Erasmus Journal for Philosophy and Economics,
Volume 11, Issue 1,
Spring 2018, pp. 1-30.
https://doi.org/10.23941/ejpe.v11i1.284

Hybrid Vigor: Coherence and Correspondence Criteria for Heuristics

PATRICIA RICH
University of Hamburg

Abstract: The ecological approach to rationality involves evaluating choice processes instead of choices themselves, and there are good reasons for doing this. Proponents of the ecological approach insist that objective performance criteria (such as monetary gains) replace axiomatic criteria, but this claim is highly contentious. This paper investigates these issues through a case study: 12 risky choice processes are simulated, and their performance records are compared. The first criterion is conformity to the Expected Utility axioms; the Priority Heuristic stands out for frequently violating Transitivity. Next, the Expected Value criterion is applied. Minimax performs especially poorly—despite never violating an axiom—highlighting the tension between axiomatic (coherence) and objective (correspondence) criteria. Finally, I show that axiom violations carry high costs in terms of expected value. Accordingly, coherence does not guarantee objectively high performance, but incoherence does guarantee diminished performance.

Keywords: ecological rationality, expected utility, transitivity, independence, priority heuristic, simulations.

JEL Classification: A12, B40, D81

1. INTRODUCTION

When it comes to the task of judging whether an agent's choices are rational, two approaches vie for dominance. The first and traditional method is to apply the Expected Utility (EU) axioms to see whether a choice pattern is *coherent* (Mas-Colell et al. 1995, Chapter 6). This is a

AUTHOR'S NOTE: The author gratefully acknowledges her Ph.D. committee at Carnegie Mellon, audiences at the 2016 SAET conference and 2017 EIPE Anniversary conference, colleagues in the Entscheidung, Wissen, Wissenschaft seminar at the University of Hamburg, three anonymous referees and the editor for valuable feedback at various stages of this project. She also acknowledges funding from the ERC project 'Epistemic Utility Theory' and DFG project 'Knowledge and Decision' while this work was carried out.

compelling method because (1) it accommodates the subjective aspect of choice goodness (via *utility*), but (2) the axiomatic test is applied to observed choices and is therefore empirically grounded, and (3) the axioms themselves are intuitive, sensible, and bolstered by many proofs and arguments (see, for instance, Gilboa 2009, Chapter 6). The second approach is advocated by many psychologists under the banner of Ecological Rationality (ER). Proponents of ER forcefully criticize EU, proposing that processes should be assessed relative to particular contexts of application (see, for example, Gigerenzer and Selten 1999, Gigerenzer et al. 2011). ER focuses especially on simple decision and inference heuristics, but most important is that the processes can be precisely described by a series of easily-programmable steps. Processes are then judged on objective scales according to how fast, frugal, and accurate they are. This approach is also quite compelling, but for different reasons: firstly because it addresses how and why people make the choices they do, secondly because it is conducive to the project of improving people's choices, and thirdly because objective success is undeniably important.

EU rationality is a *coherence* standard because it checks that choices fit together in a particular way, as captured by the EU axioms. In contrast, the objective standards that ER advocates are *correspondence* standards.[1] The coherence/correspondence divide is now central to the debate about rationality standards. The debate between the approaches is persistent, with no agreement about whether or how they might be reconciled (Wallin 2013, Berg and Gigerenzer 2006, Sturm 2012; see also Rich 2016 for more extensive discussion and detailed literature references).

This paper presents a case study to support a methodological claim about how rationality should be evaluated, in light of this debate. I propose that we should evaluate processes using a hybrid method, simulating them and applying both the relevant axioms (here, the EU axioms) and the relevant objective standards (here, wealth) to the results. Doing so, I claim, retains the advantages of both EU and ER; the method has *hybrid vigor*, just as hybrid organisms are often more robust than either parent due to their increased genetic diversity. In the present case study, I simulate choice heuristics to choose between

[1] The distinction has a long tradition in philosophy but was brought into the rational choice discussion by Hastie and Rasinski (1988); see also Berg et al. (2016, 190) and Hammond (1996, 2007).

lotteries and show that the hybrid method yields more satisfying rationality assessments than either EU or ER on its own. I have defended the basic methodology elsewhere—on both theoretical (Rich 2016), and formal (Rich 2018) grounds—and the present case study serves as a proof of concept, as both EU and ER are informative regarding the heuristics in question.

The paper is structured as follows. Section 2 provides the conceptual background of the case study, justifying each step of my analysis and explaining how the steps combine to support the hybrid method. Section 3 describes the heuristics and the lotteries. Section 4 compares the heuristics' conformity with the EU axioms. Section 5 compares the heuristics using the objective criterion of Expected Value (EV). These sections, therefore, apply a coherence and a correspondence criterion, respectively. Taken together, the most frequent axiom violators tend to leave more money on the table, but the reverse is not true. This highlights the tension between coherence and correspondence criteria. Section 6 goes on to show that the more frequent axiom violators tend to leave money on the table because incoherence and objective losses coincide in a strong sense: axiom violations are associated with significant foregone profit, over 30% in this context. Section 7 discusses both the methodological and the practical implications of the case study.

2. MOTIVATING THE THREE-PART HYBRID METHOD

2.1. Expected Utility for Processes

A superficial difference between EU and ER is that EU evaluates choice patterns, whereas ER evaluates processes. Although processes and EU are seldom combined, there is no principled reason why they should not be. It is often more useful to evaluate processes; for example, teaching people *how* to choose is more efficient than teaching them *what* to choose in each case. I, therefore, adopt a process-based approach—as recommended by ER—and hereafter leave this point implicit.

This section motivates each part of the case study in turn. For present purposes, the best way to motivate the use of EU is to explain a bit of its history. It is designed to circumvent a particular problem—namely that preference is subjective and not directly observable—that plagues more direct approaches.

The development of modern EU theory involved two key steps, first from objective value to subjective utility, and second from free-floating utility to utility grounded in preferences. In the early days of mathematical decision theory, the value of a gamble was taken to be its EV, that is, the sum of the possible outcomes, each weighted by its probability. There is a serious problem with this theory, namely that it assumes that every additional dollar is equally valuable to the agent. This assumption is false, though: people typically have diminishing marginal utility for money (also known as risk aversion). The inadequacy of EV was revealed by the well-known St. Petersburg Paradox, and Daniel Bernoulli (1853 [1738]) dissolved this paradox by explaining how value and utility could come apart, thus taking the first key step in the development of EU theory.

Although the notion of utility is intuitive, Bernoulli's version was not sufficiently scientific because it was simply posited as a quantity. Von Neumann and Morgenstern (1944, Chapter 1) solved this problem by providing a set of axioms such that an agent whose preferences satisfy those axioms is provably representable as maximizing a numerical utility function, while an agent who violates any of the axioms cannot be so represented. Their work allows the utility function to be inferred from choice data. This was the second critical step in the development of EU theory. Fishburn (1989) is an excellent historical reference on this topic with many pointers to further literature.

The point of this history is not to claim that EU theory is the best or the only way to evaluate choices. Rather, the point is that EU theory incorporates an absolutely crucial insight (that utility may legitimately differ from objective value) and solves a difficult problem (that of inferring how an agent actually values the options at hand). The insight cannot be ignored; at most it could be argued that value approximates utility well enough in some restricted context. Similarly, rejecting the axiomatic solution would require addressing the problem of inferring utility in some other way.

