Psychological Review

Choices Without Preferences: Principles of Rational Arbitrariness

Shlomi Sher, Johannes Müller-Trede, and Craig R. M. McKenzie

Online First Publication, February 17, 2025. https://dx.doi.org/10.1037/rev0000529

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

https://doi.org/10.1037/rev0000529

Choices Without Preferences: Principles of Rational Arbitrariness

Shlomi Sher[1], Johannes Müller-Trede[2], and Craig R. M. McKenzie[3]
[1] Department of Psychological Science, Pomona College
[2] IESE Business School, University of Navarra
[3] Rady School of Management and Department of Psychology, University of California San Diego

Traditional models of rational choice assume that preferences are complete, but the completeness axiom is neither normatively compelling nor psychologically plausible. Building on recent work in economics, we develop a rational analysis of decision making with incomplete preferences. The analysis sheds surprising light on a range of well-known behavioral "anomalies," including the endowment effect, status quo maintenance, the sunk cost effect, and coherent arbitrariness. We propose a two-part division of rational choice theory—into preference theory and "implementation theory"—and show how conservative and coherently arbitrary policies can effectively implement incomplete preferences. The two-part normative framework motivates a psychological distinction between evaluation and implementation phases in decision making. We argue that the endowment effect and related phenomena, which have usually been attributed to loss aversion in the evaluation phase, are better explained by conservatism in the implementation phase. The rational analysis challenges the normative adequacy of expected utility theory and raises questions about the explanatory scope of prospect theory. It illustrates the rich interplay between psychological models of value structure and normative models of rational choice.

Keywords: rationality, endowment effect, status quo maintenance, sunk cost effect, coherent arbitrariness

Supplemental materials: https://doi.org/10.1037/rev0000529.supp

Two major themes run through the psychological literature on decision making. First, human decisions are heavily *context-dependent* (Louie & De Martino, 2014). Our choices and evaluations are systematically affected by incidental features of the choice context, such as the frame (how options are described) and the response mode (how preferences are expressed). These effects violate invariance principles that are widely regarded as basic requirements of rationality. *Description invariance* states that logically equivalent descriptions of a choice problem should lead to identical decisions, while *procedure invariance* states that different methods of eliciting preferences (such as choosing between goods vs. assigning prices to them) should yield the same ordering of options. Psychologists have documented a host of framing effects that violate description invariance (Levin et al., 1998), as well as many apparent preference reversals that violate procedure invariance (Hsee et al., 2004). Tversky (1996, p. 195) voiced the standard normative assessment of these effects when he wrote, "Because the assumptions of description invariance and procedure invariance are normatively unassailable but descriptively inadequate, it is not possible to reconcile normative and descriptive accounts of individual choice."

A second major theme is the robust *history-dependence* of our decisions. Our choices are influenced by the investments we have made in the past, and by the state we happen to occupy at present, even when the future consequences of available courses of action are held fixed. We are more likely to select an option if it is designated as the status quo (Samuelson & Zeckhauser, 1988), and we are more likely to pursue a course of action if we have previously invested in it (the sunk cost effect; Arkes & Blumer, 1985). This robust effect of the past on future choices is held to be "irrational from the perspective of both classical and normative decision theories, where individuals are assumed to evaluate decision alternatives on the basis of incremental gains and losses" (Garland & Newport, 1991,

Michael R. Waldmann served as action editor.
Shlomi Sher https://orcid.org/0000-0002-4978-4534
Johannes Müller-Trede https://orcid.org/0000-0002-3669-3078
All data, materials, and analysis code have been made available at the Open Science Framework and can be accessed at https://osf.io/zbmvk/.

This research was supported by a James S. McDonnell Foundation Scholar Award in Understanding Human Cognition to Shlomi Sher, by a National Science Foundation grant (SES-2049935) to Craig R. M. McKenzie, and by Grant PID2019-104246GA-I00 from the Spanish Ministry of Science and Innovation to Johannes Müller-Trede.

Shlomi Sher played a lead role in conceptualization, investigation, methodology, writing–original draft, and writing–review and editing and an equal role in formal analysis and resources. Johannes Müller-Trede played a supporting role in conceptualization, investigation, methodology, and writing–review and editing and an equal role in formal analysis. Craig R. M. McKenzie played a supporting role in conceptualization, investigation, methodology, and writing–review and editing and an equal role in resources.

Correspondence concerning this article should be addressed to Shlomi Sher, Department of Psychological Science, Pomona College, 647 North College Way, Claremont, CA 91711, United States. Email: Shlomi.Sher@pomona.edu

pp. 55–56). And though behavior is often inconsistent across choice episodes, people strive to maintain consistency within episodes—a phenomenon known as "coherent arbitrariness" (Ariely et al., 2003). This arbitrary consistency-seeking is said to maladaptively amplify the impact of initial biases: Our choices are "highly sensitive to normatively irrelevant influences," and then, via coherent arbitrariness, "an initial choice will exert a normatively inappropriate influence over subsequent choices and values" (Ariely et al., 2003, p. 78).

In summary, the standard view of human rationality in the psychological literature is a bleak one. We fail to be consistent where we should be, we manage to be consistent where we shouldn't be, and we too often let the past stand in the way of a better future. In this article, we argue that this standard view rests on an implicit assumption that is ultimately unwarranted—the assumption that rational preferences must be *complete*. The completeness axiom (explained more fully below) states that the ranking of alternatives is always well-defined. Discarding this assumption, we show, has far-reaching implications for our understanding of the rationality and psychology of human decision making. A model of rational choice with incomplete preferences implies, for example, that invariance violations need not indicate irrationality, that coherence can be rationally mandated even when choices are arbitrary, and that it is sometimes appropriate to honor sunk costs. It also lends support to the hypothesis (Gal, 2006) that the endowment effect and related phenomena, which have commonly been attributed to *value asymmetry* (e.g., loss aversion), are instead driven by *value indeterminacy*.

In what follows, we define completeness, review its role in standard models of rational choice, and discuss its limitations. Building on recent work in economics, we next ask what rational choice looks like when preferences are incomplete. We suggest a division of rational choice theory into two components—preference theory and implementation theory. A theory of "implementation" asks how choices should be made when preferences are undefined. It turns out that, just as there are principles of rational preference, there are also principles of rational arbitrariness. In particular, we show that conservative and coherently arbitrary "implementation policies" are adaptive in a range of choice environments. These policies generate forms of rational history-dependence, including effects of endowments and sunk costs.

This article brings history-dependent choice into a framework of "rational analysis" that has been influential in many areas of cognitive science (J. R. Anderson, 1991; Chater & Oaksford, 1999; McKenzie, 2003; Tenenbaum et al., 2006). The value of a rational analysis lies in its capacity to (a) provide a unified explanation of diverse empirical findings, while (b) suggesting new testable hypotheses about cognitive processes. Correspondingly, the article is organized around two objectives. First, we show how a rational analysis of implementation offers a unified account of multiple behavioral "anomalies," including status quo maintenance, the endowment effect, coherent arbitrariness, and the sunk cost effect. In the second half of the article, we then zoom in to take a closer look at loss aversion—a central principle of prospect theory (Kahneman & Tversky, 1979), which states that losses loom larger than gains. The main evidence for loss aversion, we argue, is more plausibly explained by an alternative mechanism suggested by the rational analysis—a conservative implementation rule for imprecise valuations. We review old findings and report new experiments that favor the conservatism hypothesis, and we suggest guidelines for future tests of loss aversion versus conservatism.

The article mainly focuses on simple forms of sequential choice, such as repeated trades and discrete plans. But in the General Discussion section, we take a wider view, considering extensions of the analysis to a broader multilevel "choice hierarchy." In the upper reaches of the hierarchy, the construction of preference grades into the enactment of social roles and, ultimately, "the construction of the self." Principles of rational arbitrariness are likely to be important at these higher levels as well. In conclusion, we reflect on the rich and open-ended interplay between psychological models of value structure and normative models of rational choice.

Must Rational Preferences Be Complete?

Traditional axiomatic models of rational choice assume that preferences are both complete and transitive. The *completeness* axiom says that preferences are everywhere well-defined: For any pair, a, b, of options, either a is strictly preferred to b ($a \succ b$), b is strictly preferred to a ($b \succ a$), or the decision maker (DM) is precisely indifferent between a and b ($a \sim b$). The *transitivity* axiom rules out preference loops, requiring that if $a \succcurlyeq b$ and $b \succcurlyeq c$, then also $a \succcurlyeq c$. Decision theorists have put forward substantive arguments supporting transitivity as a requirement of rationality. For example, a stable intransitivity of preference ($c \succ a \succ b \succ c$) turns the DM into a "money pump," who will pay a premium to exchange c for b, b for a, and a for c once again (Davidson et al., 1955).[1] But, as a number of economists (e.g., Aumann, 1962; Sen, 1997) and philosophers (e.g., Raz, 1986) have argued, completeness is not a plausible requirement of rationality.

A theory of rational choice cannot dictate a DM's values. By the same token, it cannot dictate the level of precision with which those values are defined. What is rational for the DM ultimately depends on her own tastes, commitments, and aspirations. Given a choice between two options—for example, going to a lecture or a movie—which option, if either, is in the DM's best interests depends on psychological facts about what her interests happen to be. Her values may be such as to determine a definite normative comparison (e.g., if she is an aspiring film maker with no patience for lectures), but they need not do so. For example, the DM may appreciate learning and moviegoing more or less independently, without commitment to a superordinate scheme of value that entails precise trade-offs between them (cf. Walasek & Brown, 2023). If so, no normative comparison between tonight's talk and tonight's film need be defined for the DM. In this way, the normative adequacy of the completeness axiom depends on descriptive facts about the core values that define the DM's interests.

Completeness imposes logically strong, but normatively implausible, constraints on preference. It requires infinitely precise valuations for all goods ("this hat is worth \$18.32 to me, not a penny more and not a penny less"). There is no evidence that human DMs possess such valuations, nor that our core values are so constituted as to imply them. Completeness, more generally, rules out any state of indecision that is robust to small perturbations in the value of either alternative. Consider, for example, a DM mulling over two complex vacation packages—a guided tour down the Amazon (a) or a stay at a Hawaiian resort (h). Suppose the DM is undecided between a and h. Now obtain a slightly sweetened option a^+ by adding a 25¢ bonus to the Amazon tour. Despite preferring a^+ to a, the DM will likely remain undecided when

[1] For alternative perspectives on the rationality of intransitive preference, see Anand (2009), Temkin (2012), and Müller-Trede et al. (2015).

faced with a choice between a^+ and h. Such robust indecision does not, of course, establish any irrationality on the DM's part. It simply suggests that her valuation of vacation packages is not defined at the level of 25¢ increments. But this reasonable preference pattern is incompatible with traditional rational choice models that assume completeness.[2]

Completeness is not only normatively uncompelling; it is also descriptively unrealistic. Decision-making experiments rarely include measures of preference imprecision. Instead, typical experimental designs force unique choices or precise valuations. But in studies that give participants the option to express an imprecise value range rather than a sharp indifference point, they generally do so (e.g., 87% of responses in Cubitt et al.'s, 2015, study of lottery valuations). Furthermore, the average size of reported imprecision intervals is substantial, ranging between about 15% and 30% of the distance between best and worst outcomes in the lotteries studied by Butler and Loomes (2011) and Cubitt et al. (2015). Respondents report similarly wide ranges of value uncertainty in contingent valuation studies of health (Ready et al., 2001) and road safety (Dubourg et al., 1997). Of course, in some everyday choices, the ranking of alternatives is clear-cut. (The clean plate is definitely preferred to the dirty one.) But the choice problems targeted by researchers are rarely clear-cut; they usually involve conflicting values or continuous trade-offs. In such situations, completeness is likely to be the exception rather than the rule.

Why, then, is completeness usually assumed in economic models of rational choice? Not because of its normative or descriptive plausibility, but because of its formal convenience. Since the greater-than relation $\geq$ over the real numbers is complete, models that represent preference by the ordering of real-valued utilities (so that $a \succcurlyeq b$ if and only if $u(a) \geq u(b)$) must assume completeness. Indeed, the early developers of axiomatic utility theory were explicit about the motivation for the completeness axiom—its mathematical usefulness rather than its normative or descriptive appeal (von Neumann & Morgenstern, 1944).[3] But in subsequent applications of utility theory, this critical qualification has often—and in the psychological literature, almost always—been overlooked.

We are left, then, with two big questions: First, for a DM with incomplete preferences, what would rational choice look like? Second, how does rational choice with incomplete preferences compare with the actual choices of human DMs, as documented in the psychological literature? In recent years, the first question has drawn increasing interest from both economists (e.g., Cettolin & Riedl, 2019; Eliaz & Ok, 2006; Mandler, 2005; Ok et al., 2012) and philosophers (e.g., Bales et al., 2014; Hare, 2010). But these developments have had minimal impact on the psychological literature to date, and the second question above has been largely neglected. Our goal in this article is to fill this gap. Building on and extending important ideas from Mandler (2005) and others, we sketch a plausible (but tentative) answer to the first question, and then use it to address the second. As we will see, when incompleteness is incorporated, in a straightforward manner, into a model of rational choice, the gap between normative and descriptive analysis narrows in some surprising ways.

Rationality Without Completeness

The simplest way to incorporate imprecise values into a model of rational choice is to represent the DM's preferences $\succcurlyeq$ as a *preorder*. A preorder is a binary relation that is reflexive (for all a, $a \succcurlyeq a$) and transitive where it is defined, but it need not be defined everywhere. That is, there may be pairs, a, b, that are unranked by $\succcurlyeq$: it is neither the case that $a \succcurlyeq b$ nor that $b \succcurlyeq a$. Following Mandler (2005), we write $a \perp b$ in this case. The new relation $\perp$ is called *indecision*. Whereas indifference ($\sim$) indicates perfect equality of value, indecision ($\perp$) indicates imperfect comparability of value. Indecision is formally distinguished from indifference in that, while indifference is transitive ($a \sim b$ and $b \sim c$ imply $a \sim c$), indecision is not ($a \perp b$ and $b \perp c$ do not imply $a \perp c$). Intransitivity of indecision enables us to capture robust absence of preference that survives slight improvement of either alternative. For example, in the Amazon-Hawaii scenario from the last section, $a^+ \perp h$ and $h \perp a$ but $a^+ \succ a$.

