Constructed Preferences, Rationality, and Choice Architecture

Craig R. M. McKenzie
UC San Diego

Shlomi Sher
Pomona College

Lim M. Leong
UC San Diego

Johannes Müller-Trede
IESE Business School

Draft of July 3, 2018

In press, *Review of Behavioral Economics*

Abstract

Preferences must be constructed at least some of the time. This, by itself, is not problematic for rationality. At issue is whether the construction is done in a reasonable manner. The common view is that preference construction violates coherence principles that are basic requirements of rational choice. However, traditional coherence principles are static and implictly assume that the choice context provides no relevant information. In lab experiments, decision makers often evaluate or choose between options that are unfamiliar or even fictitious, and they may look to the context for choice-relevant cues that help them update their beliefs and construct their preferences. We review evidence that a number of apparent "biases" in decision making stem from adaptive sensitivity to subtle contextual cues. These context effects are dynamically coherent, in that preference-updating is coordinated with reasonable context-dependent belief-updating. This perspective on preference construction not only provides a different view of the psychology and rationality of decision making, it also suggests a different approach to choice architecture. Whereas the traditional nudge approach tries to engineer specific decision *outcomes*, often by rerouting apparent biases so that they point in desirable directions, the present approach seeks to facilitate *processes* in order to help people make rational decisions.

Constructed Preferences, Rationality, and Choice Architecture

Are people poor decision makers (DMs) who need help with satisfying their "true preferences"? Many behavioral decision researchers have argued for decades that decision making behavior is systematically biased (Gilovich, Griffin, & Kahneman, 2002; Kahneman & Tversky, 2000; Shafir & LeBoeuf, 2002) and "predictably irrational" (Ariely, 2009). These claims have motivated recent proposals that governments and businesses should "nudge" individuals to help them make better choices – choices that the individuals themselves would make if they were not systematically biased (Thaler & Sunstein, 2008). The leading proponents of nudges advocate a program of "libertarian paternalism." The idea is that "choice architects" should design choice environments that render certain options (those identified by the choice architect as best for the DM) more likely to be chosen, while at the same time making it easy for the DM to choose differently if desired. Not surprisingly, many critics of "classic" paternalism have also criticized libertarian paternalism (e.g., Glaeser, 2006).

In this article, we outline a different perspective on the psychology and rationality of decision making, and we consider its implications for choice architecture. This new perspective highlights the subtle but choice-relevant cues that are contained in many decision environments, and recasts several apparent biases as products of adaptive cue-sensitivity. It also suggests guidelines for non-paternalistic forms of choice architecture.

To introduce this alternative view, we begin where Thaler and Sunstein began *Nudge*. They opened their first chapter ("Biases and Blunders") with Shepard's (1990) famous table illusion (Figure 1). The reader is invited to "suppose that you are thinking about which [table] would work better as a coffee table in your living room," and to estimate the dimensions of the tables (p. 17). People reliably perceive the left table top to be longer and thinner than the one on

the right, even though they are the same size on the page. Thaler and Sunstein conclude that "your judgment in this task was biased, and predictably so. [...] Not only were you wrong; you were probably confident that you were right" (p. 18). The table illusion is said to "capture the key insight that behavioral economists have borrowed from psychologists": Despite the mind's impressive abilities, we can be easily fooled in simple situations, and thus "our understanding of human behavior can be improved by appreciating how people systematically go wrong" (p. 19).

We agree with Thaler and Sunstein that the table illusion captures a psychological insight that has important applications to behavioral economics. But we believe that the illusion's deep lesson is quite different from the one that they draw.