With this justification in place, the case study starts by comparing the heuristics using the EU axioms. The idea is that the more often a heuristic produces an axiom violation, the worse its choices are. This is because axiom-violating choices are guaranteed to be suboptimal from the agent's perspective.

The modern formulation of von Neumann and Morgenstern's theory, as found in Mas-Colell et al. (1995, Chapter 6), includes two axioms with implications for which choice patterns are rational: Transitivity and

Independence. Let $X \succ Y$ stand for "Lottery X is chosen over lottery Y." Let A, B, and C be arbitrary lotteries. Then we have the following axiom:

Transitivity If $A \succ B$ and $B \succ C$ then $A \succ C$

Lotteries can also be compounded by applying a probability distribution to a set of lotteries to yield a new lottery; for example, given any probability p, we can define a compound lottery $(p \cdot A\ ;\ (1-p) \cdot C)$ which gives lottery A with probability p and lottery C otherwise. Then we have the axiom:

Independence If $A \succ B$ then $(p \cdot A\ ;\ (1-p) \cdot \mathrm{C}) \succ (p \cdot B\ ;\ (1-p) \cdot \mathrm{C})$

For example, suppose that I choose \$5 over a coin flip between \$0 and \$10. Then I am offered a choice between a coin flip that pays nothing or \$5, and a coin flip that pays nothing or even chances of \$0 or \$10. Independence implies that I choose the first option because the new initial .5 chance of getting nothing is common to both options and should not reverse my initial preference.

2.2 Questioning Coherence: Ecological Rationality

Both Transitivity and Independence are taken to be normative because an agent who violates them chooses *incoherently*, as the choices seem to contradict each other. The real-world relevance of coherence has been questioned, however, and some recent criticisms come from proponents of ER. For example, Berg (2014) argues that,

> [w]ithout the link from conformity with an axiomatized rationality to an external performance metric, these rankings in the hierarchy of rationalities may not be normatively relevant (380).

Then, he suggests,

> If 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? (382).

This criticism is especially valuable because it combines skepticism about coherence with an alternative proposal, namely that we use correspondence standards as relatively direct measures of performance.

Berg cites "health, wealth, and happiness" as relevant measures of performance and suggests that we determine which heuristics are best by figuring out, essentially, which heuristics tend to yield more of specific goods that people value (with wealth being the natural candidate for lottery choices).

The foregoing makes clear two difficulties with Athis suggestion: the wealth standard may differ dramatically from the subjective standard of utility, and we lack a good way of measuring utility without using the EU axioms. The correspondence standard is therefore imperfect, but the coherence standard is imperfect too. One weakness is that, while an axiom violation indicates a suboptimal choice, more information would be needed to say how much worse the chosen option is. A more pressing concern is that EU's coherence test is very weak; a violation proves the agent *cannot* be (represented as) a utility maximizer, but it is never possible to prove that the agent is, *in fact*, maximizing utility. A perfectly coherent agent might achieve terrible health outcomes and make very little money. A single choice is always coherent. Given general facts about human psychology, then, the coherence standard does not capture everything that matters, and a correspondence standard would be a valuable supplement.

The case study, therefore, implements Berg's suggestion to evaluate heuristics based on the wealth they would produce. The most reasonable way of carrying out this proposal is to compare the EVs of the heuristics' choices, and this is the primary standard that I use. The EV metric is discussed further in Section 5.1.

2.3 Connecting Coherence and Correspondence

Given that the first two parts of my case study (motivated in sections 2.1 and 2.2) rank the heuristics according to EU and EV, the natural third step is to determine how these two standards are related, and thereby to answer the question of whether coherence is linked with objective success in a concrete application. ER proponents have suggested a negative answer in general. Berg (2014) points out a lack of evidence that real agents who violate EU will fare badly in an objective sense. In the same spirit, Arkes and colleagues survey the empirical literature, emphasize a lack of evidence that incoherent choices are costly, and decry "the widespread assumption that coherence is a universal, domain general criterion of rationality" (2016, 31). These points mirror the familiar criticism of Dutch Book arguments (see Hájek 2009 for a survey): why should we think that a real-life incoherent agent would have their incoherence exploited, or that they would lose a lot of money before they realized what was happening? The source of skepticism about coherence, then, is a lack of proof that it is strongly correlated with real-world success. The lack of proof may simply be the consequence of proof not being sought, however, and this paper's results suggest that this is the case.

While it is considered an open question whether incoherent choices are indeed bad for the chooser, it is widely recognized that coherent choices need not be good in any objective sense; for example, a Brain in a Vat may be fully coherent but is arguably wrong about everything. So, to gather the right evidence regarding the connection between coherence and correspondence, I directly measure the cost of incoherent choices in the third part of the case study. The striking result is that incoherent choices are not only objectively worse than coherent choices, but dramatically so, yielding about a third less value on average.

2.4 Related Simulation Studies

We are now in a good position to situate the present paper with respect to related work. Especially important are papers by Thorngate (1980), Johnson and Payne (1985), and Bordley (1985), which compare the performance of heuristics by simulating their lottery choices. This work can be seen as a precursor to ER in several ways: it focuses on simple heuristics, prioritizes efficient decision-making over optimal performance, addresses the importance of context in determining how well a heuristic performs and employs correspondence performance criteria. Several of the heuristics studied in this paper appear in those earlier papers.

While these earlier authors recognize that choices would ideally be compared according to their subjective utilities, they also recognize that utilities differ across agents and contexts and must be inferred from behavior. Hence, as Johnson and Payne (1985, 396-397) explain, they use EV as a substitute (just as I do in Section 5.1; see also Bordley 1985, 234). Responsibility is left to the agents to choose heuristics that suit them; the authors aim only to enable informed choices (I return to this point in Section 7.2).

Nonetheless, the drawbacks of the EV approach are clear. Descriptively, we know that agents are usually not best described as having linear utility for money; instead, the so-called "fourfold pattern" of risk preferences enjoys strong support (Markowitz 1952, Tversky and Kahneman 1992). Normatively, risk-aversion and risk-affinity are widely considered rationally permissible but are ruled out by EV. The problem is that a heuristic may perform poorly according to the EV criterion, even though it would allow agents to satisfy their preferences efficiently; it would be hard to tell whether this was the case from EV data alone. With respect to previous simulations, then, this paper is novel in that it uses an axiomatic performance standard to circumvent this problem. I present the simulation set-up in the next section.

3. SIMULATION SET-UP

3.1 The heuristics

I compare twelve heuristics, of which seven come from traditional decision theory (and of these, five are variants of the Hurwicz Criterion), four come from earlier simulation papers, and the last was developed by ER. Each heuristic takes two lotteries as input. Here are the heuristics and their definitions:

Minimax Choose the lottery with the greater minimum payoff. Be indifferent if the minima are equal.

Maximax Choose the lottery with the greater maximum payoff. Be indifferent if the maxima are equal.

Hurwicz$_\alpha$ For each lottery, multiply the minimum gain by α and the maximum gain by $(1-\alpha)$. Sum these products to get the lottery's Hurwicz$_\alpha$ value. Choose the lottery with the greater Hurwicz$_\alpha$ value, or be indifferent if they are equal. In general, $\alpha\in[0,1]$. Here, $\alpha\in\{.1,.25,.5,.75,.9\}$.