Figure 1 depicts two examples of incomplete preference relations. In Figure 1A, the DM assigns an imprecise value range, rather than a unique monetary indifference point, to a hat. In Figure 1B, the DM positively values both attributes in a two-attribute product space, but is not committed to a particular scheme for trading off the attributes. Preference is then defined by dominance ($(a_1, a_2) \succcurlyeq (b_1, b_2)$ if and only if $a_1 \geq b_1$ and $a_2 \geq b_2$). But if each product has an advantage on one attribute (e.g., $a_1 > b_1$ and $b_2 > a_2$), no preference is defined ($(a_1, a_2) \perp (b_1, b_2)$). Other incomplete preference orders can be defined, corresponding to any consistent constellation of value commitments the DM is (un)prepared to make. Such preference structures cannot be quantitatively captured by a real-valued utility function, but that does not render them "unscientific." Indeed, similar incomplete ordering structures appear in many scientific domains. Figure 1C provides one illustration from physics. Letting $\geq_T$ stand for the objective "later-than" relation over events (so that $a \geq_T b$ states that all observers agree that a comes after b in time), relativity theory has the surprising consequence that $\geq_T$, while transitive, is incomplete.[4]

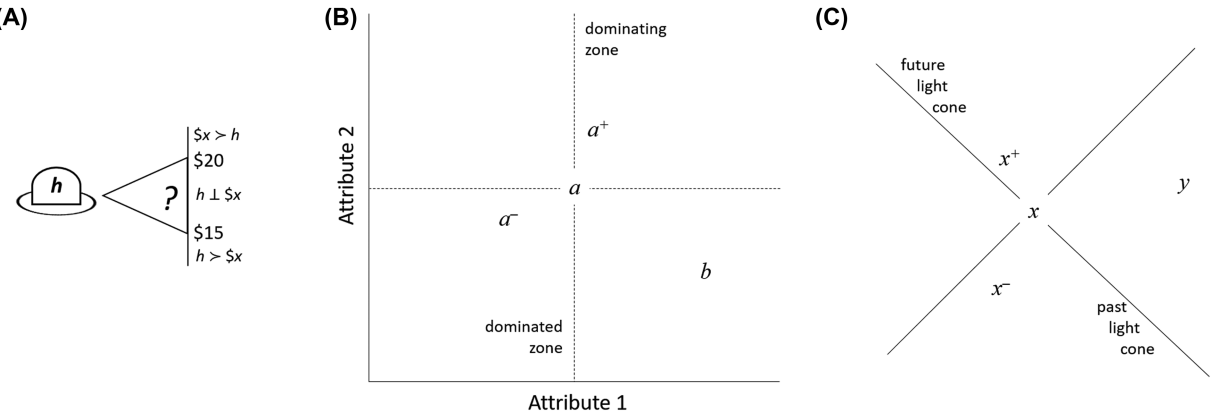
The introduction of an indecision relation $\perp$ is not the only way to incorporate value imprecision into a model of rational choice. Limitations of this modeling approach are addressed in the General Discussion section. But it is the simplest modification of the traditional rational actor model that captures robust indecision, and it is the approach we adopt here. This simple change turns out to have sweeping implications, both destructive and constructive, for our understanding of rational choice. Destructively, it invalidates the invariance principles to which psychologists commonly appeal

[2] Formally, if $a \not\succcurlyeq h$ and $h \not\succcurlyeq a$, then by completeness $a \sim h$. $a^+ \succ a$ and $a \sim h$ then jointly imply, by transitivity, that $a^+ \succ h$. Nonpreference between two options cannot survive infinitesimal enhancement of either one.

[3] von Neumann and Morgenstern (1944) remarked that their "procedure for a numerical measurement of the utilities of the individual depends, of course, upon the hypothesis of completeness in the system of individual preferences. It is conceivable—and may even be in a way more realistic—to allow for cases where the individual is neither able to state which of two alternatives he prefers nor that they are equally desirable. … How real this possibility is, both for individuals and for organizations, seems to be an extremely interesting question, but it is a question of fact. It certainly deserves further study." (p. 19) They also note that, without completeness, "a mathematical theory … is still possible" but involves "a more complicated and less satisfactory set-up." (p. 29)

[4] Though an incomplete preorder is not equivalent to the ordering induced by a single real-valued function, it can be represented by a *set* of such functions. Thus incomplete preferences can be represented by a set of utility functions, with $a \succcurlyeq b$ whenever all functions in the set agree in assigning a higher utility to a, and $a \perp b$ whenever at least two functions in the set disagree (Donaldson & Weymark, 1998). In this representation, the individual utility functions are analogous to the clocks of different observers in relativity theory (where the time order of two events is well-defined only if all clocks order the events in the same way).

Figure 1
Three Incomplete Relations



Note. (A) A DM assigns an imprecise value range to a hat (h). (B) For preferences defined by dominance in a two-attribute space, a is inferior to any option a^+ that dominates it ($a^+ \succ a$), superior to any option a^- that it dominates ($a \succ a^-$), and unranked with respect to other options ($a \perp b$). (C) An incomplete relation from physics: The "future light cone" of an event x consists of all events that can receive a physical signal sent from x. x's "past light cone" includes all events that can send a signal to x. According to relativity theory, all observers agree that event x happens before any event x^+ in its future light cone ($x^+ \succeq_T x$) and after any event x^- in its past light cone ($x \succeq_T x^-$), but its time relation to other events is not objectively defined ($x \perp_T y$). DM = decision maker.

in discussions of human rationality. Constructively, it opens up a class of new normative problems—the challenges of implementation theory—that an adaptive decision-making system must solve. While the constructive challenges are our primary interest here, we turn first to a brief review of destructive implications.

Incompleteness and Invariance

As we noted in the introduction, psychologists routinely rely on invariance principles in the normative analysis of behavior. These principles do not specify what a given DM should choose in a given situation—a judgment the researcher is rarely in a position to make. Instead, an invariance principle simply states that choices should not be affected by incidental features of the context, such as the frame or elicitation method. Invariance principles are powerful—and widely used—because they permit researchers to conclude that some choices are suboptimal, even when we have no idea what the optimal choice is for any DM.

Though invariance principles are commonly taken to be "normatively unassailable" (Tversky, 1996), their normative validity rests on two implicit assumptions that deserve close scrutiny. First, information must be held fixed across all choice contexts—that is, the DM must learn nothing from different frames, procedures, and other incidental variables. In prior research, we have argued that subtle choice-relevant information is in fact "leaked" by some common manipulations of choice context (McKenzie et al., 2006; Sher & McKenzie, 2006, 2014; Sher et al., 2022). But here we assume for the sake of argument that all context manipulations are uninformative.

Second, the standard view of invariance violations—that they imply suboptimal choice—implicitly assumes that rational preferences are complete. To see why, consider a between-subjects study in which choices are made from a finite menu M, and an incidental variable X, such as the frame or procedure, is manipulated.[5] Let $\succcurlyeq$ denote the normative ranking of options for a given DM—that is, $a \succcurlyeq b$ just in case, in a fully rational assessment of the available information, a would be judged more in accord with the DM's interests than b. Because X is a normatively irrelevant variable, it will not affect $\succcurlyeq$.

If $\succcurlyeq$ is complete, and if the options in M are complex and mutually distinct, we can often be sure that $\succcurlyeq$ will single out a *unique* option as best. Recall that rational indifference is not robust to infinitesimal enhancement of either option. Therefore, it is vanishingly improbable in a randomly selected pair of alternatives, and the uniqueness of the optimal item in a finite set can be safely assumed. If a is chosen in one condition and a distinct b in another, at least one of these choices must be suboptimal. But if $\succcurlyeq$ is incomplete, robust rational indecision is possible, and M need not have a unique optimal element. There may easily be two elements, a and b, with $a \perp b$ and no c in M such that $c \succ a$ or $c \succ b$. If a is selected in one condition and b in another, invariance is violated but no choice is suboptimal. For example, Eliaz and Ok (2006) identified conditions under which rational actors with incomplete preferences may choose a over b and yet price b over a, similar to choice-pricing "preference reversals" that violate procedure invariance (Lichtenstein & Slovic, 1971; see also Butler & Loomes, 2007).

To summarize, do effects of normatively irrelevant variables establish failures of rationality? The answer depends on whether we accept the completeness axiom. If rational preference is assumed to be complete, then any effect of a normatively irrelevant variable is normatively inappropriate: If it is not the case that X should influence choices, then it is the case that X should not influence choices. But if rational preference is allowed to be incomplete, "not-should" does not imply "should-not." Provided that no chosen option is ever ranked as inferior to an available alternative, effects of normatively irrelevant variables are themselves normatively irrelevant.

[5] The normative analysis in this section focuses on between-subjects invariance violations. A full normative analysis of the *within-subjects* case must also consider the threat of money pumps and other "incoherence traps." We address such traps in the next section, where we show how they can be avoided by a suitable implementation policy.

Importantly, this argument does not establish that subjects in, for example, typical framing experiments are making rational choices. It simply shows that invariance violations, on their own, are non-diagnostic with respect to rationality. Yet this simple destructive conclusion is a sweeping one, considering the central role invariance violations play in the literature on human rationality. As we show next, incompleteness has far-reaching constructive implications as well.

Implementation Theory

When preferences are incomplete, a theory of rational choice can be usefully divided into two components, shown in Figure 2. *Preference theory* imposes constraints on the structure of rational preferences—that is, of definite judgments that one option is better than another relative to the DM's values. *Implementation theory* asks how DMs convert preferences into choices—that is, into the behavioral selection of one option over others.[6] In this framework, note that the standard identification of preference with choice breaks down (cf. Sen, 1997). A systematic tendency to choose one option over another is ambiguous: It may arise either at the level of preference or at the level of implementation.

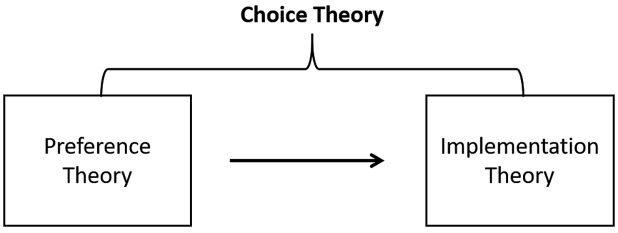
When preferences are complete, implementation theory is trivial: Choose what you prefer (and flip a coin when indifferent). When preferences are incomplete, one might suppose that implementation theory is irrelevant—that is, in the absence of a preference, you might as well choose by whim. However, this turns out not to be the case. There are better and worse policies for translating incomplete preferences into choices.

The problems of implementation theory are problems of choice composition. They arise when a collection of related choices is made. We can then evaluate the individual decisions in the collection separately, relative to the DM's preference order $\succcurlyeq$, or alternatively we can use $\succcurlyeq$ to evaluate the whole collection. The distinction is critical because, when $\succcurlyeq$ is incomplete, the whole may be less than the sum of its parts. The key challenge of rational implementation is to ensure that locally acceptable choices yield globally acceptable outcomes.

For concreteness, we compare four *implementation policies*. Each policy specifies a procedure for choosing in the absence of preference. Forced to choose between $a \perp b$, what shall the DM do? The first two policies below are relevant to the special case in which one option involves an action (i.e., a commission, such as a deliberate change to the status quo) and the other option involves inaction (i.e., an omission, or passive retention of the status quo). In this case, we have the following implementation policies:

1. In a *restless* policy, the DM acts unless she has a reason not to act.
2. In a *conservative* policy, the DM does not act unless she has a reason to act.

Figure 2
Bipartite Division of Rational Choice Theory



Turning to the more general case of unranked alternatives, all of which may involve acts that bring about similar departures from the status quo, we add two further policies:

3. In a *random* policy, the DM flips a coin for each arbitrary choice.
4. In a *coherently arbitrary* policy, the DM can make an arbitrary initial choice in any way, but this choice then establishes a precedent that future choices must respect.

Implementation theory asks: In various choice environments, how do the global outcomes of different policies for choice in the absence of preference compare? We focus on two kinds of environment here. The first involves sequences of simple stay–switch decisions. The second involves more richly structured choices among temporally extended plans. In each setting, we evaluate the normative appropriateness of different implementation policies, and we re-examine descriptive findings of history-dependence in light of the normative analysis. The discussion is largely informal, with an emphasis on psychologically relevant illustrations and applications. Formal treatments may be found in Mandler (2005), regarding the advantages of conservatism, and Danan (2010), regarding the drawbacks of randomness. We suggest extensions of their conclusions to coherent arbitrariness and the sunk cost effect.[7]

Staying or Switching

We start with the simplest type of choice problem in which challenges of intertemporal coordination arise. The DM occupies some state, and sequentially encounters offers to either retain her state or switch to another. The DM thus faces the timeless question, "Should I stay or should I go?" When the answer is arbitrary, what is a rational DM to do?

We assume a stable but incomplete preference relation $\succcurlyeq$ defined over a set of mutually exclusive states, one of which is occupied at each time. At various times, the DM in state S faces a menu of possible options, which includes S (i.e., she can always retain the status quo). Given a sequence of state transitions, $S_1 \rightarrow S_2 \rightarrow \ldots \rightarrow S_N$, we can then evaluate transitions locally, comparing S_i with S_{i+1}, or globally, comparing S_1 with S_N. When preferences are both complete and transitive, such sequences can never take DMs from more to less preferred states. But when preferences are incomplete, the wrong implementation policy can have just this effect.

Repeated Moves

Consider first the simple case of a DM choosing where to live. The DM may reside in either of two apartment buildings, one with a

[6] The preference–implementation distinction should not be confused with Gollwitzer and Bayer's (1999) distinction between "deliberative" and "implementational mindsets" in goal pursuit. The latter concerns the distinction between goal selection and subsequent goal execution, whereas the former distinguishes between the judgment that an option is best and (in the absence of such a judgment) the selection of a particular option.

[7] Our conclusions are also related to Cubitt and Sugden's (2001) analysis. Employing a more general framework for representing extended choice behavior, they show that history-dependent choice can avoid money pumps even when standard consistency assumptions are violated.

classic (C) and the other with a modern (M) style. The DM prefers apartments on higher floors in each building but lacks well-defined rankings across buildings. Letting C_i (M_i) denote the ith floor of the classic (modern) building, the DM's preferences are then given by the following relations: (a) For all $i > j$, $C_i \succ C_j$ and $M_i \succ M_j$; and (b) for all i, j, $C_i \perp M_j$. This preference structure is visualized in Figure 3.

