The priority heuristic: making choices without trade-offs. *Psychological Review*, 113 (2): 409–432.

Camerer, Colin F., and George Loewenstein. 2004. Behavioral economics: past, present, future. In *Advances in behavioral economics*, eds. Colin F. Camerer, George Loewenstein, and Matthew Rabin. Princeton: Princeton University Press, chapter 1, 3–51.

Conee, Earl, and Richard Feldman. 1998. The generality problem for reliabilism. *Philosophical Studies*, 89 (1): 1–29.

Dutilh Novaes, Catarina. 2015. A dialogical, multi-agent account of the normativity of logic. *Dialectica*, 69 (4): 587–609.

Džbánková, Zuzana, and Pavel Sirůček. 2015. Rationality and irrationality in economics (selected problems). In *The 9th International Days of Statistics and Economics*. Prague, September 10-12, 2015.

Echenique, Federico, Sangmok Lee, and Matthew Shum. 2011. The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119 (6): 1201–1223.

Gigerenzer, Gerd, and Daniel G. Goldstein. 1996. Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, 103 (4): 650–669.

Gigerenzer, Gerd, Ralph Hertwig, and Thorsten Pachur, eds. 2011. *Heuristics: the foundations of adaptive behavior*. New York: Oxford Universtity Press.

Gigerenzer, Gerd, and Reinhard Selten, eds. 1999. *Bounded rationality: the adaptive toolbox*. Cambridge (MA): MIT Press.

Gilboa, Itzhak. 2009. *Theory of decision under uncertainty*. Cambridge: Cambridge University Press.

Goldman, Alvin, and Bob Beddor. 2016. Reliabilist epistemology. In *The Stanford encyclopedia of philosophy*, ed. Edward N. Zalta. Metaphysics Research Lab, Stanford University. Winter 2016 edition.

Goldman, Alvin I. 1967. A causal theory of knowing. *The Journal of Philosophy*, 64 (12): 357–372.

Grice, Herbert Paul. 1975. Logic and conversation. In *Syntax and semantics: speech acts*, eds. P. Cole and J. Morgan. London: Academic Press, volume 3.

Grove, Adam. 1988. Two modellings for theory change. *Journal of Philosophical Logic*, 17 (2): 157–170.

Grüne-Yanoff, Till. 2010. Rational choice and bounded rationality. In *Religion, economy and cooperation*, ed. Ilkka Ryysiainen. Berlin: De Gruyter, 61–81.

Hammond, Kenneth R. 1990. Functionalism and illusionism: can integration be fully achieved? In *Insights in decision making: a Tribute to Hillel J. Einhorm*, ed. R.M. Hogarth. Chicago: University of Chicago Press, 227–261.

Hammond, Kenneth R. 1996. *Human judgment and social policy: irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford: Oxford University Press.

Hammond, Kenneth R. 2007. *Beyond rationality: the search for wisdom in a troubled time*. Oxford: Oxford University Press.

Hands, D. Wade. 2014. Normative ecological rationality: normative rationality in the fast-and-frugal-heuristics research program. *Journal of Economic Methodology*, 21 (4): 396–410.

Hastie, Reid, and Kenneth A. Rasinski. 1988. The concept of accuracy in social judgment. In *The social psychology of knowledge*, eds. Daniel Bar-Tal and Arie W. Kruglanksi. New York: Cambridge University Press, 193–208.

Hertwig, Ralph, and Gerd Gigerenzer. 1999. The 'conjunction fallacy' revisited: how intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, 12 (4): 275–305.

Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden. 2016. Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioral welfare economics. *Journal of Economic Methodology*, 23 (1): 1–25.

Kahneman, Daniel. 2003. Maps of bounded rationality: psychology for behavioral economics. *American Economic Review*, 93 (5): 1449–1475.

Kahneman, Daniel. 2011. *Thinking, fast and slow*. New York: Macmillan.

Kahneman, Daniel, and Amos Tversky. 1973. On the psychology of prediction. *Psychological Review*, 80 (4): 237–251.

Kitcher, Philip. 1992. The naturalists return. *Philosophical Review*, 101 (1): 53–114.

Klein, Gary. 2001. The fiction of optimization. In *Bounded rationality: the adaptive toolbox*, eds. Gerd Gigerenzer and Reinhard Selten. London: MIT Press, 103–121.