Equiprobable Average the outcome values for each lottery. Choose the lottery with the greater average. Be indifferent if these are equal.[2]

Probable Define the 'probable' outcomes for a given lottery as those outcomes with a probability of at least (1/the number of outcomes). Choose the lottery with the greater average of these 'probable' outcomes, and be indifferent when these averages are equal.

Least Likely Choose the lottery with the smaller probability attached to its minimum payoff. Be indifferent when these are equal.

Most Likely Choose the lottery whose most probable outcome is greater; use the average when there are multiple outcomes with

[2] This is equivalent to treating all outcomes as equiprobable, as the name suggests.

maximal probability. When this quantity is equal for two lotteries, be indifferent between them.

Priority Heuristic If the difference between the minimum gains of the lotteries differs by at least 10% of their maximum gain, choose the lottery with the greater minimum. Else, if the probabilities attached to these minima differ by at least .1, choose the lottery with the smaller probability of getting the minimum. Else, if the lotteries' maximum gains differ by at least 10% of the overall maximum gain, choose the lottery with the greater maximum. Else, choose the lottery with the higher probability attached to its maximum gain. Be indifferent if these are equal.

The first three heuristics are old staples in decision theory, proposed for decision under uncertainty (when the outcome probabilities are unknown). As such, they ignore probabilities entirely. Minimax simply chooses the lottery with the best worst-case outcome, while Maximax chooses the lottery with the best best-case outcome. These are limiting cases of the Hurwicz Criterion, which compromises by assigning weight $\alpha \in [0,1]$ to a lottery's worst outcome and $\beta = 1 - \alpha$ to its best outcome, choosing the lottery with the greater weighted sum. Since Hurwicz is characterized by its parameters in this way, I test five different Hurwicz criteria, with $\alpha \in \{.1, .25, .5, .75, .9\}$. Intuitively, then, these heuristics cover the spectrum from extreme caution or pessimism to extreme risk-affinity or optimism. (See Luce and Raiffa 1957, Chapter 13.2 for an overview.)

Equiprobable, Probable, Most Likely, and Least Likely all appear in earlier simulations (Thorngate 1980, Johnson and Payne 1985, Bordley 1985).[3] Note, however, that Equiprobable is essentially the "principle of insufficient reason" that appears in the literature on uncertainty (Luce and Raiffa 1957, Chapter 13.2). Like the other heuristics, these ignore a lot of available information: Equiprobable ignores probabilities entirely, while the others use probabilities (outcomes) in a limited way to determine which outcomes (probabilities) to attend to.

The Priority Heuristic (PH) is of special interest as one of the hallmarks of the ER program. While its precise thresholds are an

[3] This paper studies pairwise choices between lotteries with all non-negative outcomes, since the PH is designed for such choices and they are the most commonly studied. Applicability to this task rules out some heuristics studied elsewhere, such as Tversky's 'Elimination by Aspects' (Tversky 1972) and the 'Better-than-Average heuristic' (Thorngate 1980).

idealization facilitating implementation, it is psychologically quite plausible. Its creators argue that it provides a compelling explanation for many observed patterns in lottery choice, including the paradoxical Allais pattern, the 'fourfold pattern of risk', and the 'Certainty' and 'Possibility' effects (Brandstätter et al. 2006).

To illustrate the heuristic, let us see how it reproduces the typical response pattern in the Allais situation. The first choice is between the lotteries we will call *A* and *B*:

A	$1 million for sure
B	$1 million with probability .89, $5 million with probability .10, and nothing with probability .01

The second choice is between the lotteries referred to as *C* and *D*:

C	$1 million with probability .11 and nothing with probability .89
D	$5 million with probability .10 and nothing with probability .90

The PH is lexicographic, which means that it considers a series of possible reasons for choice, in order, until one of those reasons is decisive. The first of the PH's reasons is the minimum gain. This reason decides in favor of Lottery *A* over Lottery *B* in the Allais case: *A*'s worst outcome of $1 million is compared to *B*'s worst outcome of nothing; this difference exceeds 10% of $5 million, so *A* is chosen. (Put simply, *A* guarantees a good outcome.) In contrast, when the PH compares lotteries *C* and *D*, the lotteries have the same worst outcomes (nothing), and so this reason is not decisive. The probabilities of the minima are compared next, but these are too similar (.89 vs. .90). The PH, therefore, compares the lotteries' maxima, and Lottery *D* is chosen because its maximum ($5 million) is sufficiently large in comparison to *C*'s ($1 million).

Although the PH is intended as a descriptive heuristic, ER proponents repeat that the program's goals are threefold, "descriptive, normative, and engineering" (see, for example, Gigerenzer et al. 2011, xix), so it is fair to subject the PH to normative appraisal. Nonetheless, I am not committed to the PH as the correct explanation. Regardless of its descriptive status, the PH is of interest because it is prominent in the literature and provides a useful comparison between lexicographic and one-step heuristics. Having described the heuristics, I turn next to the lotteries they choose between.

3.2 The Lotteries

A formal lottery precisely represents a risky option; the possible outcomes (here given in dollar values) are listed along with the objective probability with which each outcome occurs (recall the Allais lotteries in Section 3.1). The heuristics are tested on lotteries appearing in the decision science literature, specifically Allais (1953), Brandstätter et al. (2006), Binmore (2009, 50-52), and Kahneman and Tversky (1979). Additionally, around 45% of the lotteries are randomly generated and come from the Technion Prediction Tournament (2008), a competition between algorithms to best predict human lottery choices. These sources provide an initial set of 171 unique lotteries. The heuristics are simulated to make choices on every pair of lotteries from this initial set.[4]

The main requirement for this test set is that it be sufficiently large and diverse to ensure that the results are not an artifact of some feature of the test set that a broader sample of lotteries would not share. The lotteries are diverse in terms of the possible outcomes: Most (including most Tournament lotteries) have two possible outcomes, but some have more, and quite a few have five; some also offer a particular outcome for certain. The outcomes vary from $0 to millions, with the entire range from $0 to thousands well represented (the Tournament lotteries generally have lower values). The fact that all outcomes are non-negative simplifies the analysis at little cost; for example, the loss version of the PH perfectly mirrors the gain version so that replacing all gains ($*x*) with their negation (−$*x*) would not influence the results.[5]

The lotteries also cover a broad spectrum of within-lottery outcome variances: low-variance lotteries are especially prevalent (a natural consequence of including riskless options), but relatively high variance lotteries are well-represented and the range in between is covered. Similarly, a disproportionate number of lotteries have no variance in the probability distributions (as when there is a sure payoff or outcomes are equiprobable), but apart from this, the representation of the range of possible variances is roughly even.

---

[4] Upon request, the author can share the lottery tables, the spreadsheets used to produce the choices, the code used to analyze them, and so forth.

[5] While mixed lotteries—those in which both losses and gains are possible—are potentially interesting, a different set of heuristics would be relevant in that setting.

Additionally, it is critical that nearly half of the lotteries were randomly generated for a prediction tournament because this guards against the concern that there is something peculiar about the lotteries that decision scientists invent and test, and indeed we know that lotteries are often designed to elicit particular responses, such as axiom violations. (I generate more lotteries in Section 6.3 to guard against this concern with respect to Independence.) While 171 lotteries may seem meager, pairing each lottery with every other lottery gives 29,070 choice pairs. This pairing method is another important safeguard against researcher-designed choices, because even when the original lottery pairs were designed to elicit a particular response, the present analysis ignores the intended pairings. It is true that offering every pair of lotteries results in some trivial choices—$1 for sure versus $1 million for sure—but these easy choices won't obscure behavior in more interesting cases, and indeed we will see that the PH in particular makes some surprising choices in cases that we might have considered uninteresting.