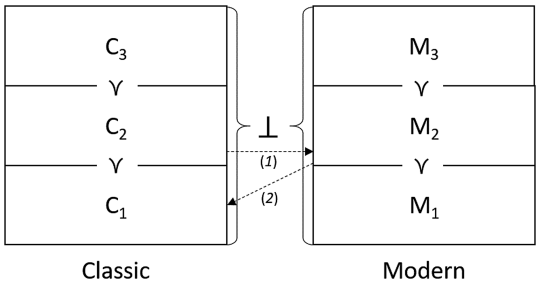
Suppose the DM, currently living on the second floor of the classic building, learns of an opening on the second floor of the modern building. Given that $C_2 \perp M_2$, it is acceptable but not necessary to move. A DM with a restless implementation policy will thus move, and a DM with a random policy will do so with probability ½. Now suppose the DM subsequently learns of an opening on the first floor of the classic building. Since $C_1 \perp M_2$, the restless DM will, and the random DM may, move again. The restless DM thus ends up in a clearly inferior apartment on a lower floor of her original building, and the random DM meets this fate with probability ¼. Though each individual move is consistent with the DM's incomplete preference order, the combined move ($C_2 \rightarrow C_1$) is not.

It is easy to see that a DM with a conservative or coherently arbitrary policy never falls for this "incoherence trap." The conservative DM, refraining from arbitrary moves not required by her preference order, will simply retain her original apartment. In a coherently arbitrary policy (described in more detail below), an arbitrary initial decision to move to M_2 establishes the precedent $M_2 \succcurlyeq C_2$. This precedent in turn implies $M_2 \succ C_1$, ensuring the DM does not subsequently move back to an inferior apartment in the old building. The simple example of repeated moves illustrates a lesson that will recur frequently in what follows: When preferences are incomplete, rational implementation is history-dependent.

Repeated Trades

We now turn to a traditional staple of normative decision theory: a market in which durable goods can be bought and sold. In each round, the DM can engage in trades of goods and/or money, but she always has the option to maintain the status quo. An analysis of implementation in this environment sheds surprising light on empirical disparities between buying and selling prices (e.g., Kahneman et al., 1990).

Figure 3
Incomplete Preferences That Rank Apartments Within but Not Across Buildings



Note. A restless DM will, and a random DM may, make move (1) followed by (2). Each move is locally acceptable, but their combined result is globally unacceptable. A conservative DM will make neither move. If a coherently arbitrary DM makes move (1), the resulting precedent blocks move (2). DM = decision maker.

Consider the DM in Figure 1A, who assigns a value between \$15 and \$20 to a hat (h), and suppose that a trader offers to sell the hat to the DM for \$19. A restless DM *will*, and a random DM *may*, agree. Later, the trader returns and offers to buy back the hat for \$16. A restless DM will again say yes, while a random DM will do so half the time. While both local trades (buy for \$19; sell for \$16) are consistent with the DM's incomplete preference order, they combine to yield an unacceptable global transaction (throw away \$3).

A conservative implementation policy is not susceptible to incoherence traps of this kind, as Mandler (2004) observed. The conservative DM only buys for prices below \$15 and only sells for prices above \$20. The DM thus only accepts unambiguous upward trades, and no sequence of voluntary trades can take her down from a state x to an inferior state $y \prec x$.

However, Mandler (2005) also showed that, while a conservative policy never falls for incoherence traps when a *single* good is held at each time, it is not strictly optimal in more complex environments where *multiple* goods can be independently traded. Intuitively, while the conservative DM will never trade down to an inferior state in those environments, she may miss rather subtle, multistep opportunities to trade up to a superior state.[8] Nonetheless, when stay–switch preferences are incomplete, status quo maintenance may often be "a simple and effective way to make self-interested decisions," as Mandler (2004, p. F531) argued.[9]

Next, we consider a coherently arbitrary implementation policy. Denote the DM's "true" incomplete preference order by $\succcurlyeq_0$. Forced to choose between $a \perp_0 b$, the DM makes an arbitrary choice in some way—say a is selected. In a coherently arbitrary policy, the DM then adopts the precedent that a is at least as good as b, a precedent that will bind her future choices in the same environment. Formally, the DM "updates" to a new working preference relation $\succcurlyeq_1$ with the following properties:

1. $\succcurlyeq_1$ extends $\succcurlyeq_0$ (e.g., if $x \succ_0 y$, then also $x \succ_1 y$).
2. $\succcurlyeq_1$ includes the new precedent $a \succcurlyeq_1 b$.
3. $\succcurlyeq_1$ is transitive.

Condition (3) means that $\succcurlyeq_1$ must include all rankings that result from combining old preferences with the new precedent (e.g., if $b \succcurlyeq_0 c$ and $a \succcurlyeq_1 b$, then $a \succcurlyeq_1 c$). This process is iterated, so that, when the nth arbitrary choice in a choice sequence is made, the working preference order $\succcurlyeq_{n-1}$ is updated to yield a new working order $\succcurlyeq_n$, which includes the new precedent along with its transitive implications.[10]

[8] For example, suppose the DM holds a bundle of two independent goods $<a, b>$, which are neither substitutes nor complements. If $a \perp b^+$ and $a^+ \perp b$, a conservative policy will refuse an initial offer to trade a for b^+ and will then refuse a subsequent offer to trade b for a^+. The DM has thus missed an opportunity to trade up to $<b^+, a^+>$, which dominates the initial bundle $<a, b>$.

[9] Note that this analysis assumes a simple environment in which "stay" and "switch" options do not systematically differ, except in their relation to the status quo. When staying and switching consistently differ in subtle ways (such as in effort and/or learning), other normative issues arise; these are addressed in the General Discussion section.

[10] We note that, while condition (2) only requires the DM to add a precedent corresponding to her arbitrary choice to $\succcurlyeq_n$, it does not rule out the addition of further precedents consistent with $\succcurlyeq_{n-1}$. For reasons that go beyond the scope of this article, such further precedents may serve to simplify the preference structure, facilitating computation of transitive implications. In contexts of risky choice (unlike the riskless situations we focus on here), precedents may also have to respect conditions of probabilistic coherence (e.g., the independence axiom of expected utility theory).

A coherently arbitrary implementation policy is vulnerable neither to downward trades nor to the kinds of subtle missed opportunities that Mandler (2005) identified. In the market example, the coherently arbitrary DM may initially choose to purchase the hat for \$19. But in doing so, she establishes the new precedent that $h \succcurlyeq \$19 \succ \$16$. This precedent ensures that she will decline the crafty trader's subsequent offer to buy back the hat for \$16. More generally, at each moment in time, the DM's past choices are behaviorally indistinguishable from those which would result from any *complete* transitive preference order $\succcurlyeq_\infty$ which extends the DM's present working order $\succcurlyeq_n$. Since $\succcurlyeq_\infty$ never falls for incoherence traps, $\succcurlyeq_n$ will not have fallen for such traps either.

These lessons generalize to any "stay–switch" problem in which a DM can willfully transition between mutually incompatible states, some of which are mutually unranked: Restless and random DMs are susceptible to a series of voluntary state transitions $S_1 \rightarrow S_2 \rightarrow \ldots \rightarrow S_N$, in which, for all i, either $S_{i+1} \succcurlyeq S_i$ or $S_{i+1} \perp S_i$ (i.e., each local state transition is acceptable), but $S_1 \succ S_N$ (i.e., the global state transition is unacceptable). But in repeated trades, repeated moves, and other sequences of incompletely ranked state transitions, conservative and precedent-setting implementation policies never transition from a superior to an inferior state.

Behavioral Findings

Studies of buying and selling confirm the pricing gaps predicted by a conservative implementation policy. In the endowment effect, participants demand more to sell a good that they have been endowed with than they would have paid to purchase it (Kahneman et al., 1990). Similar disparities between low willingness-to-pay (WTP) and high willingness-to-accept (WTA) prices are well-documented in studies of public as well as private goods (e.g., Banford et al., 1980). The rational analysis suggests that such conservative pricing gaps are adaptive. If human DMs were restless rather than conservative in arbitrary choice, we would exhibit a maladaptive "reverse endowment effect," buying at the high end and selling at the low end of our imprecise value range. Conservative implementation can also explain a host of related findings, in which behavior seems to exhibit a bias toward inaction (C. J. Anderson, 2003), including status quo maintenance (Samuelson & Zeckhauser, 1988), default effects (Johnson & Goldstein, 2013), omission bias (Spranca et al., 1991), and the acceptance–elimination effect (Yaniv & Schul, 2000). In a later section ("Evaluation or Implementation?"), we take a closer look at this body of evidence. We argue that these effects, which are often attributed to loss aversion, are more parsimoniously explained by conservative implementation.

Behavioral manifestations of coherent arbitrariness have also been documented. For example, an arbitrary initial refusal to accept \$2 to attend a poetry recital keeps people from later agreeing to attend for free, even if they might otherwise have made that choice (Ariely et al., 2006). Similarly, when an incidental contextual variable, such as a numerical anchor, affects pricing judgments for one item, prices for clearly inferior or superior items may subsequently shift in a consistent way (Ariely et al., 2003; cf. Vlaev et al., 2009).[11] Further evidence comes from choice-induced preference paradigms, in which an arbitrary choice generates congruent shifts in ratings of selected versus rejected alternatives (Enisman et al., 2021). Empirical patterns of history-dependent choice resemble those favored by a rational analysis of implementation.

Plan Selection and Execution

We turn now to a richer environment, in which causally interconnected costs and benefits are distributed over time. Humans are planning creatures. We select and execute temporally extended act-sequences for the benefits they are ultimately expected to yield. In the situations discussed above, DMs could switch between simple enduring states. In the case we consider now, DMs can switch between complex unfolding plans. When plan preferences are incomplete, how can intertemporal coordination of action be achieved?

The "sunk cost effect" occurs when a past investment in a course of action makes future investment more likely, even when expected costs and benefits are held fixed. Sensitivity to sunk costs is generally regarded as irrational. Because choices affect the future but not the past, it is thought that only anticipated future contingencies should impact current choices. Arkes and Ayton's (1999, p. 591) description of the sunk cost effect is representative:

> The sunk cost effect is a maladaptive economic behavior that is manifested in a greater tendency to continue an endeavor once an investment in money, effort, or time has been made. … A prior investment should not influence one's consideration of current options; only the incremental costs and benefits of the current options should influence one's decision. Nevertheless, several researchers have shown that people do attend to prior investments as they consider what course of action to take.

This standard view of the sunk cost effect, however, implicitly assumes that available plans can be *completely* ranked by desirability. If Plan A's expected outcomes are clearly superior to those of Plan B ($A \succ B$), then of course A should be preferred, regardless of past investment. But if the DM does not have a definite ranking of the plans based on expected outcomes ($A \perp B$), then we must ask about implementation rather than preference. We argue that a rational implementation policy for arbitrary planning decisions will sometimes honor sunk costs.

To motivate the argument, we begin with a simple example: Imagine a DM standing at a fork in the road. The DM faces a choice between two diverging paths, covering different distances and leading to different destinations. Suppose the DM is undecided between the paths but, for any fixed destination, prefers shorter to longer distances. The DM arbitrarily selects one path, taking a few steps down it. But at this point, even though the DM has learned nothing new, she reconsiders. She could continue on the present path or retrace her initial steps and take the other path instead. These two options, differing only marginally from the options she faced at first (i.e., the current path is now but a few steps shorter, the other path but a few steps longer), may also be unranked. If the DM ignores the past, the new choice is arbitrary, and she has no reason not to switch paths. But then the DM ends up with a total journey (the second path, together with the steps taken and then retraced along the first path) dominated by one available to her at the outset. In this situation, a history-blind implementation policy is plainly suboptimal, giving rise to wasteful vacillation between unranked

[11] Some failures to clearly replicate Ariely et al.'s (2003) anchoring effects have been reported (Fudenberg et al., 2012; Maniadis et al., 2014), though other studies have found anchoring effects in similar pricing judgments (Bergman et al., 2010; Sugden et al., 2013; Yoon et al., 2019). In relation to the normative analysis, the relevant question is not whether anchoring effects reliably occur, but whether, when they do, the effects translate into consistent shifts in the prices assigned to clearly related options in the same pricing context, as Ariely et al. found.

paths. It is wiser to commit to one path at the outset and then, barring relevant new information, stick with that arbitrary choice.[12]

What holds for paths holds, more generally, for plans. A generic plan can be represented as a series of small costly steps, none of which would be desirable in isolation but which jointly yield a desired reward. For example, in a home improvement plan, the steps might include the purchase of materials, meetings with a contractor, and hours of painting, while the reward is a more pleasant living environment. Because different available plans may incur different kinds of costs and yield different kinds of rewards, their ranking need not be well-defined. Table 1 lists two generic plans, A and B, involving different sets of costly steps ($\{s_i^A\}$ and $\{s_i^B\}$) and yielding distinct rewards (R^A and R^B). We assume that costs and benefits are known in advance, that the DM's preferences $\succcurlyeq$ over plans are stable but incomplete, and that the removal of a costly step always makes the resulting plan more desirable. Thus, in Table 1, we have $A^+ \succ A$, where A^+ is the marginally enhanced version of A obtained by removing its first small step s_1^A.

Suppose the DM is robustly undecided between A and B—that is, $A \perp B$ and also $A^+ \perp B$. Faced with a choice between A and B, she arbitrarily selects and initiates A. But once she has carried out s_1^A, she faces a forward-looking choice between A^+ and B. If her implementation policy is history-independent, she has no reason to make either choice. She is thus free to switch to B, but this results in a total sequence (B plus the wasted cost of s_1^A) that dominates a sequence (B alone) available to her initially. A committal implementation policy—in which the initiation of a plan's first costly step establishes a standing commitment to it—avoids this trap. Formally, commitment is simply a special case of coherent arbitrariness. This is because, in a coherently arbitrary policy, the initial selection of A over B establishes the precedent that $A \succcurlyeq B$, which (together with $A^+ \succ A$) implies that $A^+ \succ B$. The new precedent ensures that the DM will not switch to B (i.e., select B over A^+). It is, in effect, a commitment to the continuation of what has been begun. The coherently arbitrary DM is thus spared the wasted effort that results when arbitrary plan selection is followed by arbitrary plan switching.

In the context of plan selection and execution, coherent arbitrariness means that sunk costs will sometimes be honored. Consider a between-subjects experiment with two conditions, in which a hypothetical population of rational DMs face a choice from $\{A^+, B\}$, either (1) after having made a prior investment s_1^A that reduced A to A^+ or (2) in the absence of any prior investment. Suppose further that some DMs in the population have the incomplete preferences assumed above. In condition (2), such DMs are free to select B, but a coherently arbitrary implementation policy rules out this choice in condition (1). As a result, the rate of A^+ selection will be higher in condition (1) than in condition (2). It can be rational to honor sunk costs.