What is left unsaid in *Nudge* is *why* people perceive the left table as longer and thinner. The explanation is simple and instructive. The visual system's central challenge is to solve the "inverse problem": Given the limited information in the (2D) retinal array, what configuration of objects in the (3D) world is most likely to have caused it? Because the purpose of the visual system is to help us navigate in the 3D world (and not to show us what our retinas look like), it is the likeliest 3D interpretation of the retinal input that we should see. And because retinal inputs are ambiguous (e.g., a long line far away and a short line nearby cast the same projection on the retina), the visual system can only make a "best guess" about the configuration of objects in the 3D world. Importantly, the visual system draws on contextual cues within the image to construct this guess. In Figure 1, multiple cues conspire to suggest that the long edge of the left table is receding into the distance, and hence must be longer. That is, in a normal 3D world, the table projecting the image on the left very likely *would* be longer and thinner than the table projecting the image on the right. And since it's probably the 3D table, not the 2D image of the table, that you'll be fitting into your living room, that's what you *should* see.

When you consider *why* it arises, it becomes clear that the table illusion hardly illustrates shortcomings of our cognitive system. Instead, it showcases its sophistication and adaptiveness. The fact that we perceive a 3D world accurately (and rapidly, and without conscious reasoning) when our retinas only receive ambiguous 2D information is nothing short of amazing. The (2D) illusion helps us understand how the system gets things right, not wrong, in the real (3D) world.

Interestingly, Fine et al. (2003) describe a patient (MM) who perceives the table tops in Figure 1 to be the same size. MM was blinded in an accident at age 3, and regained vision forty years later through surgery. His long-term visual deprivation makes it difficult for MM to interpret retinal images in 3D. The fact that MM does not fall prey to the table illusion is indicative of an impaired visual system, not an ideal one.

In this article, we argue that interpretations of people's decision making as biased and irrational are sometimes misguided in ways that are analogous to the interpretation of the table illusion as indicating a shortcoming of the visual system. In these cases, what appear to be violations of rational choice theory turn out, upon closer inspection, to reflect adaptive responses to relevant information in the choice context.

Preference Construction in Context

According to the standard behavioral view of decision making that motivates nudges and libertarian paternalism, preferences are actively constructed in the act of choice, and constructed preferences are systematically incoherent. While these two claims -- constructed preference on the one hand, and incoherence on the other -- are often conflated, it is important to distinguish them. First, there is no normative requirement that a person have predetermined utility for every possible object, event, or attribute. Second, preferences could be constructed online in a manner

that is consistent, unbiased, and satisfies relevant coherence principles. Preference construction need not imply incoherence or irrationality.

In practice, however, the evidence indicating that preferences are constructed comes from violations of coherence principles (e.g., Slovic, 1995). Framing effects, for instance, violate the principle of *description invariance*, according to which logically equivalent descriptions should lead to identical decisions (Levin, Schneider, & Gaeth, 1998). Similarly, default effects (Johnson & Goldstein, 2003) and other apparent preference reversals (Hsee, Zhang, & Chen, 2004) violate *procedure invariance*, according to which different procedures for eliciting preferences should yield the same ordering of options. And the attraction and compromise effects (Huber, Payne, & Puto, 1982; Simonson, 1989) violate the principle of *independence of irrelevant alternatives*, according to which the ranking of any two options should not be affected by the presence or absence of a third. The pervasive context-dependence of human decision making is most naturally explained by a constructive model, and it also runs afoul of widely accepted principles of rational choice (Tversky, 1996).

These coherence principles, however, implicitly assume that the choice context provides no choice-relevant information, whereas the context can in fact provide information that would be irrational to ignore. Participants in decision making experiments are often asked to evaluate or choose between options that they are unfamiliar with (and that are sometimes defined in terms of attributes invented by the experimenter). When their knowledge is incomplete, participants may look to the task "context" for relevant information, and update their beliefs accordingly. These revised beliefs may in turn translate into revised preferences. In the next section, we argue that, in the natural ecology of decision making, various features of the choice context – including the frame, the default, and the menu of available options – are typically cues to choice-relevant

features of the larger environment in which the choice problem arises. As a result, rational decision making has much in common with optimal perception. The subjective value of an option, like the perception of the length of a line, *should* be sensitive to subtle contextual cues that are logically ambiguous but probabilistically informative.