It is standard practice to study formalized lottery choice because lotteries capture the essential features of options, even if in many real-world situations those features can only be estimated. (Extrapolation of the results to more common situations is discussed in Section 7.) This paper is atypical in taking such a diverse set of lotteries and pairing each with each. It is more common to consider a restricted problem set in which all choices have a common feature (for instance, a fixed sum is compared to a risky lottery with similar EV); the purpose of this is to determine how well a given heuristic performs for different kinds of problems.

Since the purpose of this paper is different, it makes sense for the problem set to be different as well. The primary goal is to defend a methodological position about how heuristics should be compared, and not to characterize the circumstances in which any given heuristic should be used. To get results that hold broadly, the choice set must be correspondingly diverse. Perhaps the most significant division between lotteries is the magnitudes of the potential gains. I consider sub-contexts with respect to this by breaking the results down by choice EV in Section 5 and checking that payoff magnitude does not drive the results in Section 6.

Proponents of the PH in particular may object that the heuristic is meant to explain so-called 'hard' choices, meaning those in which the available lotteries have similar EVs. The important point here is that this paper is concerned with normative choice, and especially with determining the value of EU conformity once we grant that EV is relevant. From this viewpoint, the similar-EV choices that the PH best explains are unhelpful, precisely because the options are similarly good by design. Especially for the advocate of correspondence standards, the performance differences between heuristics will not reveal themselves on 'hard' choices. These choices are included in the test set, but they cannot comprise it.

4. EVALUATION BY THE EXPECTED UTILITY AXIOMS

4.1 Detecting Violations

As noted, two of the EU axioms—Transitivity and Independence—constrain the (rational) choice patterns of our heuristics. I, therefore, find all the violations of these two axioms for each heuristic, using a program created for the purpose. I discuss my methodological choices in more detail now. Sections 4.2 and 4.3 then present the results—the heuristics' axiom violation rates—for Transitivity and Independence, respectively.

The meaning and justification of Independence is taken to depend on its formulation, and concerns about the standard formulation have been raised which are relevant here. Specifically, there is the question of whether the lotteries in question are multi-stage lotteries or single-shot lotteries. Segal (1992) investigates the distinction thoroughly; in his terms, I use the "Mixture Independence Axiom" (171). Segal finds this version, which pertains to single-shot lotteries, to be less descriptively accurate and not as easily justified from a normative perspective. Nonetheless, the version I use is of interest because it is the standard version, and Section 6.3 provides the axiom with new support. Furthermore, the heuristics are only applicable to single-shot lotteries, so alternative formulations of Independence are not readily evaluated here.

Another methodological point is in order. I test the Independence and Transitivity axioms individually, rather than performing one test for compatibility with some EU hypothesis.[6] It is possible in principle for a choice pattern to violate EU Theory without explicitly violating any particular axiom; the violation may be implicit, as in "Zeckhauser's

[6] For those specifically interested in the PH, the author can provide recipes for holistic EU violations for lottery choices with up to three possible outcomes.

Paradox" (Jeffrey 1988).[7] Some axiom must be violated for EU to be violated, but the violation may be implicit. One might, therefore, worry that by testing Transitivity and Independence separately, I risk missing some violations.

Despite this, the results of this analysis are informative, and in some respects, the axiom-by-axiom method is preferable for present purposes. There are several reasons for this. Firstly, I reduce the risk of missing violations by directly testing for an important class of implicit violations, namely implicit Independence violations, as exemplified by the Allais Paradox.[8] Secondly, the sheer size of the heuristics' choice records means that each heuristic has ample opportunity to show its true colors in my tests. Thirdly, much of the discussion of EU's normative import focuses on particular axioms. For example, some people reject Independence, and Transitivity is often singled out in the debate between coherence and correspondence standards. The axiom-by-axiom analysis, therefore, facilitates engagement with the literature by revealing the performance of the heuristics with respect to particular axioms, and later, the costs of violating individual axioms.

4.2 Transitivity

This section shows the results of counting each heuristic's violations of Transitivity:

Transitivity If $A \succ B$ and $B \succ C$ then $A \succ C$

A set of choices violating Transitivity is a cycle. Cycles are impossible for every heuristic except for the PH, for the simple reason that each orders the lotteries according to a single number, such as the average of the 'probable' outcomes. Since it is lexicographic, the PH can violate Transitivity, and in fact, does so frequently: there are 101,253 total violations in the 876,044 cases of $A \succ B \succ C$ (so $C \succ A$ around 12% of the time).

[7] The Paradox is basically this: Suppose you are compelled to play Russian Roulette with a six-chamber revolver. Consider how much you would pay to remove the only bullet (guaranteeing your life), and how much you would pay to remove one bullet when five chambers are loaded. EU requires that you give the same price for each bullet, which contradicts typical price reports, but the violation must be derived.

[8] The Allais choices (see section 3.1) do not explicitly violate the axiom as written, but the violation can be quickly derived. Note that neither choice pair yields the other with some probability *p*; rather, the first pair has the form $(p \cdot A\ ;\ (1-p) \cdot C) \succ (p \cdot B\ ;\ (1-p) \cdot C)$ while the other has the form $(p \cdot A\ ;\ (1-p) \cdot D) \succ (p \cdot B\ ;\ (1-p) \cdot D)$. That $A \succ B$ is implicit.

This is already a striking finding because cycles are a cause for serious concern, but the source of these violations is also noteworthy; the PH produces many violations because it often makes choices that seem utterly unreasonable. Figure 1 shows a typical example (see the Appendix for more):

	Lottery
A	\$10.60
B	\$11.40 · (.97) ; \$1.90 · (.03)
C	\$310 · (.15) ; \$230 · (.15); \$170 · (.15); 130 · (.20) ; \$0·(.35)

Figure 1 Lotteries between which the Priority Heuristic cycles

The PH chooses *A* over *B* because *A*'s minimum is nearly the maximum outcome for that pair, while *B*'s minimum is relatively small. Between *B* and *C*, the minima are similar and so they are not decisive. Instead, the heuristic checks the probability of those minima; the difference of .32 between the probabilities exceeds the threshold of .10, and so $B \succ C$ due to *B*'s smaller chance of earning the minimum. But the heuristic also chooses *C* over *A*: the minimum gains are not decisive, and *C* has a much more attractive probability of minimum gain (.35 instead of 1). Hence, the PH cycles.

This violation seems unrealistic, and the PH is billed as explanatory for choices between lotteries with similar EVs (Brandstätter et al. 2006, 24)—unlike many of the choices evaluated here. It bears repeating that I evaluate the heuristics in the abstract for a broad range of choice problems, and dissimilar-EV choices are normatively more interesting. I give practical conclusions for the PH in Section 7.2.

4.3 Independence

The test for Independence violations yields more mixed results. Recall the axiom:

Independence If $A \succ B$ then $(p \cdot A \,;\, (1-p) \cdot C) \succ (p \cdot B \,;\, (1-p) \cdot C)$

As noted above, two algorithms are used to detect both explicit and implicit violations. I only count strict Independence violations: it is not counted as a violation if $A \succ B$ and $A' \sim B'$, even if $A \succ B$ implies (via Independence) that $A' \succ B'$. This is a reasonable way to proceed because, first, a strict preference for the "wrong" lottery ($B' \succ A'$) is plausibly serious in a way that indifference between them is not. Moreover, EU states requirements on choices, and when a heuristic is indifferent a person applying it would choose one of the lotteries based on additional considerations; this choice simply goes beyond what the heuristic determines.