This does not, of course, mean that it is always rational to honor sunk costs. In the foregoing discussion, we have assumed that the DM's underlying true preference order $\succcurlyeq$ is stable. But if new information is received, a rational DM may need to revise her preferences; and if implementation is coherently arbitrary, this may require revision to precedents as well. In particular, if, after the initiation of A in Table 1, new information is received which suffices to establish that $B \succ A^+$ (i.e., B is superior relative to the DM's personal values) where the two alternatives were formerly unranked, then the DM should switch to B. A failure to switch to the clearly superior alternative would indicate excessive regard for sunk costs. But the present analysis shows that, when plan preferences are incomplete, it is possible to give too much *or* too little deference to sunk costs. There are dangers of over- and undercommitment alike.

Table 1
Three Plans With an Asymmetric Dominance Structure

Plan	Series of costly steps	Reward
A	$s_1^A s_2^A \ldots \ldots \ldots \ldots \ldots \ldots s_m^A$	R^A
B	$s_1^B s_2^B \ldots \ldots \ldots \ldots \ldots \ldots s_n^B$	R^B
A^+	$s_2^A \ldots \ldots \ldots \ldots \ldots \ldots s_m^A$	R^A

Does the literature allow us to conclude that people generally give excessive deference to sunk costs? We do not think so. First, as we detail in a footnote, the evidence for some of the most striking experimental sunk cost effects is, on close inspection, equivocal.[13] Second, while some experiments include new information which impressionistically seems to weigh against continuation of the present course of action, it is unclear whether this information should establish a definite ranking (i.e., of $B \succ A^+$, in the language of Table 1) for all participants. Researchers rarely bother to argue that it should, because they assume, incorrectly, that *any* effect of sunk costs must be irrational. While it would surely be unreasonable to assume that people have optimal sensitivity to sunk costs, the current literature does not clearly establish that sensitivity is "on the whole" higher or lower than it should be.

It is noteworthy, however, that the rational analysis seems to corroborate the reasons participants give for their deference to sunk costs. Arkes and Blumer (1985) found that these justifications often refer to a wish to avoid waste. Researchers generally take these justifications to be misguided, because they refer to the past rather than the future. But when preferences are incomplete, a rational implementation policy must be concerned with the coordination of past and future choice. The virtue of a policy that honors arbitrary commitments is precisely that it does not waste effort and other resources in the way that purely future-oriented policies do.

The present analysis is related to McClennen's (1990) notion of "resolute choice" (see also Machina, 1989). While McClennen focuses on incoherence traps arising from failures of independence[14] rather than completeness, his conclusion—that a DM may avoid

[12] For a rather different normative analysis of a problem with a similar structure, see Broome (2001).

[13] Two especially influential effects are the Michigan–Wisconsin ski trip study (Experiment 1) and the theater-ticket field study (Experiment 2) of Arkes and Blumer (1985). But it is doubtful whether the former is a sunk cost effect at all, because (as found in Experiment 9 of the same article) participants chose the more expensive trip about as often regardless of whether they paid the costs themselves or simply learned the prices from a travel agent (in which case there are no sunk costs). In the analysis of the theater-ticket study, sunk cost (three levels) and half of the season (first or second) are included as factors. Neither the main effect of sunk cost nor (apparently) the interaction is significant. Despite the nonsignificant interaction, the sole evidence for a sunk cost effect comes from comparisons in the first half of the season, which are significant in one-tailed tests. While this observation is suggestive, we do not think it can support the theoretical weight this study has been given in the literature.

[14] This condition arises in decision under risk and uncertainty. In one formulation, it says that if the DM prefers a to b, she prefers a gamble offering a with probability p to an otherwise matched gamble offering b with probability p.

such traps by resolutely binding her future self to comply with a selected plan of action—aligns broadly with ours. Bratman (1987) drew a similar conclusion in a philosophical theory of planning. He argues that, for creatures with cognitive resource limitations and intertemporal coordination requirements, effective intention requires some level of resistance to reconsideration. While Bratman emphasizes resource limitations, we show that commitment can be advantageous even for unboundedly rational DMs, provided that plan preferences are incomplete.[15]

Summary

When preferences are incomplete, a theory of rational choice requires an analysis of rational implementation. If implementation is purely forward-looking (i.e., blind to the DM's past decisions and present state), locally acceptable choices may combine to yield globally unacceptable outcomes. A sequence of permissible state transitions may substitute a less for a more preferred state, and plan execution may devolve into a globally wasteful patchwork of locally acceptable plan-parts. But if implementation is sensitive to the DM's past choices and/or present state, locally arbitrary acts can be integrated in globally appropriate ways. Rational implementation thus anchors the choice of future acts in the history of past decisions.

A rational analysis of implementation offers a unified account of a range of well-known behavioral findings, including the endowment effect, coherent arbitrariness, and the sunk cost effect. The rational analysis also motivates new empirical hypotheses. In the next section, we contrast cognitive processes of evaluation versus implementation. We argue that effects usually attributed to loss-averse evaluation are better explained by conservative implementation.

Evaluation or Implementation?

In a passage from the novel *Severina* by Rey Rosa (2011/2014, p. 14), a character stumbles across a seemingly trite aphorism of the 14th-century monk Yoshida Kenkō:

> "What are you reading?" she asked me.
>
> "Kenko, aphorisms."
>
> "Can I see?"
>
> I handed her the book. She opened it at random, somewhere in the middle.
>
> She read: "It is best not to change something if changing it will not do any good."
>
> "Well, that seems obvious," she said.
>
> "Aphorisms often do, don't they?"

The reader may share the character's quick reaction to Kenkō's aphorism—that it is so obvious as to be hardly worth stating. From the standpoint of traditional rational choice theory, however, with its perfect symmetry of action and inaction, the aphorism is anomalous. Kenkō could just as well have written that it is best to change something if changing it will not do any harm. But when preferences are incomplete, the direction of the aphorism is natural, both normatively and psychologically: Normatively, because changes that are harmless in isolation may be harmful in combination. Psychologically, because of our propensity toward conservatism in arbitrary choice.

This propensity is empirically well-documented, with many manifestations of omission bias (Spranca et al., 1991) and related forms of inertia (C. J. Anderson, 2003). As we noted above, conservatism also offers a simple explanation of status quo maintenance and the endowment effect (Gal, 2006). Yet here we confront an explanatory puzzle. These effects are commonly attributed not to conservatism but to loss aversion. Loss aversion, a basic tenet of prospect theory (Kahneman & Tversky, 1979), states that losses (defined relative to a psychologically salient reference point) loom larger than objectively equivalent gains. Loss aversion can explain the endowment effect, as it implies that the seller's loss of a good will be more painful than the buyer's gain is pleasurable. It can also account for status quo maintenance, if departures from the status quo are psychologically coded as mixtures of prospective gains and losses (Tversky & Kahneman, 1991). These effects may thus be explained in two different ways—either in terms of an asymmetry of evaluation (the subjective value of losses relative to gains) or, more directly, in terms of an asymmetry of implementation (a reluctance to act in the absence of a reason to act; i.e., Kenkō's maxim).

Why is the evaluation-based explanation generally favored (and the implementation-based explanation largely ignored) by psychologists and behavioral economists? We believe that, somewhat ironically, this explanatory bias is a vestige of expected utility theory, the traditional economic theory that prospect theory challenged. Because it assumes completeness, issues of implementation do not arise in expected utility theory: All outcomes are associated with precise probabilities and precise utilities; hence every option has a precise expected utility, and the ordering of options is always well-defined. As we explain below, prospect theory has a similar overall form, and similarly delivers a definite ordering of available options by subjective value (though this ordering depends on the reference point, which may differ across choice problems). Value imprecision—and the implementation problems it creates—thus have no more place in prospect theory than in expected utility theory. When one considers prospect theory's original aims, this similarity is unsurprising. As Kahneman (2000, p. x) later recounted in describing the theory's early development, "The goal we set for ourselves was to assemble the minimal set of modifications of expected utility theory that would provide a descriptive account of everything we knew about … choices between simple monetary gambles." This natural strategy no doubt contributed to prospect theory's great influence in economic applications. But it also imported an unrealistic idealization of expected utility theory into the psychology of decision making—the assumption of precise subjective values and complete rankings.

In this section, we critically examine the implications of this assumption. We show that implementation policies for incomplete preference, such as those discussed earlier, are bound to be misinterpreted in the standard prospect theory framework. We then discuss methods and findings that can empirically distinguish between loss aversion and conservatism.

[15] Recently, some researchers have argued that sunk cost effects can also be adaptive for other reasons, involving strategic signaling to one's future self (i.e., using an initial investment to signal one's current belief in the likely success of a plan; Hong et al., 2019), or the protection of one's reputation and self-image as a capable agent and reliable interaction partner (Doody, 2020).

Value Precision in Prospect Theory

Prospect theory posits two stages in the decision-making process—an "editing phase," in which the representation of options is variously structured and simplified, and an "evaluation phase," in which outcomes are assigned subjective values and probabilities are assigned subjective weights.[16] The weighted subjective values that are the outputs of the evaluation phase determine the DM's choice. Loss aversion arises in the evaluation phase, because the function that maps objective outcomes to subjective values is asymmetric, with a steeper slope for losses than gains. In addition to explaining the endowment effect, this asymmetry can explain why people tend to decline "symmetric mixed gambles" offering equal probabilities of equal gains and losses (an effect that can also be explained by conservatism, as we discuss below).

Because prospect theory's value function is perfectly sharp (i.e., every change from the reference point is associated with a well-defined subjective value), as is its weighting function (every outcome-probability is assigned a well-defined subjective weight), the evaluation phase always results in a definite ranking of alternatives. Within this framework, *any* systematic tendency to select one option over another can *only* be attributed to an underlying difference in evaluation. Of course, the sharpness of prospect theory's value function (like the assumption of completeness in expected utility theory) is only an idealization. For some purposes, it is surely a useful idealization. But it also introduces a distortion into psychological explanations which has often been overlooked: It guarantees that implementation effects, whenever they occur, will be misidentified as evaluation effects. In the prospect-theoretic formalism, the only way to accommodate effects of conservatism—a reluctance to act when imprecise values fail to dictate choice—is by introducing an asymmetry into the value function (i.e., loss aversion).[17]

Disentangling Loss Aversion and Conservatism

A more realistic extension of prospect theory would make two adjustments. First, some form of value imprecision would be incorporated into the evaluation phase, with the result that it sometimes but not always yields a definite ranking of options. Second, alongside the editing and evaluation phases, an "implementation phase" would be added to the decision process. In this enriched framework, status quo maintenance and the endowment effect have two possible explanations: loss aversion in the evaluation phase or, in line with Gal's (2006) suggestion,[18] conservatism in the implementation phase. These two pathways are psychologically distinct, though they are not mutually exclusive. The conservatism hypothesis is not tied to any specific assumption about the psychophysics of value; it neither assumes nor excludes an asymmetry in the perception of gains and losses. But importantly, conservative implementation predicts both status quo maintenance and the endowment effect, *even if* there is no evaluation asymmetry. That is, even if buyers and sellers were assumed to have identical imprecise value ranges for a good, conservative buyers would nonetheless buy low and conservative sellers would sell high.

How, then, can we empirically distinguish between loss-averse evaluation and conservative implementation? To do so, we must turn to other paradigms, in which predictions of loss aversion and conservatism potentially come apart. The remainder of this section reviews relevant findings, old and new, from such paradigms (see Table 2, for an overview). In our view, they yield a balance of evidence that strongly corroborates conservatism, but provides less support for loss aversion.

Active Versus Passive Status Quo Maintenance

One useful source of evidence comes from paradigms that unconfound inaction from the status quo. Ritov and Baron (1992) examined problems in which a commission is required to preserve a status quo that would otherwise change. They observed an omission bias in this case: Participants tended *not to act* to preserve the status quo. In addition, they expected to feel worse in the event that active retention of the status quo (as opposed to passive acquiescence in change) led to a bad outcome. Likewise, while people demand high prices to sell an endowment, they are not willing to pay similarly high prices to retain an endowment that would otherwise be lost (Gal & Rucker, 2018; Smitizsky et al., 2021).[19] If prospective losses of current possessions bring outsized pain, one would expect owners both to demand more to sell and also to pay more to retain. But only the former effect is robust, a result readily explained by conservatism. A general tendency toward inaction discourages arbitrary transactions of all kinds, whether they involve selling or paying to retain; it thus drives up selling prices while driving down retention prices for the same endowment. To be sure, a loss-aversion explanation of these findings is possible, but it requires the ad hoc auxiliary hypothesis that, when the inaction-associated state is distinct from the status quo, the former serves as the reference point. This auxiliary hypothesis has the counterintuitive implication that the default loss in the pay-to-retain paradigm is not a psychological loss at all, with retention of one's current possession instead counting as a gain. (In this paradigm, Smitizsky et al., 2021, found that participants do not describe their situation in these terms.)

[16] The editing phase includes operations such as re-coding outcomes as gains or losses relative to a reference point, and canceling components shared by all options. Note that, in prospect theory, editing occurs prior to evaluation; it is only in the latter phase that options are subjectively valued and a preference order is generated.

[17] Different probability weighting functions for gains and losses (Tversky & Kahneman, 1992) cannot provide a general account of such effects, which are observed in riskless as well as risky choice.

[18] Gal (2006) attributed inaction bias to a basic "psychological law of inertia," according to which, much as physical acceleration requires force, psychological change requires motivation. More focused mechanisms, including decision heuristics à la Kenkō, could also yield conservative implementation. And in a rather different theoretical framework, conservative implementation could be achieved by a rejection-biased setting of the starting point for a drift diffusion process, a "prevaluation bias" suggested by Zhao et al. (2020). To the extent that conservatism is a broadly adaptive response to value uncertainty, it would not be surprising if multiple mechanisms, suited to diverse task contexts, converge in favoring inaction when imprecise valuation makes choice arbitrary.

[19] In explaining this finding, Smitizsky et al. (2021) proposed that it reflects overgeneralization of high-pricing strategies that usually make strategic sense for sellers but that do not apply in the "pay-to-retain" paradigm. While this proposal accounts for the pay-to-retain findings, taken in isolation, it does not explain the related finding that, when participants are randomly endowed with one of two goods which they can trade for the other, rates of trade are less than 50% (Knetsch, 1989). Conservative implementation offers a common explanation for these similar effects.