Lee, Carole J. 2007. The representation of judgment heuristics and the generality problem. In *Proceedings of the Cognitive Science Society*, eds. Danielle S. McNamara and J. Gregory Trafton. Austin: Cognitive Science Society, 1211–1216.

Machina, Mark J. 1982. "Expected utility" analysis without the independence axiom. *Econometrica*, 50 (2): 277–323.

Machina, Mark J. 1983. Generalized expected utility analysis and the nature of observed violations of the independence axiom. In *Foundations of utility and risk theory with applications*, eds. B.P. Stigum and F. Wenstop. Dordrecht: D. Reidel Publishing Company, 263–293.

Makinson, David. 1985. How to give it up: A survey of some formal aspects of the logic of theory change. *Synthese*, 62 (3): 347–363.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic theory*. New York: Oxford University Press.

Okasha, Samir. 2016. On the interpretation of decision theory. *Economics and Philosophy*, 32 (3): 409–433.

Pope, Devin G., and Maurice E. Schweitzer. 2011. Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *American Economic Review*, 101 (1): 129–157.

Quine, Willard V. O., and Joseph S. Ullian. 1970. *The web of belief*. New York: Random House.

Rich, Patricia. 2014. Comparing the axiomatic and ecological approaches to rationality: Fundamental agreement theorems in SCOP. *Synthese*, doi:10.1007/s11229-014-0584-1.

Ross, Don. 2014. *Philosophy of economics*. New York: Palgrave Macmillan.

Savage, Leonard J. 1954. *The foundations of statistics*. New York: Dover Publications.

Schervish, Mark J., Teddy Seidenfeld, and Joseph B. Kadane. 2000. How sets of coherent probabilities may serve as models for degrees of incoherence. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8 (3): 347–355.

Sen, Amartya. 1997. Maximization and the act of choice. *Econometrica*, 65 (4): 745–779.

Simon, Herbert A. 1956. Rational choice and the structure of the environment. *Psychological Review*, 63 (2): 129–138.

Simon, Herbert A. 1957. A behavioral model of rational choice. In *Models of man: social and rational–mathematical essays on rational human behavior in a social setting*. New York: John Wiley & Sons.

Staffel, Julia. 2015. Measuring the overall incoherence of credence functions. *Synthese*, 192 (5): 1467–1493.

Stalnaker, Robert. 1998. Belief revision in games: Forward and backward induction. *Mathematical Social Sciences*, 36 (1): 31–56.

Stalnaker, Robert. 2008. Iterated belief revision. *Erkenntnis*, 70 (2): 189–209.

Stenning, Keith, and Michiel van Lambalgen. 2008. *Human reasoning and cognitive science*. Cambridge (MA): MIT Press.

Sturm, Thomas. 2012. The 'rationality wars' in psychology: where they are and where they could go. *Inquiry*, 55 (1): 66–81.

Tversky, Amos, and Daniel Kahneman. 1974. Judgment under uncertainty: heuristics and biases. *Science*, 185 (4157): 1124–1131.

Tversky, Amos, and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *Science*, 211 (4481): 453–458.

Tversky, Amos, and Daniel Kahneman. 1983. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review*, 90 (4): 293–315.

Van Dalen, Dirk. 1994. Intuitionistic logic. In *Logic and Structure*. Berlin: Springer, 157–190.

van Rooij, Iris, Cory D. Wright, and Todd Wareham. 2012. Intractability and the use of heuristics in psychological explanations. *Synthese*, 187 (2): 471–487.

von Neumann, John, and Oskar Morgenstern. 1953. *Theory of games and economic behavior*, [3rd ed.]. Princeton: Princeton University Press.

Zenker, Frank. 2012. Logic, reasoning, argumentation – insights from the wild. In *International Conference on Logic & Cognition*. Poznan.

Zynda, Lyle. 1996. Coherence as an ideal of rationality. *Synthese*, 109 (2): 175–216.

Patricia Rich is a postdoctoral researcher in philosophy, currently at the University of Bristol and soon to be at the University of Hamburg. Her main research areas are epistemology and game and decision theory, with a focus on rationality. In addition to the present topic, she has written on belief revision in perfect information games and reputation in signaling games.
Contact e-mail: <pr15102@bristol.ac.uk>