Minimax, Maximax, Most Likely, and Probable never produce strict Independence violations; the other heuristics do. The results of the evaluation are summarized in Figure 2, for both axioms. For reference, there are twelve unique opportunities to violate Independence.

Heuristic	Transitivity	Independence
Priority Heuristic	101,253	4
Minimax	0	0
Maximax	0	0
Hurwicz, α=.1	0	1
Hurwicz, α=.25	0	1
Hurwicz, α=.5	0	2
Hurwicz, α=.75	0	3
Hurwicz, α=.9	0	3
Equiprobable	0	4
Most Likely	0	0
Least Likely	0	7
Probable	0	0

Figure 2 Total axiom violations for each process. The maximum possible number of violations is 876,044 for Transitivity, and 12 for Independence.

On the one hand, some heuristics violate Independence at a high rate—for Least Likely, more than half the time. On the other hand, this rate could be misleading because the violations occur in cases designed to elicit them. Another reason for caution is the small sample: Independence places relatively few constraints on choices from the initial test set, and so the heuristics have relatively few violation opportunities. I avoid these problems when I measure the cost of Independence violations (see Section 6.3). Recall that Hurwicz takes the weighted average of the minimum and maximum outcomes, with weight α on the minimum. Figure 3 shows an example of a set of lotteries on which Hurwicz violates Independence for all the tested α values:

	Lottery
A	$2500 · (.33) ; $2400 · (.66) ; $0 · (.01)
B	$2400
C	$2500 · (.33) ; $0 ·(.67)
D	$2400· (.34) ; $0 · (.66)

Figure 3 Lotteries on which the Hurwicz Criterion violates Independence

For every α, *B*≻*A* (because 2400 is greater even than a .9 weighting of 2500) while *C*≻*D* (because 2500 is greater than 2400 no matter what their weighting). This choice pattern is an example of the 'Certainty Effect' (see Brandstätter et al. 2006, 11 for discussion). Independence requires, in contrast, that *A*≻*B if and only if C*≻*D*.

5. EVALUATION BY EXPECTED GAINS

5.1 Expected Value

I now compare the heuristics according to an objective wealth standard by comparing choice EVs. Specifically, given each pair of lotteries, the benchmark is the choice with the greater EV, along with the magnitude of that EV. When a heuristic is indifferent between two lotteries, we can simply average their EVs.

EV is an appropriate benchmark because it tells us the *expected* monetary value of each lottery, or equivalently, its cash equivalent for a risk-neutral individual, or its average payoff if it were played repeatedly. The greater the number of choices to be made, the less relevant variance becomes, and EV can be expected to coincide more closely with actual earnings. We need not simulate the lotteries themselves: by the Law of Large Numbers, the total profits for a heuristic's choices on the test set will be very close to the sum of the EVs of the chosen lotteries. This makes EV especially apt for processes that will be used repeatedly. Moreover, as an objective standard, there is no better option: accounting for an individual's attitude towards the variance would mean looking at their subjective preferences.

We can compare both the aggregate EVs (the total earnings each heuristic is expected to produce) and the average EVs (the percentage of the available EV that each heuristic realizes, averaged over all the choices). These quantities differ because in the aggregate case the relative impact of an individual choice depends on how much money is at stake, whereas in the average case each choice is equally important.[9] Nonetheless, the two measures support similar qualitative judgments, as Figures 4 and 5 demonstrate.

The presence of high-EV lotteries could distort our view because choosing a lottery with an EV of $500,000 when the other has an EV of $1 million leaves so much money on the table that it can overshadow performance on more modest (but perhaps more common) choices. The charts, therefore, break down the comparisons according to the maximum EV for each choice pair.

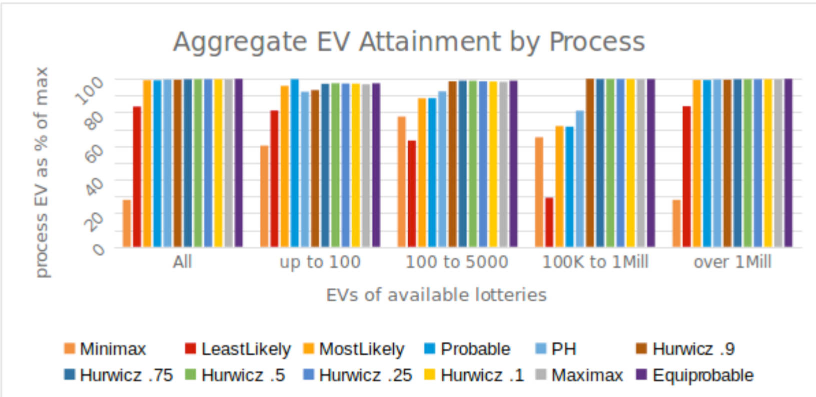

Figure 4 Aggregate EV attainment by process

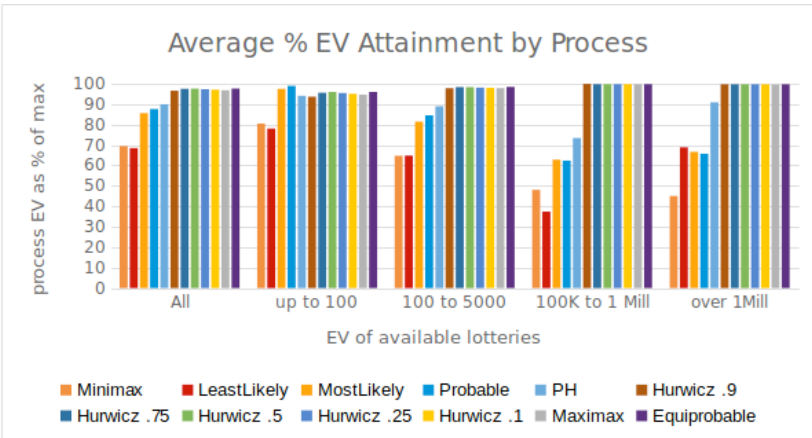

Figure 5 Average percentage EV attainment by process

These charts show that Minimax and Least Likely leave the most money on the table almost across the board; Minimax yields only 28% of

[9] For example, if a heuristic chooses a lottery with an EV of $3 million over one with an EV of $4 million, this registers as a $1 million loss in the aggregate and as a 25% loss as a percentage. In contrast, a choice of a $3 EV lottery over a $4 EV lottery has the same impact on the percentage performance, but a relatively tiny impact on the aggregate performance.

the total available EV, about 70% per choice on average, and leaves more as the lotteries get larger. This is to be expected, as these heuristics only consider the worst case and will avoid lucrative but riskier lotteries.

Most Likely, Probable, and the PH are overall the middle performers; they are noticeably better than Minimax and Least Likely, but noticeably worse than the rest. Most Likely and Probable are outperformed by Least Likely for the highest-valued choices, while Most Likely and Probable are the top performers for lotteries up to $100.