Table 2
Effects Explained by Conservatism and/or Loss Aversion

Effect	Explained by conservatism?	Conservatism account	Explained by loss aversion?	Loss aversion account
Endowment effect	Yes	Refrain from transactions in imprecise value range	Yes	Seller's loss is more painful than buyer's gain is pleasurable
Rejection of symmetric mixed gambles (yes/no)	Yes	Gamble not clearly superior to status quo	Yes	Gamble clearly inferior to status quo
Low payments to retain	Yes	Refrain from transactions in imprecise value range	Unclear	Depends on ancillary assumptions about reference points
Reluctance to exchange equivalent items	Yes	Refrain from arbitrary trades	Unclear	
Acceptance–elimination effect	Yes	Refrain from arbitrary inclusions, exclusions	Unclear	
(In)action framing	Yes	2AFC frame obscures action–inaction distinction	No	Predicts gamble rejection in both frames
Assertion gap	Yes	Refrain from arbitrary assertions, denials	No	NA: No reference-dependent losses
Endowment-like effect with no endowment	Yes	Refrain from arbitrary assertions of preference	No	

Note. 2AFC = two-alternative forced choice; NA = not applicable.

Conservatism, by contrast, requires no special auxiliary hypotheses to explain these effects.

Exchanging Equivalent Items

Conservatism also provides a simple explanation for people's reluctance to trade a lottery ticket with which they have been endowed for another economically equivalent ticket (Bar-Hillel & Neter, 1996), despite the fact that both tickets offer equal probabilities of equal gains relative to the ticket holder's present state. Though probability distortion (Risen & Gilovich, 2007) or regret aversion (van de Ven & Zeelenberg, 2011) may contribute to this reluctance, it can be parsimoniously explained as a special case of a more general reluctance to engage in arbitrary exchanges. This reluctance is also found in trades involving goods that are hard to compare (such as mugs and Swiss chocolate bars; Knetsch, 1989) or strictly identical (such as two quarters; Gal, 2006), situations in which neither probability estimation nor expected regret seem relevant.

(In)Action Framing

A third important line of evidence comes from (in)action framing. As we noted above, the tendency to reject symmetric gambles (offering a 50% chance to win $\$x$ and a 50% chance to lose $\$x$) is often cited as evidence for loss aversion. Kahneman and Tversky (1979, p. 279) initially motivated loss aversion with this observation, noting that "most people find symmetric bets of the form (x, .50; $-x$, .50) distinctly unattractive." However, despite the major role this stylized observation has played in behavioral decision theory, the empirical literature on symmetric gambles is surprisingly mixed. Reviewing the literature, Yechiam and Hochman (2013) found that many studies fail to observe loss aversion in such choices (see their Table 1).[20] In many cases, people are about equally likely to choose a symmetric gamble or a sure $0 alternative, and when risk aversion is found, it is often modest in size and comparable to the level of risk aversion observed when all outcomes are uniformly translated into the gain domain, by the addition of a common baseline sum to all of them. Why, then, is a marked aversion to symmetric gambles commonly presupposed?

An intriguing clue comes from an experiment conducted by Ert and Erev (2013). Attitudes toward symmetric gambles may be examined in a yes–no task, in which participants are asked whether or not they would play the gamble, or in a two-alternative forced choice (2AFC) task, in which participants are asked to choose between a symmetric gamble and an explicit $0 alternative. Notably, published failures to find aversion to symmetric gambles (e.g., Battalio et al., 1990) have generally employed 2AFC tasks. Employing a symmetric gamble with high nominal payoffs in an experimental currency, Ert and Erev observed a high rate of rejection for the gamble only in a yes–no task but not in a 2AFC task. If a conservative implementation tendency is assumed, this observation is unsurprising. The yes–no framing clearly singles out one action (playing the gamble) for an up-or-down choice, while the 2AFC framing at least partly blurs the distinction between action and inaction, presenting the $0 selection as a second available act. Loss aversion, by contrast, would predict that symmetric gambles should be rejected both in yes–no and in 2AFC tasks.

Because Ert and Erev's observation is potentially critical, but the gamble they used is somewhat unusual (involving high nominal payoffs in an experimental currency, translating to low actual payoffs), we conducted a conceptual replication of their study, using standard U.S. currency at two different levels of stakes (±$10 or ± $100) in a hypothetical choice task. Participants were 472 workers

[20] Yechiam and Hoffman consider both description-based and experienced-based choice studies (Hertwig & Erev, 2009). Loss aversion is not consistently observed even in description-based paradigms, which are most relevant to the current discussion.

Table 3
Risk Seeking Rates for Two Symmetric Gambles and Two Action Frames

Format	(heads, +$10; tails, −$10)	(heads, +$100; tails, −$100)
Yes/no	37.5% (120)	28.4% (116)
2AFC	57.1% (119)	41.0% (117)

Note. Each gamble was judged alone (yes/no) or explicitly compared to a $0 option (2AFC) in a between-subjects design (condition sample size in parentheses). 2AFC = two-alternative forced choice.

on Amazon Mechanical Turk (251 male, 220 female, one did not disclose, $M_{age} = 39.8$), recruited with CloudResearch (Litman et al., 2017).[21] They were randomly assigned to make either a yes–no decision regarding a symmetric gamble (whose payoff was said to be determined by a coin toss) or a 2AFC between the gamble and an explicit null option (in which nothing was to be won or lost regardless of which side the coin landed on). Full materials, data, and analysis code for all studies reported in this article are available at https://osf.io/zbmvk/. The studies were not preregistered. The research was certified as exempt by the UC San Diego Institutional Review Board.[22]

The results are shown in Table 3. In the yes–no format, there is a strong tendency to reject symmetric gambles, both for stakes of ± $10 (37.5% acceptance, $p < .01$, two-tailed binomial test) and for stakes of ±$100 (28.4% acceptance, $p < .001$). But rates of gamble selection rise substantially in the 2AFC format, both for lower stakes, $\chi^2(1, N = 239) = 9.25$, $p = .002$, and for higher stakes, $\chi^2(1, N = 233) = 4.06$, $p = .044$. The gamble selection rate does not significantly differ from 50% in either 2AFC condition (p's ≥ .064), though it is lower when the stakes are higher, $\chi^2(1, N = 236) = 6.13$, $p = .013$, consistent with past findings that risk aversion increases when stakes are scaled up (Hogarth & Einhorn, 1990; Weber & Chapman, 2005).

These results replicate Ert and Erev's (2013) finding that reactions to symmetric gambles are highly sensitive to the (in)action framing suggested by the task. When the action–inaction distinction is blurred by the 2AFC format, the aversion to symmetric gambles predicted by prospect theory does not appear to be robust. But if implementation is conservative, framing an act as omission or commission may have a decisive influence on arbitrary choices that could go either way, even in tasks (like this one) with only one plausible reference point.

Acceptance Versus Elimination

A related example of (in)action framing is the acceptance–elimination effect (Huber et al., 1987; McDonald et al., 2014; Yaniv & Schul, 1997, 2000), in which an option is more likely to be retained in a choice set when undesirable options are actively eliminated than when desirable options are actively selected. A loss aversion account of the effect would require special assumptions about reference points (which reference points are relevant for the task, and how and why do they differ between acceptance and elimination?). But the effect is directly predicted by conservatism, with no need for ad hoc assumptions. A bias toward inaction in "borderline" cases (i.e., those in which the participant lacks a definite preference regarding the item's placement) should result in more inclusions when the act is exclusive and more exclusions when the act is inclusive.

No-Loss Paradigms

The scope of conservatism is wide, extending well beyond the settings usually studied by decision researchers. We face myriad possibilities of arbitrary action—where there is no clear reason to act or not to act—even in domains where acts are not mentally coded in terms of prospective gains or losses. Conservatism thus predicts effects in "no-loss paradigms" which closely parallel familiar findings in the decision-making literature, but which cannot be explained by loss aversion. In this section, we focus on linguistic assertion, where a speaker may lack a clear reason to affirm or deny a statement, but no losses are on the line. First, we report an assertion-based analogue of the acceptance–elimination effect. Next, we turn to assertion gaps that mirror standard WTP–WTA pricing gaps, including an endowment-like effect in the absence of any endowment.

Active Versus Passive Assertion. Just as borderline options are retained passively but not actively in the acceptance–elimination effect, conservatism predicts that borderline assertions will be passively but not actively endorsed. Table 4 reports the results of a study in the domain of attitude expression. We asked 160 undergraduates at UC San Diego (49 males, 111 females, M_{age} = 20.0) whether a list of 16 attitude statements in a paper-and-pencil questionnaire applied to them. Half of the participants were instructed as follows: "If a statement is true for you, write a check (✓) to the left of the statement. If a statement is not true for you, just leave the statement unmarked." For the remaining participants, the response procedure was inverted, as follows: "If a statement is not true for you, cross out the statement (like ~~this~~). If a statement is true for you, just leave the statement unmarked." It is hard to see how loss aversion could be relevant here, as no clear reference-dependent losses are at stake. Nonetheless, conservatism predicts that attitudes are more likely to be expressed via inaction (failing to cross out a statement) than via action (checking a statement). This is what we found. For every statement, the proportion not crossing out exceeded the proportion checking, and this difference was individually significant for 11 items. Arbitrary assertion, like arbitrary choice, appears to be conservative.

Assertion Gaps. A similar conservatism has been documented in discourse with vague language. Speakers refrain from applying both vague terms (e.g., "tall") and their negations ("not tall") to borderline cases (a man who is 5 ft. 10 in. in height). This results in the "assertion gaps" reported by Bonini et al. (1999), in which the shortest man who is said to be "tall" is taller than the tallest man said to be "not tall" (cf. Alxatib & Pelletier, 2011). Assertion gaps mirror standard pricing gaps in WTP–WTA studies, in which the lowest price for which a good is sold exceeds the highest price for which it would be bought. But while conservatism predicts both effects, loss

[21] Six additional participants were excluded, because they failed an attention check, along with six further responses with duplicate IP addresses or Mechanical Turk IDs.

[22] For all studies reported here, the rule for terminating data collection was determined in advance (a fixed sample size for the first two studies; a minimum sample size followed by a time-based stopping rule for the third). For all studies, we report all data exclusions, all manipulations, and all measures.

Table 4
Acceptance Versus Nonrejection of Attitude Statements

Attitude statement	% checking	% not crossing out	*p*
I like running	41%	57%	.048
I like thrillers	59%	68%	.251
I like Diet Pepsi	10%	23%	.032
I like hamsters	53%	61%	.264
I like cooking	75%	91%	.006
I like heavy metal	6%	10%	.385
I like motorcycles	20%	43%	.002
I like poetry	36%	68%	<.001
I like calligraphy	45%	70%	.001
I like rollercoasters	63%	70%	.316
I like Jerry Seinfeld	9%	29%	.001
I like hiking	64%	90%	<.001
I like modern art	39%	75%	<.001
I like boxing	21%	41%	.006
I like rain	69%	84%	.026
I like science fiction	65%	71%	.396

Note. For each condition (checking, crossing out), $n = 80$. A single response to the "running" item in the crossing-out condition was unclear and excluded from the analysis. Proportions compared with χ^2 tests.

aversion cannot explain assertion gaps. In a companion article (Sher et al., 2024), we extend the rational analysis to vague discourse. We argue that conservatism and coherent arbitrariness may also be adaptive in this domain, blocking logical contradictions in vague discourse in much the way they block money pumps in sequential choice.

Notably, assertion gaps can be leveraged to generate an endowment-like effect in the absence of any endowment. Table 5 presents the results of a study, conducted on undergraduates at the Rady School of Management, that we report in more detail in online Supplemental Material. Adapting a task from Bonini et al. (1999), we elicited monetary cutoff points in hypothetical choices involving a UC San Diego mug. Two *trade* conditions tested for an endowment effect: Some participants stated the smallest amount for which they would sell the mug, while others stated the largest amount for which they would buy the mug. In two further *choice* conditions, participants imagined choosing between the mug and some amount of money. Those in the "choose-money" condition stated the smallest amount for which they would choose the money over the mug, while those in the "choose-mug" condition stated the largest amount for which they would choose the mug over the money.

Note that the choice conditions offer gains but no losses. Yet along with an endowment effect for buyers and sellers, conservatism predicts an endowment-like effect for choosers. If speakers refrain from arbitrary assertion, they should assert that they would choose the money only for amounts *above* the mug's imprecise value range (analogous to the selling price in the endowment effect). But they should assert that they would choose the mug only for amounts *below* the range (analogous to the buying price).

Table 5
Value Gaps in Trading and Choice

Task	Stated monetary value	*N*	*M* (*SD*)	*p*
Trade	Min. selling price	115	$11.23 (7.62)	<.001
	Max. buying price	118	$7.02 (4.89)	
Choice	Min. "choose-money" price	116	$9.50 (6.94)	.01
	Max. "choose-mug" price	115	$7.36 (5.55)	

Note. Min. = minimum; Max. = maximum.

As Table 5 shows, both predictions were confirmed. In the trade conditions, minimum selling prices exceeded maximum buying prices. A parallel gap was seen in choice, where the minimum amount preferred to the mug exceeded the maximum amount to which the mug was preferred. (See online Supplemental Material for a full analysis of these effects.) Both effects are predicted by a reluctance to engage in arbitrary action, encompassing assertions ("I would choose $\$x$ over a mug," in the range of uncertain valuation) as well as trades (selling the mug for $\$x$ in the same range). Conservatism can generate endowment-like effects with no endowment and no prospect of a loss.

Summary

When imprecise evaluations are assumed to be precise, implementation effects will inevitably be misidentified as evaluation effects. Conservative implementation will look like loss aversion. The two accounts make identical predictions in the best-known paradigms, which feature a single focal act (buying, selling, or playing a symmetric gamble) bringing about a departure from a single plausible reference point (the status quo). But in other paradigms, where the implications of loss aversion and conservatism come apart, conservatism appears to have greater explanatory power: When omission is unconfounded from the status quo, effects that are directly predicted by conservatism can be explained by loss aversion only with the aid of ad hoc auxiliary hypotheses about the reference point. (In)action framing has robust effects which are unlikely to be explained by reference-dependent evaluation. And conservatism explains parallel findings in no-loss paradigms, where loss aversion makes no predictions.

But as we noted above, loss aversion and conservative implementation are not mutually exclusive. And with one exception (the 2AFC findings in Table 3), the phenomena reviewed above are not strictly inconsistent with loss aversion, even if they are not clearly predicted by it. Some researchers find independent support for loss aversion in more general findings of "negativity bias" (i.e., greater attention and responsiveness to negatively valenced stimuli; Baumeister et al., 2001; Rozin & Royzman, 2001). But recently, other researchers have argued that many of these effects are attributable to asymmetries in the real-world distribution of positive and negative stimuli (Shin & Niv, 2021), not to an internal bias in evaluative processing (Unkelbach et al., 2021).

Some specific evidence for conservatism can also be explained by other hypotheses. For example, Ert and Erev (2008) proposed a "lemon avoidance heuristic" to explain the effects of (in)action framing. In some social environments, direct yes–no offers may signal attempts at manipulative salesmanship, and participants may have internalized a heuristic to decline such offers (thus avoiding potential "lemons"—i.e., defective or otherwise undesirable products). But while other cognitive processes may well contribute to these effects, we are not aware of an alternative hypothesis that can explain all of them (e.g., a lemon avoidance heuristic would not explain the findings in Tables 4 and 5). In our view, considerations of parsimony favor a conservatism account of these effects.

Given the ubiquity of inaction biases not explained by loss aversion, we suggest that future tests of loss aversion should focus on predictions that cannot also be derived from simple conservatism in the implementation phase. Specifically, research

should strategically target loss aversion predictions for forced-choice situations in which there is no psychologically plausible distinction between greater and lesser degrees of action. We have seen that conservatism makes successful predictions in no-loss paradigms. The critical question is whether loss aversion has predictive power in paradigms with no action–inaction distinction. Tests of this kind would be ideally suited to clarify whether there is an evaluation asymmetry that subjectively magnifies losses, over and above an implementation asymmetry that favors inaction when choices are arbitrary.

General Discussion

The simple observation that human values are imprecise, and human preferences are incomplete, has complex and wide-ranging implications for the rationality and psychology of decision making. At the normative level, it implies that rational choice need not be invariant to frames, procedures, and other incidental features of the decision context; and should not be invariant to the DM's history, even when future costs and benefits are held fixed. When preferences are incomplete, history-blind decision making exposes the DM to incoherence traps, which are evaded by implementation policies that take the DM's past choices and/or present state into account. A rational analysis of implementation provides a unified perspective on a range of well-known behavioral "anomalies," including status quo maintenance, coherent arbitrariness, and the sunk cost effect. It also raises questions about the explanatory scope of prospect theory. Because it inherits expected utility theory's idealization of precise values, prospect theory is bound to mistake conservative implementation for loss aversion. Yet we have argued that the endowment effect and the tendency to reject symmetric gambles are more parsimoniously explained by conservatism—a hypothesis that also accounts for low payments-to-retain, the acceptance–elimination effect, effects of (in)action framing, and parallel findings in no-loss paradigms.

While normative problems of incomplete preference have received growing attention in economic theory (e.g., Mandler, 2005) and philosophy (e.g., Bales et al., 2014), these problems have usually been overlooked by psychologists. Thus, while our framework builds on a long tradition of "rational analysis" in cognitive science, it also goes beyond that tradition in an important way. J. R. Anderson (1991, p. 472) noted that the first step in a rational analysis is "to specify the goals being optimized by the system." From Anderson's pioneering work to the present (e.g., Lieder & Griffiths, 2020), the system's goals are generally represented by a real-valued function that assigns precise utilities to outcomes. As Gershman et al. (2015, p. 273) put it, "maximizing some measure of expected utility provides a general-purpose ideal for decision-making under uncertainty." Proponents of rational analysis have thus mostly shared with critics of rationality the assumption that the normative ranking of outcomes (specifying which outcomes are better or worse vis-à-vis the system's goals) must be complete. We have seen that, when this assumption is relaxed, the rational analysis acquires a new shape, with a bifurcation into distinct problems of preference and implementation (Figure 2). Patterns of history-dependence that are puzzling at the level of preference, such as endowment and sunk cost effects, may be natural at the level of implementation.

In an insightful recent contribution, Walasek and Brown (2023) also criticized the utility representation of preference. They focus on cases of "incommensurability," in which (a) there is a complete ranking of options within each of several "covering values," but (b) no comparisons are defined across covering values (similar to the two-attribute preference structure shown in Figure 1B). Our approach agrees with Walasek and Brown's in the central role it assigns to noncomparability, but it also differs in important ways. First, the preorder representation we employ is more general. It encompasses incommensurable values, yet generalizes to other forms of noncomparability in which there is either *less* structure (such as an imprecise value range on a single dimension; Figure 1A) or *more* structure (such as trade-off weights between dimensions that are partially but imprecisely specified). Second, Walasek and Brown suggest that, when values are incommensurable, resulting choices must be inconsistent. In contrast, our analysis of rational implementation shows that costly inconsistencies can be systematically avoided.

We are not, of course, claiming that all human decisions (or all instances of history-dependence) are rational. They are not. To recognize this obvious fact, no special training in decision theory is needed; it suffices to open a newspaper. But we *are* claiming that, in our assessments of human rationality, psychologists often use the wrong normative benchmarks. These benchmarks are derived from a theory of rational choice that assumes completeness. When we employ them, we will mistake even an innocuous effect of context (e.g., a frame that leads a DM to arbitrarily select a rather than b, when $a \perp b$) for a major failure of rationality. They can also lead to perverse "debiasing" prescriptions. For example, a decision researcher—operating under the standard assumption that sunk costs should never be honored—may design nudges or training to eliminate sensitivity to sunk costs. But if such techniques were to be truly successful, they would undermine commitment whenever planning decisions have an element of arbitrariness (i.e., much of the time), resulting in capricious plan abortion and vacillation and considerable waste in time and effort.

Nonetheless, findings of robust context- and history-dependence—even if they do not demonstrate irrationality—*do* challenge the descriptive validity of traditional economic theory. Because it assumes completeness, it cannot explain effects of frame, procedure, present state, and past history, when anticipated costs and benefits are fixed. Yet insofar as these "anomalies" simply reflect the incompleteness of human preferences, they likely would not have surprised the original developers of axiomatic utility theory, who acknowledged that completeness is only a mathematically useful idealization (von Neumann & Morgenstern, 1944; cf. Footnote 3 above).

In the remainder of this discussion, we first consider extensions of the present normative framework to other environments and types of decisions, at the individual and social level. We then address limitations of the framework and consider alternative representations of incomplete preference. We conclude with a comment on the broader relationship between empirical studies of value structure and normative theories of rational choice.

Extensions

Our analysis has focused on two simple kinds of choice problem, involving transitions between incompletely ranked states and the execution of incompletely ranked plans. In future research, it may be possible to extend the analysis to a richer range of problems and situations.

Asymmetric (In)Action Effects

In the choice problems we have examined, individual options available for action do not systematically differ from those associated with inaction. The same goods may be traded or retained, and more generally the local costs and benefits of staying versus switching are not different in kind. But in some real-world environments, actions (compared to inaction) may be associated with distinctive subtle costs and/or benefits. For example, actions may differentially incur small costs in physical or mental effort. On the other hand, if the status quo is better known than its alternatives, action may on average yield more valuable information than inaction (Cohen et al., 2007). Small consistent costs and benefits of this kind would in effect create systematic differences between the subspaces of available action-options and available inaction-options, when options are fully described to explicitly represent all of their attendant consequences. If these asymmetric consequences are small, they may fail to dictate choice in most local instances of robust indecision (e.g., a slight gain in information may not suffice to resolve preference uncertainty in an individual case). Yet they may be important in a global assessment of many local choices. For example, a conservative policy may turn out to be cumulatively maladaptive in an environment where action offers small consistent information gains and the long-term premium on learning is high. Depending on the (in)action cost–benefit asymmetries that characterize a given environment, a normative analysis of conservatism may thus arrive at very different conclusions. The analysis of subtle (in)action asymmetries in real-world choice environments is an important direction for future research.

Decisions From Experience

The rational analysis can also be extended to "decisions from experience," in which probabilities and outcomes are learned through repeated sampling rather than by explicit description. Notably, risk attitudes predicted by prospect theory are not always observed in decisions from experience, where rare outcomes appear to be under- rather than overweighted (Hertwig & Erev, 2009). At a more fundamental level, Erev and Marx (2023) challenged the assumptions—usually taken for granted in decision research—that (a) participants "read, understand, and believe" descriptions of options (Erev, 2020), and (b) an initial judgment phase is followed by a separate decision-making process (Erev & Plonsky, 2022). While we have relied on these assumptions in interpreting empirical findings, we note that the rational analysis does not ultimately depend on them. It applies whenever values are imprecise, leading to robust indecision and the need for an adaptive implementation policy. For example, we predict that participants will continue to report wide ranges of value imprecision (e.g., in Cubitt et al.'s, 2015, procedure) when gamble properties are learned via sampling. Conservative implementation may then be adaptive in this setting as well, consistent with demonstrations of an endowment effect in decisions from experience (Pachur & Scheibehenne, 2017). However, because decision-from-experience paradigms involve repeated sampling and learning, behavior is likely to be jointly determined by several factors, including sampling strategies (Teodorescu & Erev, 2014) and sample-based inferences (Sher et al., 2022), alongside implementation rules for arbitrary choice. Future research should clarify how these factors interact when a DM confronts "indecision from experience."

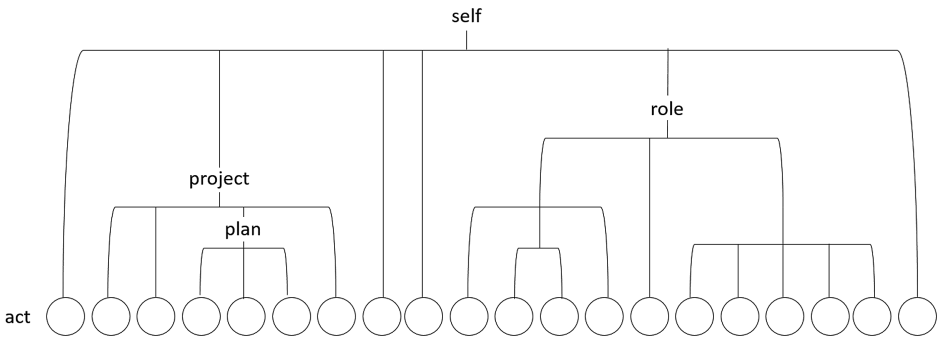
Computational Limits

The rational analysis may also be enriched by incorporating computational constraints (Lieder & Griffiths, 2020). For example, although a coherently arbitrary implementation policy has subtle advantages over conservatism (cf. Footnote 8), it is more computationally demanding and may be less robust to error (since it requires accurate memory of, and inference from, precedents). Trade-offs of this kind can be investigated in computer simulations that compare DMs with shared incomplete preferences, who employ different implementation policies in matched trading environments, subject to various performance limits (e.g., computing costs and/or nonzero error rates). In online Supplemental Material, we report one illustrative set of such simulations. They show that, when memory for past precedents is imperfect, a conservative policy can outperform coherent arbitrariness. Further simulations, exploring other choice environments and computational limits, may motivate more focused predictions about which implementation rules are best-adapted to specific contexts.

A Choice Hierarchy

In addition to studying action in other environments and subject to computational constraints, we can also consider action on multiple scales. Human choices can be arranged in a hierarchy (Figure 4), in which individual acts at the lowest level are organized,

Figure 4
A Hierarchy of Constructive Choice



at progressively higher levels, into plans, projects, roles, and finally the life history of a narrative self (cf. Bruner, 2004). The foregoing analysis has been limited to the lowest levels of the hierarchy, where acts transition between states and merge into plans. But a global evaluation of many local acts can also be undertaken at higher levels in the hierarchy. Here, instead of asking whether they yield an undominated state or plan, we ask how individual acts collectively cohere in the enactment of a social role or in "the construction of the self."

The self-construction of a personal identity poses complex problems that are unlikely to succumb to a neat decision-theoretic treatment. Nonetheless, they may be the most significant problems of preference construction in human life. Low-level acts must be selected in a way that preserves the integrity and stability of the far larger scale structures they ultimately compose. Buchanan (1979) noted that active self-construction challenges the dominant explanatory paradigm in economics. Because it assumes a stable utility function that is fixed prior to choice, this paradigm cannot account for human projects of self-improvement, in which we attach value to the prospective transformation of our own values. As Buchanan wrote (p. 94), "We are, and will be, at least in part, that which we make ourselves to be. We construct our own being, again within limits." As a result, "We are not, and cannot be, the 'same person' in any utility-maximization sense." (p. 100)

A normative analysis of self-construction would have to address issues well beyond the scope of this article (cf. Callard, 2018; Korsgaard, 2009; Oretga y Gasset, 1941). But though far from the whole story, we believe coherent arbitrariness is likely to play a part in it. At higher levels of the choice hierarchy, competing roles and narratives may embody values that appear to be incommensurable (Chang, 2015), and a given role is bound to be multiply realizable. Choices among roles and their realizations may then involve elements of arbitrariness. But because of their complexity, longevity, and social embeddedness, roles and narratives may make especially stringent demands of action coordination, inter- as well as intrapersonal. At these higher levels, the "coherence" in "coherent arbitrariness" must integrate considerations that go far beyond the pricing of goods and the execution of plans. Rather, considerations of *narrative coherence* may be paramount. The adaptive value of narrative coherence, even where it is partly arbitrary, may perhaps be relevant to more complex forms of behavioral consistency-seeking that appear when the self-concept is at stake (Aronson, 1999; Stone & Cooper, 2001).

A historical illustration of arbitrary implementation across the choice hierarchy comes from Descartes' (1637/1985) *Discourse on Method*. Descartes reports how, even as he undertook to subject his entire belief system to systematic doubt, he resolved in his personal conduct:

> to be as firm and decisive in my actions as I could, and to follow even the most doubtful opinions, once I had adopted them, with no less constancy than if they had been quite certain. In this respect I would be imitating a traveler who, upon finding himself lost in a forest, should not wander about turning this way and that, and still less stay in one place, but should keep walking as straight as he can in one direction, never changing it for slight reasons even if mere chance made him choose it in the first place; for in this way, even if he does not go exactly where he wishes, he will at least end up in a place where he is likely to be better off than in the middle of a forest. (Descartes, 1637/1985, p. 123)

Coherent arbitrariness is the best policy for the lost traveler. Descartes suggests that it may also straighten our ways in the journey of life.

Social Decision Making

The present framework may also be extended from individual to social decision making. One must then globally coordinate, not just the different actions of a single DM across time, but the actions of different DMs within an institution, organization, or social group. Because group goals often inherit, and sometimes surpass, the incompleteness of individual preference, systems of group decision making must also solve problems of rational arbitrariness.

Such problems are common in the law. Like personal values, legal statutes and norms are generally incomplete (Pistor & Xu, 2003). Legal incompleteness springs from several sources, including the incomplete agreements that legislators manage to reach, the vagueness of the language in which those agreements are expressed, and the potential incommensurability of legally relevant considerations (Sunstein, 1994). At the same time, consistency in the application of the law is critical for multiple reasons, relating to justice, predictability, and efficiency. In law as in life, inconsistency can be costly.