The overall top performers, then, are Equiprobable, Maximax, and the Hurwicz criteria, which all attain nearly maximal EV.[10] It is striking that all of them ignore probabilities entirely. Equiprobable does much better than Probable except in smaller lotteries; the only difference between them is that Probable ignores 'improbable' outcomes, but this leads it to ignore, for example, the 40% outcome in a 60/40 pair. Similarly, all instantiations of the Hurwicz Criterion do far better than Minimax, which shows that putting even a small amount of weight on the best possible outcome (as opposed to the worst) is sufficient to counteract Minimax's caution and completely change the results.

Compared to the axiomatic performance metric, the major difference is that Minimax looks worst according to EV (whereas it was in perfect conformity with the EU axioms). Looking especially at Figure 5, the breakdown of EV attainment by EV bracket provides a partial explanation: as the maximum EV of the lotteries increases, Minimax becomes much less likely to choose the lottery with the higher EV. This is exactly what we would expect to see from a risk averse agent, for whom objective differences in EVs, especially high EVs, have much less subjective relevance than the differences between the minimum gains that can be guaranteed.

Comparing our two performance metrics, then, we see that it is possible to be perfectly coherent and yet not earn much (Minimax), perfectly coherent and a top earner (Maximax), and to violate occasionally and still be a top earner (Hurwicz). The least coherent heuristics—the PH and Least Likely—are only a moderate and a low earner, respectively. This suggests that incoherence is costly, but we need more data to prove this.

[10] This high performance is partly explained by the fact that many choices will involve simple dominance, and the heuristics will choose optimally in these cases. Such choices are 'easy' for all the heuristics, though, and will not affect the heuristics' rankings.

5.2 Two Alternatives

One might not be entirely satisfied with the above evaluation. Firstly, one might think that Minimax is too unrealistic to be of much interest. Secondly, there is an alternative to EV that would better reflect typical preferences. This section briefly describes some supplementary analysis that addresses these points.

Although some people may be Minimax choosers, the vast majority are nowhere near so risk averse: this heuristic would choose a sure $1 over a lottery paying nothing or $1 million with equal probability. We can modify Minimax to create a more realistic conservative heuristic; call this TEV, EV combined with a threshold for the minimum gain. The intuition is this: an agent wants to ensure that their minimum gain meets a certain threshold—I use $1,000—so that they can pay a debt or take a trip, but after this aspiration is taken into account, they maximize EV. In my implementation, TEV chooses the lottery that guarantees at least $1,000 if only one does, and otherwise chooses according to EV. This hybrid process goes much of the way towards closing the gap between Minimax and the other processes because EV is often—but not always—the deciding factor; it yields 88% of the available EV in the aggregate, and 99% on average. So this sensible compromise between caution and EV maximization performs quite well.

As an alternative to EV, we could use a plausible utility function; the logarithmic utility function—proposed by Bernoulli (1954 [1738]), and frequently used in modern economics—is a natural candidate. It defines utility as $U(x)=ln(x)$, and since $ln(0)$ is undefined, I endow our hypothetical chooser with $10,000 in prior wealth to which their lottery earnings are added. The notable result is that Minimax performs much better, attaining approximately 98% of the available utility, because this utility function implies significant risk aversion. However, since this utility function is much more forgiving, it is also much less discriminating, and hence does not enable us to distinguish the heuristics very well by their performance.

6. REALIGNING COHERENCE AND CORRESPONDENCE

6.1 Overview

The axiom-based evaluation method and the correspondence evaluation method only partially agree about the ranking of the heuristics. What can be said about the precise relationship between these performance standards? From the coherence of a process, we can infer nothing about its EV attainment. The previous section raised the possibility that, nonetheless, incoherence might lead to diminished payoffs.

This section confirms that incoherence is costly by looking at the relationship between axiom violations and EV attainment for the most frequent axiom violators: the PH and Least Likely. Each choice that these heuristics make is taken as a data point, with two important attributes: first, is this choice associated with a violation (a binary variable); second, what percentage of the maximal EV does this choice attain in expectation? I evaluate Transitivity and Independence separately.

6.2 Transitivity

The correlation between cyclic choices and those that leave EV on the table tells a clear story. Only the PH is tested since it alone violates Transitivity. When the PH chooses $A \succ B$ and $B \succ C$, a choice of $C \succ A$ is associated with an EV loss of approximately 31%; the result is highly significant ($p<.001$) and the 95% confidence interval around the coefficient is quite tight. Controlling for other factors, such as minimum and maximum choice EV, does not change the result. (See the Appendix for more detail for both axioms.) Descriptive statistics tell the same story: when the PH violates Transitivity, the mean EV attainment of the choice is 64% of the maximum, while for non-violations the mean attainment is 95%. (Despite this significant finding, it is not the case that low-EV choices always coincide with a violation; the minimum attainment in each case is a small fraction of a percent. The point is that violations incur an EV loss, not the reverse.)

	Lottery	**EV**
A	\$10.60	\$10.60
B	\$11.40·(.97); \$1.90·(.03)	\$11.12
C	\$310·(.15); \$230·(.15); \$170·(.15); 130·(.20); \$0·(.35)	\$126

Figure 6 Priority Heuristic cycle with EVs

The statistical result also reflects what we observe when looking at violations. Recalling the example in Figure 6, the PH violates Transitivity relatively often because it is prone to making highly dubious choices; here, the $B \succ C$ choice is particularly costly. Again, the example is typical; the Appendix contains additional examples.

In fact, these statistics underestimate the cost of violating Transitivity because any given cycle involves three choices—hence three data points—but only one of those choices need be the 'mistake' responsible for the EV loss. This accounts for the fact that among both violating and non-violating choices, the median EV attainment is 100%. The 25th percentile of EV attainment among violating choices is only 8.5%, however, which fits perfectly with the hypothesis that the typical cycle contains two reasonable choices and one poor one.

6.3 Independence

The Independence violations from the initial task are too few for meaningful analysis. To remedy this, I construct new lotteries for which Independence has specific implications by compounding the lotteries in the initial set, following a (set of) patterns. Specifically, for each lottery A in the initial set, I create new lotteries $A' := pA\ ;\ (1-p)C$ for probabilities $p \in \{.1, .25\}$ and outcomes $C \in \{0, 25, 500, 5000, 5000000\}$. This means that each lottery A is associated with 10 additional lotteries A' to which it bears an Independence relationship, and every choice $A \succ B$ in the original set implies 10 additional choices $A' \succ B'$. The heuristics therefore have ample opportunity—up to 290,700 opportunities each—to violate Independence, providing enough data points to measure the cost of violations. This new data also provides a better test of Independence since the choice pairs were not designed to generate violations by human subjects.

Apart from this, the Independence analysis mirrors that for Transitivity. Over the 10 A' variations tested, the PH produces between 866 and 11,546 violations per variation (out of 29,070 choices for each). For Least Likely, the minimum is 106 and the maximum is 8,504 (note that Least Likely is more often indifferent between lotteries, so it has fewer opportunities for violations).

For both heuristics, violations are costly. A PH violation is associated with an EV cost of about 32%, and the result is highly significant ($p<.001$) with the 95% confidence interval narrowly around the coefficient. PH choices that violate Independence (that is, $B' \succ A'$ when $A \succ B$) yield only 66% of the available EV on average, while non-violations yield 99%. The median choice attains 93% of the available EV among violating choices, and 100% among non-violating choices. An Independence violation by Least Likely is associated with an EV cost of about 41%, again with $p<.001$ and a narrow confidence interval. Violating choices yield 46% of EV on average (and 41% at the median), while non-violating choices yield 87% (and a median of 100%). As with Transitivity, violations are the driver of cost even when EV is controlled for.