In jurisprudence, the principle of *stare decisis* (i.e., respect for judicial precedent) is, in effect, an institutional norm of coherent arbitrariness, favoring consistency in the application of disputed or incompletely specified law. Precedents, in legal as in individual decision making, help to globally coordinate local decisions that inevitably include an element of arbitrariness. To be sure, adherence to precedent may be too rigid, in judicial as in individual choice. The proper scope of *stare decisis* is, accordingly, a matter of continuing controversy. But even critics who allege that judicial deference to precedent is over-rigid generally concede that *some* degree of deference is necessary. As Justice William Douglas, a prominent critic of excessive obedience to precedent, wrote, "*Stare decisis* serves to take the capricious element out of law and to give stability to a society. It is a strong tie which the future has to the past." (Douglas, 1949, pp. 735–736) Because it is irreducibly incomplete, law without precedent (implicit or explicit) would be inescapably inconsistent, hence unpredictable and unjust.

A similar analysis may apply to other organizations with ambiguous collective goals (March & Olsen, 1975). Perhaps it could also shed light on some forms of stable arbitrariness in cultural traditions (Shils, 1981). In adaptive decision-making systems, individual or social, for which action (a) is partly arbitrary but (b) demands intertemporal coordination, past acts will, in part, constrain future decisions.

Limitations

Following the standard practice in economic models of incomplete preference, we have employed a preorder, in which a transitive ordering of options is sometimes defined (e.g., $a \succcurlyeq b$) and sometimes undefined ($a \perp b$). This is the simplest modification of traditional rational choice theory that accommodates robust indecision, and as such is a natural step toward greater realism in the representation of value. But the preorder representation has important limitations, relating to higher-order value imprecision, value conflict, and epistemic incompleteness.

Higher-Order Value Imprecision

A preorder relaxes the unrealistic assumption of precise monetary indifference points by substituting an indecision range (Figure 1A). In doing so, it retains the "second-order" assumption that the boundaries of the indecision range are precise. But the same psychological considerations that make sharp indifference points implausible also make sharp second-order indecision boundaries implausible.

Higher-order value imprecision has received little attention in economic models of incomplete preference. Yet it is the decision-theoretic counterpart of a problem which has attracted considerable attention in the literature on vague concepts—the problem of "higher-order vagueness" (Sainsbury, 1996). In a companion article (Sher et al., 2024), we argue that the problems of higher-order value imprecision and higher-order vagueness are not really distinct problems, because value imprecision is, at root, a species of vagueness. But while formal representations of higher-order vagueness have been proposed (e.g., fuzzy set theory; Zadeh, 1965), none to our knowledge is fully satisfactory (cf. Osherson & Smith, 1981). Importantly, higher-order imprecision complicates the measurement of indecision. If second-order indecision boundaries are imprecise, attempts to precisely locate them are bound to be sensitive to the elicitation method (Dubourg et al., 1997; Nielsen & Rigotti, 2022), much like attempts to precisely measure first-order preference (Fischhoff, 1991). How best to model and measure the "fine texture" of value imprecision is an important challenge for future research.

Value Conflict

A related limitation is that a preorder representation treats arbitrariness as an all-or-none construct. If no definite ranking of alternatives is specified, the choice is purely arbitrary. For when $a \perp b$, the preorder contains no information that could be relevant to rational deliberation as to whether a or b should be chosen.

But indecision can come in many forms (e.g., ambivalence; Schneider & Schwarz, 2017), some of which raise more complex problems. The setting of a precise buying price (Kahneman et al., 1990) or the choice of whether to attend a poetry recital (Ariely et al., 2006) may confront the DM with relatively pure forms of arbitrariness. But in cases of deep internal conflict (as when two internalized value systems cannot be reconciled), the DM is unlikely to regard her choice as arbitrary, even if her values do not deliver a clear verdict for action (Sher & McKenzie, 2022). A well-known example is Sartre's (1946/2007) student, who was torn between leaving the country to fight the Nazis and staying to support the mother who depended on him. In such cases, the choice of an act is tangled up with the choice of an ethic (Levi, 1986). Even if one regards the best choice as ill-defined, the decision is anything but trivial. A far richer representation of value is then needed, with more information than a simple preorder can contain, suited to the higher levels of the choice hierarchy (Figure 4).

Normative and Epistemic Incompleteness

The traditional theory of rational choice assumes that the DM always knows her preferences. Similarly, our rational analysis assumes that the DM always knows when she has no preference. If $a \perp b$, the implementation policy automatically kicks in. But for human DMs with limited cognitive capacity, it may not always be easy to distinguish arbitrariness (no option is best) from ignorance (the best option is well-defined but unknown).

A rational analysis must then consider both *normative* and *epistemic* incompleteness. We have focused on normative incompleteness, in which the DM's values do not logically imply a ranking. In epistemic incompleteness, by contrast, one option is truly best vis-à-vis the DM's values, but the DM does not know which. When the boundedly rational DM is unsure whether incompleteness is normative or epistemic, complex issues arise that lie beyond the scope of our analysis. For example, although we have seen that coherent arbitrariness is broadly adaptive in the case of normative incompleteness, it can have the maladaptive effect of entrenching initial errors when incompleteness is merely epistemic.

The boundedly rational DM also faces the problem of "deciding when to decide" (Patalano & Wengrovitz, 2007): Is further deliberation likely to resolve preference uncertainty, or is it unresolvable? If the DM is too reluctant to terminate inconclusive deliberation, the benefits of an optimal selection (when it is well-defined) may be outweighed by the costs—measured in time, emotion, and missed opportunities—of extreme indecision (which has been associated with compulsive perfectionism; Frost & Shows, 1993). At some point, the indecisive DM may be well-advised to treat the choice as "for all practical purposes" arbitrary, employing principles of rational arbitrariness to make consistent choices and avoid wasteful plans. But a full treatment of this problem would have to explicitly account for the risk of error and the cost of deliberation, along with the pitfalls of incoherence.

Conclusion

As the preceding discussion makes clear, our rational analysis of incomplete preferences is far from the final word. It is itself incomplete. Open questions remain regarding the shape and texture of vague values, as well as the environments and tasks in which DMs must act on those values. Deeper insights into these empirical questions may require substantial revisions to the rational analysis.

In the psychological literature on decision making, two implicit assumptions, it seems to us, play a pervasive background role in guiding inquiry. First, the normative (*ought*) and the descriptive (*is*) are regarded as independent conceptual realms. Second, it is assumed that, while descriptive theories of human decision making are perennial works in progress, the theory of rational choice is a finished product: At the normative level, expected utility theory, in its standard axiomatization, *is* the final word. Both of these background assumptions, we believe, are mistaken.

Normative and descriptive analyses certainly can, and do, diverge in myriad ways. We do not always do what we should, and our errors can be systematic and consequential. Yet there is a deeper conceptual *inter*dependence between the normative and descriptive levels of analysis. Ultimately, what is rational depends on what is sought and what is known. Normative principles of rational choice thus depend on descriptive assumptions, often implicit, about the structure of human beliefs and values. And because the first background assumption (normative-descriptive independence) is wrong, the second assumption (rational choice theory as a finished product) is likely to be wrong as well. Beliefs and values form part of the subject matter of human psychology. Theories of belief and value structure, like theories in any branch of empirical psychology, are perennially tentative and open to

revision; hence so too is the theory of rational choice. Expected utility theory is not the final word.

We have seen that a simple revision to our descriptive representation of human value—dispensing with the idealization of sharp values that is standard in both classical and behavioral economics—has a cascade of surprising normative implications. By the same token, future developments in the psychological understanding of human value may reinforce or challenge the tentative normative conclusions we have reached here. Normative inquiry, like the empirical inquiry on which it depends, is open-ended.

References

Alxatib, S., & Pelletier, F. J. (2011). The psychology of vagueness: Borderline cases and contradictions. *Mind & Language*, *26*(3), 287–326. https://doi.org/10.1111/j.1468-0017.2011.01419.x

Anand, P. (2009). Rationality and intransitive preference: Foundations for the modern view. In P. Anand, P. Pattanaik, & P. Puppe (Eds.), *The handbook of rational and social choice* (pp. 156–172). Oxford University Press.

Anderson, C. J. (2003). The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Bulletin*, *129*(1), 139–167. https://doi.org/10.1037/0033-2909.129.1.139

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(3), 471–485. https://doi.org/10.1017/S0140525X00070801

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, *118*(1), 73–106. https://doi.org/10.1162/00335530360535153

Ariely, D., Loewenstein, G., & Prelec, D. (2006). Tom Sawyer and the construction of value. *Journal of Economic Behavior & Organization*, *60*(1), 1–10. https://doi.org/10.1016/j.jebo.2004.10.003

Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin*, *125*(5), 591–600. https://doi.org/10.1037/0033-2909.125.5.591

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, *35*(1), 124–140. https://doi.org/10.1016/0749-5978(85)90049-4

Aronson, E. (1999). Dissonance, hypocrisy, and the self-concept. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 103–126). American Psychological Association. https://doi.org/10.1037/10318-005

Aumann, R. (1962). Utility theory without the completeness axiom. *Econometrica*, *30*(3), 445–462. https://doi.org/10.2307/1909888

Bales, A., Cohen, D., & Handfield, T. (2014). Decision theory for agents with incomplete preferences. *Australasian Journal of Philosophy*, *92*(3), 453–470. https://doi.org/10.1080/00048402.2013.843576

Banford, N. D., Knetsch, J. L., & Mauser, G. A. (1980). Feasibility judgements and alternative measures of benefits and costs. *Journal of Business Administration*, *11*, 25–35.

Bar-Hillel, M., & Neter, E. (1996). Why are people reluctant to exchange lottery tickets? *Journal of Personality and Social Psychology*, *70*(1), 17–27. https://doi.org/10.1037/0022-3514.70.1.17

Battalio, R. C., Kagel, J. H., & Jiranyakul, K. (1990). Testing between alternative models of choice under uncertainty: Some initial results. *Journal of Risk and Uncertainty*, *3*(1), 25–50. https://doi.org/10.1007/BF00213259

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Bergman, O., Ellingsen, T., Johannesson, M., & Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters*, *107*(1), 66–68. https://doi.org/10.1016/j.econlet.2009.12.028

Bonini, N., Osherson, D., Viale, R., & Williamson, T. (1999). On the psychology of vague predicates. *Mind & Language*, *14*(4), 377–393. https://doi.org/10.1111/1468-0017.00117

Bratman, M. (1987). *Intention, plans, and practical reason*. Harvard University Press.

Broome, J. (2001). Are intentions reasons? And how should we cope with incommensurable values? In C. Morris & A. Ripstein (Eds.), *Practical rationality and preference: Essays for David Gauthier* (pp. 98–120). Cambridge University Press. https://doi.org/10.1017/CBO9780511570803.006

Bruner, J. (2004). Life as narrative. *Social Research*, *71*(3), 691–710. https://doi.org/10.1353/sor.2004.0045

Buchanan, J. M. (1979). Natural and artifactual man. In J. M. Buchanan (Ed.), *What should economists do?* (pp. 93–112). Liberty Fund.

Butler, D. J., & Loomes, G. C. (2007). Imprecision as an account of the preference reversal phenomenon. *The American Economic Review*, *97*(1), 277–297. https://doi.org/10.1257/aer.97.1.277

Butler, D. J., & Loomes, G. C. (2011). Imprecision as an account of violations of independence and betweenness. *Journal of Economic Behavior & Organization*, *80*(3), 511–522. https://doi.org/10.1016/j.jebo.2011.05.008

Callard, A. (2018). *Aspiration: The agency of becoming*. Oxford University Press. https://doi.org/10.1093/oso/9780190639488.001.0001

Cettolin, E., & Riedl, A. (2019). Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization. *Journal of Economic Theory*, *181*, 547–585. https://doi.org/10.1016/j.jet.2019.03.002

Chang, R. (2015). Value incomparability and incommensurability. In I. Hirose & J. Olson (Eds.), *The Oxford handbook of value theory* (pp. 205–224). Oxford University Press.

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*(2), 57–65. https://doi.org/10.1016/S1364-6613(98)01273-X

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Cubitt, R. P., Navarro-Martinez, D., & Starmer, C. (2015). On preference imprecision. *Journal of Risk and Uncertainty*, *50*(1), 1–34. https://doi.org/10.1007/s11166-015-9207-6

Cubitt, R. P., & Sugden, R. (2001). On money pumps. *Games and Economic Behavior*, *37*(1), 121–160. https://doi.org/10.1006/game.2000.0834

Danan, E. (2010). Randomization vs. selection: How to choose in the absence of preference? *Management Science*, *56*(3), 503–518. https://doi.org/10.1287/mnsc.1090.1116

Davidson, D., McKinsey, J., & Suppes, P. (1955). Outlines of a formal theory of value, I. *Philosophy of Science*, *22*(2), 140–160. https://doi.org/10.1086/287412

Descartes, R. (1985). Discourse on the method of rightly conducting one's reason and seeking truth in the sciences. In J. Cottingham, R. Stoothoff, & D. Murdoch (Trans.), *The philosophical writings of Descartes*, Vol. 1 (pp. 111–151). Cambridge University Press. (Original work published 1637)

Donaldson, D., & Weymark, J. A. (1998). A quasiordering is the intersection of orderings. *Journal of Economic Theory*, *78*(2), 382–387. https://doi.org/10.1006/jeth.1997.2360

Doody, R. (2020). The sunk cost 'fallacy' is not a fallacy. *Ergo*, *6*, 1153–1190. https://doi.org/10.3998/ergo.12405314.0006.040

Douglas, W. O. (1949). Stare decisis. *Columbia Law Review*, *49*(6), 735–758. https://www.scribd.com/document/375986796/W-O-Douglas-Stare-Decisis

Dubourg, W. R., Jones-Lee, M. W., & Loomes, G. (1997). Imprecise preferences and survey design in contingent valuation. *Economica*, *64*(256), 681–702. https://doi.org/10.1111/1468-0335.00106

Eliaz, K., & Ok, E. A. (2006). Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences. *Games and Economic Behavior*, *56*(1), 61–86. https://doi.org/10.1016/j.geb.2005.06.007

Enisman, M., Shpitzer, H., & Kleiman, T. (2021). Choice changes preferences, not merely reflects them: A meta-analysis of the artifact-free free-choice paradigm. *Journal of Personality and Social Psychology*, *120*(1), 16–29. https://doi.org/10.1037/pspa0000263

Erev, I. (2020). Money makes the world go round, and basic research can help. *Judgment and Decision Making*, *15*(3), 304–310. https://doi.org/10.1017/S1930297500007129

Erev, I., & Marx, A. (2023). Humans as intuitive classifiers. *Frontiers in Psychology*, *13*, Article 1041737. https://doi.org/10.3389/fpsyg.2022.1041737

Erev, I., & Plonsky, O. (2022). The J/DM separation paradox and the reliance on small samples hypothesis. In K. Fiedler, P. Justlin, & J. Denrell (Eds.), *Sampling in judgment and decision making* (pp. 115–130). Cambridge University Press.

Ert, E., & Erev, I. (2008). The rejection of attractive gambles, loss aversion, and the lemon avoidance heuristic. *Journal of Economic Psychology*, *29*(5), 715–723. https://doi.org/10.1016/j.joep.2007.06.003

Ert, E., & Erev, I. (2013). On the descriptive value of loss aversion in decisions under risk: Six clarifications. *Judgment and Decision Making*, *8*(3), 214–235. https://doi.org/10.1017/S1930297500005945

Fischhoff, B. (1991). Value elicitation: Is there anything in there? *American Psychologist*, *46*(8), 835–847. https://doi.org/10.1037/0003-066X.46.8.835

Frost, R. O., & Shows, D. L. (1993). The nature and measurement of compulsive indecisiveness. *Behaviour Research and Therapy*, *31*(7), 683–692. https://doi.org/10.1016/0005-7967(93)90121-A

Fudenberg, D., Levine, D. K., & Maniadis, Z. (2012). On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal. Microeconomics*, *4*(2), 131–145. https://doi.org/10.1257/mic.4.2.131

Gal, D. (2006). A psychological law of inertia and the illusion of loss aversion. *Judgment and Decision Making*, *1*(1), 22–32. https://doi.org/10.1017/S1930297500000322

Gal, D., & Rucker, D. D. (2018). The loss of loss aversion: Will it loom larger than its gain? *Journal of Consumer Psychology*, *28*(3), 497–516. https://doi.org/10.1002/jcpy.1047

Garland, H., & Newport, S. (1991). Effects of absolute and relative sunk costs on the decision to persist with a course of action. *Organizational Behavior and Human Decision Processes*, *48*(1), 55–69. https://doi.org/10.1016/0749-5978(91)90005-E

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278. https://doi.org/10.1126/science.aac6076

Gollwitzer, P. M., & Bayer, U. C. (1999). Deliberative versus implemental mindsets in the control of action. In S. Chaiken (Ed.), *Dual-process theories in social psychology* (pp. 403–422). Guilford Press.

Hare, C. (2010). Take the sugar. *Analysis*, *70*(2), 237–247. https://doi.org/10.1093/analys/anp174

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523. https://doi.org/10.1016/j.tics.2009.09.004

Hogarth, R. M., & Einhorn, H. J. (1990). Venture theory: A model of decision weights. *Management Science*, *36*(7), 780–803. https://doi.org/10.1287/mnsc.36.7.780

Hong, F., Huang, W., & Zhao, X. (2019). Sunk cost as a self-management device. *Management Science*, *65*, 2216–2230.

Hsee, C. K., Zhang, J., & Chen, J. (2004). Internal and substantive inconsistencies in decision making. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 360–378). Blackwell. https://doi.org/10.1002/9780470752937.ch18

Huber, V. L., Neale, M. A., & Northcraft, G. B. (1987). Decision bias and personnel selection strategies. *Organizational Behavior and Human Decision Processes*, *40*(1), 136–147. https://doi.org/10.1016/0749-5978(87)90009-4

Johnson, E. J., & Goldstein, D. G. (2013). Decisions by default. In E. Shafir (Ed.), *The behavioral foundations of public policy* (pp. 417–427). Princeton University Press.

Kahneman, D. (2000). Preface. In D. Kahneman & A. Tversky (Eds.), *Choices, values, and frames* (pp. ix–xvii). Cambridge University Press.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, *98*(6), 1325–1348. https://doi.org/10.2307/j.ctv550cbm.30

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Knetsch, J. L. (1989). The endowment effect and evidence of nonreversible indifference curves. *The American Economic Review*, *79*(5), 1277–1284.

Korsgaard, C. M. (2009). *Self-constitution: Agency, identity, and integrity*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199552795.001.0001

Levi, I. (1986). *Hard choices: Decision making under unresolved conflict*. Cambridge University Press. https://doi.org/10.1017/CBO9781139171960

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149–188. https://doi.org/10.1006/obhd.1998.2804

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*, 46–55. https://doi.org/10.1037/h0031207

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, Article e1. https://doi.org/10.1017/S0140525X1900061X

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Louie, K., & De Martino, B. (2014). The neurobiology of context-dependent valuation and choice. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (2nd ed., pp. 455–476). Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00024-3

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, *27*(4), 1622–1668.

Mandler, M. (2004). Status quo maintenance reconsidered: Changing or incomplete preferences? *Economic Journal*, *114*(499), F518–F535. https://doi.org/10.1111/j.1468-0297.2004.00257.x

Mandler, M. (2005). Incomplete preferences and rational intransitivity of choice. *Games and Economic Behavior*, *50*(2), 255–277. https://doi.org/10.1016/j.geb.2004.02.007

Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *The American Economic Review*, *104*(1), 277–290. https://doi.org/10.1257/aer.104.1.277

March, J. G., & Olsen, J. P. (1975). The uncertainty of the past: Organizational learning under ambiguity. *European Journal of Political Research*, *3*(2), 147–171. https://doi.org/10.1111/j.1475-6765.1975.tb00521.x

McClennen, E. (1990). *Rationality and dynamic choice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511983979

McDonald, R. I., Newell, B. R., & Denson, T. F. (2014). Would you rule out going green? The effect of inclusion versus exclusion mindset on

pro-environmental willingness. *European Journal of Social Psychology*, *44*(5), 507–513. https://doi.org/10.1002/ejsp.2040

McKenzie, C. R. M. (2003). Rational models as theories—Not standards—Of behavior. *Trends in Cognitive Sciences*, *7*(9), 403–406. https://doi.org/10.1016/S1364-6613(03)00196-7

McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, *17*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2006.01721.x

Müller-Trede, J., Sher, S., & McKenzie, C. R. M. (2015). Transitivity in context: A rational analysis of intransitive choice and context-sensitive preference. *Decision*, *2*(4), 280–305. https://doi.org/10.1037/dec0000037

Nielsen, K., & Rigotti, L. (2022). *Revealed incomplete preferences*. PsyArXiv. https://doi.org/10.48550/arXiv.2205.08584

Ok, E. A., Ortoleva, P., & Riella, G. (2012). Incomplete preferences under uncertainty: Indecisiveness in beliefs versus tastes. *Econometrica*, *80*(4), 1791–1808. https://doi.org/10.3982/ECTA8040

Oretga y Gasset, J. (1941). *Toward a philosophy of history*. University of Illinois Press.

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *9*(1), 35–58. https://doi.org/10.1016/0010-0277(81)90013-5

Pachur, T., & Scheibehenne, B. (2017). Unpacking buyer-seller differences in valuation from experience: A cognitive modeling approach. *Psychonomic Bulletin & Review*, *24*(6), 1742–1773. https://doi.org/10.3758/s13423-017-1237-4

Patalano, A. L., & Wengrovitz, S. M. (2007). Indecisiveness and response to risk in deciding when to decide. *Journal of Behavioral Decision Making*, *20*(4), 405–424. https://doi.org/10.1002/bdm.560

Pistor, K., & Xu, C. (2003). Incomplete law. *New York University Journal of International Law & Politics*, *35*, 931–1014.

Raz, J. (1986). Value incommensurability: Some preliminaries. *Proceedings of the Aristotelian Society*, *86*(1), 117–134. https://doi.org/10.1093/aristotelian/86.1.117

Ready, R. C., Navrud, S., & Dubourg, W. R. (2001). How do respondents with uncertain willingness to pay answer contingent valuation questions? *Land Economics*, *77*(3), 315–326. https://doi.org/10.2307/3147126

Rey Rosa, R. (2014). *Severina* . (C. Andrews, Trans.). Yale University Press. (Original work published 2011)

Risen, J. L., & Gilovich, T. (2007). Another look at why people are reluctant to exchange lottery tickets. *Journal of Personality and Social Psychology*, *93*(1), 12–22. https://doi.org/10.1037/0022-3514.93.1.12

Ritov, I., & Baron, J. (1992). Status-Quo and omission biases. *Journal of Risk and Uncertainty*, *5*, 49–61. https://link.springer.com/article/10.1007/BF00208786

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Sainsbury, R. M. (1996). Concepts without boundaries. In R. Keefe & P. Smith (Eds.), *Vagueness: A reader* (pp. 251–264). MIT Press.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59. https://doi.org/10.1007/BF00055564

Sartre, J.-P. (2007). *Existentialism is a humanism*. Yale University Press. (Original work published 1946)

Schneider, I. K., & Schwarz, N. (2017). Mixed feelings: The case of ambivalence. *Current Opinion in Behavioral Sciences*, *15*, 39–45. https://doi.org/10.1016/j.cobeha.2017.05.012

Sen, A. (1997). Maximization and the act of choice. *Econometrica*, *65*(4), 745–779. https://doi.org/10.2307/2171939

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition*, *101*(3), 467–494. https://doi.org/10.1016/j.cognition.2005.11.001

Sher, S., & McKenzie, C. R. M. (2014). Options as information: Rational reversals of evaluation and preference. *Journal of Experimental Psychology: General*, *143*(3), 1127–1143. https://doi.org/10.1037/a0035128

Sher, S., & McKenzie, C. R. M. (2022). Incomplete preferences and rational framing effects. *Behavioral and Brain Sciences*, *45*, Article e240. https://doi.org/10.1017/S0140525X2200111X

Sher, S., McKenzie, C. R. M., Müller-Trede, J., & Leong, L. (2022). Rational choice in context. *Current Directions in Psychological Science*, *31*(6), 518–525. https://doi.org/10.1177/09637214221120387

Sher, S., Müller-Trede, J., & McKenzie, C. R. M. (2024). *Vagueness, value, and constructive cognition* [Unpublished manuscript]. Department of Psychological Science, Pomona College.

Shils, E. (1981). *Tradition*. University of Chicago Press.

Shin, Y. S., & Niv, Y. (2021). Biased evaluations emerge from inferring hidden causes. *Nature Human Behaviour*, *5*(9), 1180–1189. https://doi.org/10.1038/s41562-021-01065-0

Smitizsky, G., Liu, W., & Gneezy, U. (2021). The endowment effect: Loss aversion or a buy-sell discrepancy? *Journal of Experimental Psychology: General*, *150*(9), 1890–1900. https://doi.org/10.1037/xge0000880

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, *27*(1), 76–105. https://doi.org/10.1016/0022-1031(91)90011-T

Stone, J., & Cooper, J. (2001). A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology*, *37*(3), 228–243. https://doi.org/10.1006/jesp.2000.1446

Sugden, R., Zheng, J., & Zizzo, D. J. (2013). Not all anchors are created equal. *Journal of Economic Psychology*, *39*, 21–31. https://doi.org/10.1016/j.joep.2013.06.008

Sunstein, C. R. (1994). Incommensurability and valuation in law. *Michigan Law Review*, *92*(4), 779–861. https://doi.org/10.2307/1289693

Temkin, L. S. (2012). *Rethinking the good: Moral ideals and the nature of practical reasoning*. Oxford University Press.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318. https://doi.org/10.1016/j.tics.2006.05.009

Teodorescu, K., & Erev, I. (2014). On the decision to explore new alternatives: The coexistence of under-and over-exploration. *Journal of Behavioral Decision Making*, *27*(2), 109–123. https://doi.org/10.1002/bdm.1785

Tversky, A. (1996). Rational theory and constructive choice. In K. J. Arrow, E. Colombatto, M. Perlman, & C. Schmidt (Eds.), *The rational foundations of economic behavior* (pp. 185–197). Macmillan.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323. https://doi.org/10.1007/BF00122574

Unkelbach, C., Koch, A., & Alves, H. (2021). Explaining negativity dominance without processing bias. *Trends in Cognitive Sciences*, *25*(6), 429–430. https://doi.org/10.1016/j.tics.2021.04.005

van de Ven, N., & Zeelenberg, M. (2011). Regret aversion and the reluctance to exchange lottery tickets. *Journal of Economic Psychology*, *32*(1), 194–200. https://doi.org/10.1016/j.joep.2010.11.008

Vlaev, I., Seymour, B., Dolan, R. J., & Chater, N. (2009). The price of pain and the value of suffering. *Psychological Science*, *20*(3), 309–317. https://doi.org/10.1111/j.1467-9280.2009.02304.x

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Walasek, L., & Brown, G. D. (2023). Incomparability and incommensurability in choice: No common currency of value? *Perspectives on Psychological Science*, *19*(6), 1011–1030. https://doi.org/10.1177/17456916231192828

Weber, B. J., & Chapman, G. B. (2005). Playing for peanuts: Why is risk seeking more common for low-stakes gambles? *Organizational Behavior and Human Decision Processes*, *97*(1), 31–46. https://doi.org/10.1016/j.obhdp.2005.03.001

Yaniv, I., & Schul, Y. (1997). Elimination and inclusion procedures in judgment. *Journal of Behavioral Decision Making*, *10*(3), 211–220. https://doi.org/10.1002/(SICI)1099-0771(199709)10:3<211::AID-BDM250>3.0.CO;2-J

Yaniv, I., & Schul, Y. (2000). Acceptance and elimination procedures in choice: Noncomplementarity and the role of implied status quo. *Organizational Behavior and Human Decision Processes*, *82*(2), 293–313. https://doi.org/10.1006/obhd.2000.2899

Yechiam, E., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin*, *139*(2), 497–518. https://doi.org/10.1037/a0029383

Yoon, S., Fong, N. M., & Dimoka, A. (2019). The robustness of anchoring effects on preferential judgments. *Judgment and Decision Making*, *14*(4), 470–487. https://doi.org/10.1017/S1930297500006148

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zhao, W. J., Walasek, L., & Bhatia, S. (2020). Psychological mechanisms of loss aversion: A drift-diffusion decomposition. *Cognitive Psychology*, *123*, Article 101331. https://doi.org/10.1016/j.cogpsych.2020.101331

Received September 5, 2023
Revision received October 16, 2024
Accepted November 4, 2024 ■