Although Independence violations can be famously compelling (as with the Allais paradox), the reason for their cost is straightforward. Suppose a lottery *A* has greater EV than another *B*, and in fact *A* is chosen. Now, for any *p* and *C* used to make compounds *A*' and *B*', *A*' will have a higher EV than *B*'. An Independence violation, therefore, *guarantees* that one choice fails to maximize EV; the same is true for Transitivity violations. While it is prima facie legitimate to question money pump and Dutch Book arguments, these arguments are essentially elaborations of this observation. The significance of the results presented here lies in the magnitude of the cost, not its mere existence.

7. DISCUSSION

7.1 How to Evaluate Heuristics

This case study demonstrates that the hybrid approach combining EU and ER avoids their individual problems and is more informative regarding the performance of heuristics. Aside from making a small concession to ER by evaluating heuristics, an otherwise pure EU approach would simply rank the heuristics according to how often they generate incoherent (EU-violating) choices; thus, perfectly coherent heuristics would be deemed perfectly rational. In contrast, a pure ER approach would rank the heuristics strictly according to objective performance criteria such as their EV attainment. Additional rankings might be produced to account for additional virtues such as speed. These rankings would similarly be based on objective measurements, such as the average number of computational steps. (Here, the heuristics are all so fast that speed is essentially irrelevant.) The most ecologically rational heuristics would achieve the best EV/speed combination.

Both pure approaches yield evaluations with critical flaws. For example, EU judges both Maximax and Minimax to be perfectly rational. These heuristics imply very different preferences, though, and so each will be wrong for many people. For some, Minimax would guarantee inadequate earnings, while for others Maximax would involve an unacceptable risk of the same fate. ER judges heuristics more favorably the more closely they coincide with EV maximization. Again, this is wrong for agents who are not risk-neutral, for example those for whom

$2 million and $4 million have practically equal utility. For them, a conservative heuristic would be more appropriate.

By combining the EU and ER tests, we achieve a fuller picture of the heuristics' performance and can avoid both kinds of mistakes. Furthermore, by assessing the cost of incoherence according to the correspondence standard, we can determine how relevant the EU standard is even for those who prioritize objective success. The best implementation of the hybrid approach is therefore to perform all three tests (when possible), thereby extracting all of the potentially valuable information from the choice data so that theorists and agents can make informed decisions about which heuristics to use, endorse, and teach.

Yet there is another equally important aspect of the hybrid approach, and an equally important lesson to be learned from this case study. In the case of lottery choices, both coherence and correspondence standards are readily available. For many problems of interest, however—and especially for the kinds of real-world problems of interest to ER proponents—correspondence standards are harder to come by. Coherence can then serve as a proxy—just as Hammond (1996, 2007) argues—bolstered by the demonstrated connection between incoherence and diminished performance. (Of course, the strength of this connection may vary with context, which is why the connection itself should be tested whenever possible.)

As an example, many real-life decision problems involve not risk but rather uncertainty, where the probabilities of the possible outcomes are not known and can only be estimated with more or less confidence. Even if agents have valid subjective probabilities, these are unknown to (and hence unusable by) the theorist in comparing the heuristics, just as with subjective utility. Such decision problems are less amenable to simulation and objective ranking. Nonetheless, the coherence test provides a way to compare possible heuristics, and the connection between coherence and correspondence in the case of risk—especially since the connection is demonstrated for a very broad context—provides evidence that less coherent heuristics would yield objectively worse results in the case of uncertainty as well. Let us now turn to the relevance of the hybrid approach for people making real-world choices.

7.2 How To (Help People) Choose

These simulations do not permit fine distinctions regarding heuristic performance in specific contexts. Further studies would do so, but no study would determine the 'best' heuristic, even relative to a context. Instead, agents must choose heuristics that align with their expectations, preferences, and aspirations in their particular choice context. This fits well with the motivation expressed by Thorngate (1980), Johnson and Payne (1985), and Bordley (1985)—and the intervening years have seen growing interest in helping the public in this way.

Let us first consider how we should respond to the PH's performance. The PH performs poorly here, but it is hypothesized that people use it for similar-EV and therefore less critical choices. This illustrates an important point, which is that the first step of any attempt to improve people's choices should be to determine whether their current choices are especially problematic. It is only worth investing limited time and resources to teach people new heuristics in cases where their existing choice processes are likely to serve them especially poorly. Absent evidence that people are often unhappy with the outcomes of their PH choices, we should not see the use of this heuristic as especially problematic.

Some of the heuristics assessed here—namely those from traditional decision theory—have long been evaluated on their theoretical merits and through examples and intuition, but the simulation method allows us to assess them according to how well we can actually expect them to perform. The results indicate that simpler is often better, Maximax being the most extreme example; this is convenient because simpler is also easier to learn. In contrast, the PH underperforms in an important sense because it is more complicated: its lexicographic nature enables it to make costly intransitive choices.

At this point, one might ask why agents should not simply be taught to maximize EV, at least as a first step. Estimating and attempting to maximize EU is probably not feasible for people without significant formal training, but the arithmetic required to calculate EV is simple, and the exercise would provide a valuable safeguard against very bad choices. EV also basically dominates Equiprobable. While people ought to learn the basics of EV calculation in school along with some fundamentals of probability, there are broader advantages to learning simple heuristics too.

Again, an important consideration is that uncertainty is more common than risk in everyday life, but EV cannot be calculated under uncertainty. Subjective probabilities may be inaccurate, incoherent, or inaccessible. Heuristics become more attractive as choices get more complicated in this way. Since many of the heuristics studied here perform well given a range of probability and outcome distributions, and the top performers make no use of probabilities anyhow, we can extrapolate from their performance here and expect those heuristics to do well in situations of true uncertainty.

Considering situations of uncertainty makes Minimax look even more appealing. In situations of risk, this heuristic epitomizes the tension between coherence and correspondence. In situations of radical uncertainty, however, EV is irrelevant (note that this also makes Equiprobable more reasonable). For an agent whose priority is to make conservative choices, Minimax could be an excellent choice: it involves practically no effort, it minimizes risk, and it will not lead the agent into incoherence (which would entail a cost). Minimax could also easily be used conditionally—as in TEV—by an agent who is risk averse only below a certain aspiration level, or when losses are possible. For those seeking a less conservative heuristic, the Hurwicz Criterion could be very useful. It allows the agent to choose exactly how much weight to put on the worst outcome, and how much on the best; this balance could even be varied contextually. While this heuristic is not perfectly coherent, it can accommodate conservative preferences to a high degree and promises much higher earnings than Minimax. By evaluating these heuristics with the hybrid approach, we are in the best position to help choosers to find their preferred balance.

APPENDIX: VIOLATIONS AND THEIR COSTS

Transitivity

Additional examples of Transitivity violations by the Priority Heuristic:

A	\$15.50	*A*	\$3,000 · .002 ; \$0 · .998	*A*	\$15.50
B	\$18.90 · .9 ; \$6.70 · .1	*B*	\$10.60	*B*	\$18.90 · .9 ; \$6.70 · .1
C	\$1000 · .5 ; \$0 · .5	*C*	\$17.90 · .92 ; \$7.20 · .08	*C*	\$5000000 · .1 ; \$0 · .9

Figure 7 Priority Heuristic cycles

The correlations described in Section 6.2 are based on the following table:

Call:				
glm (formula = PhexpPrcntgofMax ~ PHtransViolYN)				
Deviance Residuals:				
Min	1Q	Median	3Q	Max
-94.868	5.132	5.132	5.132	35.814
Coefficients:				
	Estimate	Std. Error	t value	Pr (> | t |)
(Intercept)	94.86825	0.02756	3441.9	<2e – 16 ***
PHtransViolYN	-30.68178	0.08103	-378.6	<2e – 16 ***

Figure 8 Priority Heuristic Transitivity regression table

Due to the very large sample size, use the standard z^* for the 95% confidence interval as the critical value. Let β^* be the correlation coefficient and *se* the standard error. Then the above yields $\beta \in [\, \beta^* \pm z^* \cdot$ se$] = [-30.68 \pm 1.96 \cdot .08] = [-30.83\ ,\ -30.52]$ as the 95% confidence interval around the regression coefficient of -30.68 (in other words, a violation is associated with a decrease of 30.68% EV).

Independence

The correlations reported in Section 6.3 are based on the regression shown in Figures 9 and 10.

Again use the standard z* for the 95% confidence interval as the critical value. Let β^* be the correlation coefficient and *se* the standard error. Then the above yields $\beta \in [\, \beta^* \pm z^* \cdot \text{se}\,] = [\, -31.94 \pm 1.96 \cdot .09\,] = [\, -32.12\ ,\ -31.76]$ as the 95% confidence interval around the regression coefficient of -31.94 (i.e. a violation is associated with a decrease of 31.9% EV), for the PH. For Least Likely, the calculation is $\beta \in [\, \beta^* \pm z^* \cdot se] = [\, -40.74 \pm 1.96 \cdot .18\,] = [\, -40.09\ ,\ -40.38\,]$.

Call:				
lm (formula = PrcntEV ~ ViolCode)				
Residuals:				
Min	1Q	Median	3Q	Max
-98.503	1.497	1.497	1.497	33.439
Coefficients:				
	Estimate	Std. Error	t value	Pr (> | t |)
(Intercept)	98.50283	0.04148	2374.6	<2e – 16 ***
ViolCode	-31.94203	0.09345	-341.8	<2e – 16 ***

Figure 9 Priority Heuristic Independence regression table

Call:				
lm (formula = LL_PrcntEV ~ ViolCode)				
Residuals:				
Min	1Q	Median	3Q	Max
-86.578	6.622	13.422	13.422	54.162
Coefficients:				
	Estimate	Std. Error	t value	Pr (> | t |)
(Intercept)	86.57767	0.06247	1386	<2e – 16 ***
ViolCode	-40.74004	0.17951	-227	<2e – 16 ***

Figure 10 Least Likely Independence regression table

REFERENCES

Allais, Maurice. 1953. "Le Comportement de L'Homme Rationnel Devant Le Risque: Critique Des Postulates et Axiomes de L'Ecole Americaine." *Econometrica* 21 (4): 503–46.

Arkes, Hal R., Gerd Gigerenzer, and Ralph Hertwig. 2016. "How Bad Is Incoherence?" *Decision* 3 (1): 20–39.

Berg, Nathan. 2014. "The Consistency and Ecological Rationality Approaches to Normative Bounded Rationality." *Journal of Economic Methodology* 21 (4): 375–95.

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?" In *Is There Value in Inconsistency?*, edited by C. Engel. and L. Daston. Baden-Baden: Nomos.

Bernoulli, Daniel. "Exposition of a New Theory on the Measurement of Risk." *Econometrica* 22 (1): 23–36.

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

Bordley, Robert F. 1985. "Systems Simulation Comparing Different Decision Rules." *Behavioral Science* 30 (4): 230–39.

Brandstätter, Eduard, Gerd Gigerenzer, and Ralph Hertwig. 2006. "The Priority Heuristic: Making Choices Without Trade-Offs." *Psychological Review* 113 (2): 409–32.

Fishburn, Peter C. 1989. "Retrospective on the Utility Theory of Von Neumann and Morgenstern." *Journal of Risk and Uncertainty* 2 (2): 127–158.

Gigerenzer, Gerd, Ralph Hertwig, and Thorsten Pachur 2011. *Heuristics: The Foundations of Adaptive Behavior*. New York: Oxford University Press.

Gigerenzer, Gerd, and Reinhard Selten. 1999. *Bounded Rationality: The Adaptive Toolbox*. Cambridge: The MIT Press.

Gilboa, Itzhak. 2009. *Theory of Decision Under Uncertainty*. Cambridge: Cambridge University Press.

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.

Hastie, Reid, and Kenneth A. Rasinski. 1988. "The Concept of Accuracy in Social Judgment." In *The Social Psychology of Knowledge*, edited by Daniel Bar-Tal and Arie W. Kruglanksi, 193-208. New York: Cambridge University Press.

Hájek, Alan. 2009. "Dutch Book Arguments." In *The Handbook of Rational and Social Choice*, edited by Paul Anand, Prasanta Pattanaik, and Clemens Puppe, 173–96. Oxford: Oxford University Press.

Jeffrey, Richard. 1988. "Biting the Bayesian Bullet: Zeckhauser's Problem." *Theory and Decision* 25: 117–22.

Johnson, Eric J., and John W. Payne. 1985. "Effort and Accuracy in Choice." *Management Science* 31 (4): 395–414.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263–92.

Luce, R. Duncan, and Howard Raiffa. 1957. *Games and Decisions: A Critical Survey*. Mineola: Dover Publications.

Markowitz, Harry. 1952. "The Utility of Wealth." *Journal of Political Economy* 60 (2): 151–58.

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

Neumann, John von, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Rich, Patricia. 2016. "Axiomatic and Ecological Rationality: Choosing Costs and Benefits." *Erasmus Journal for Philosophy and Economics* 9 (2): 1–33.

Rich, Patricia. 2018. "Comparing the Axiomatic and Ecological Approaches to Rationality: Fundamental Agreement Theorems in SCOP." *Synthese* 195 (2): 529–47.

Segal, Uzi. 1992. "The Independence Axiom Versus the Reduction Axiom: Must We Have Both?" In *Utility Theories: Measurements and Applications*, edited by Ward Edwards, 165–83. Boston: Kluwer Academic Publishers.

Sturm, Thomas. 2012. "The 'Rationality Wars' in Psychology: Where They Are and Where They Could Go." *Inquiry* 55 (1): 66–81.

The Technion Prediction Tournament, 2008. Organized by Ido Erev, Eyal Ert and Alvin E. Roth. Accessed June 2015. Retrieved from <tx.technion.ac.il/erev/COMP/>.

Thorngate, Warren. 1980. "Efficient Decision Heuristics." *Behavioral Science* 25 (3): 219-25.

Tversky, Amos. 1972. "Elimination by Aspects: A Theory of Choice." *Psychological Review* 79 (4): 281–99.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5: 297–323.

Von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Wallin, Annika. 2013. "A Peace Treaty for the Rationality Wars? External Validity and Its Relation to Normative and Descriptive Theories of Rationality." *Theory & Psychology* 23 (4): 458-78.

Patricia Rich is a researcher on the DFG project *Knowledge and Decision* (P.I. Moritz Schulz) in the Philosophy Department at the University of Hamburg. Her main research areas are epistemology and game and decision theory, with a focus on interdisciplinary methodology. In addition to the present topic, she has written on belief revision in games and reputation in signaling games.
Contact e-mail: <patricia.rich@uni-hamburg.de>