If patterns of context-dependent preference that have been taken as evidence for irrationality are effects of adaptive cue-sensitivity, what are the implications for "nudging" DMs? Consider again the table illusion. Without understanding why the illusion occurs, we might be tempted to "nudge" the visual system to avoid seeing it. Such an intervention would, of course, be misguided. (One is reminded of the parable in which a monkey tries to "save" a fish from "drowning", but by doing so ends up killing it.)

But choice architecture is important, even in cases where human decision making is approximately rational. The cues embedded in a choice problem, like the cues in an image, can be valid, but they can also be misleading. Furthermore, "inattentional blindness" (Mack & Rock, 1998) is as relevant to decision making as it is to visual perception – in both domains, we are largely blind to what we do not attend to. Thus, even a choice architect who rejects paternalism, and who respects the DM's capacity for rational choice, will think carefully about precisely which information to convey and how, and about the intended and unintended contextual inferences the DM is likely to draw. We explore these implications in more detail later in the article. First, we review empirical evidence for adaptive cue-sensitivity in preference construction.

Cues in the Choice Context

We have highlighted parallels between percepts and preferences. Both perception and preference are fundamentally constructive, and in both cases the constructive process is highly sensitive to subtle contextual features. In neither case, however, does the constructive nature of the process imply incoherence, systematic bias, or suboptimal performance. On the contrary, the working assumption that constructive processes are approximately Bayes-optimal has often proved productive in studies of visual perception (e.g., Kersten, Mamassian, & Yuille, 2004; Knill & Richards, 1996). In this section, we argue that a kindred analysis sheds valuable light on some familiar "biases" in decision making that are widely viewed as "blunders."

In the table illusion (Figure 1), the legs and orientation of the tables may seem to be mere "distracters" that prevent the visual system from functioning well (Thaler & Sunstein, 2008, p. 18). For the vision scientist, they are instead regarded as *cues*, within the image, that provide information relevant to the likeliest 3D interpretation of other aspects of the 2D image (e.g., the areas of the table tops). The decision scientist can ask a similar question: What cues are available within typical "2D" descriptions of choice problems, and what information might they convey about the larger "3D" world in which the choice problem is likely to have arisen? In this section, we focus on three contextual cues that are often available in the choice problems that people face – the frame, the default, and the menu of options. These three cues are normally correlated with relevant features of the larger environment, including the distribution of options in the choice space and the opinions of the speaker or the choice architect.

When contexts are informative, the relevant test of coherence is *dynamic* rather than *static* (Sher & McKenzie, 2014). The critical question then is not whether choices are invariant to frame, procedure, and menu (the static test), but whether preference-updating is coherently

coordinated with belief-updating across contexts (the dynamic test). That is, do the effects of context on choice match the combination of (1) the effects of context on beliefs and (2) the effects of beliefs on preferences? In what follows, we show that some well-known effects of frame, default, and menu – which violate the traditional coherence principles of invariance and independence – satisfy the more nuanced, and in our view more appropriate, test of dynamic coherence.

Cue 1: The Frame

Framing effects are said to occur when redescribing objects or outcomes in equivalent ways affects behavior. Such effects violate description invariance, a coherence principle that is widely regarded as "normatively unassailable" (Tversky, Slovic, & Kahneman, 1990, p. 214). Based in part on framing effects in risky choice, Tversky and Kahneman (1986) claimed that "no theory of choice can be both normatively adequate and descriptively accurate" (p. S251). Because framing effects are empirically robust, any descriptive theory must allow them, but if framing effects are normatively unacceptable, no rational theory could allow them. Hence, there can be descriptive theories, and there can be rational theories, but never the twain shall meet.

There are different types of framing effects. One type, known as an "attribute framing effect", occurs when logically equivalent redescriptions of an object or event along a single dimension affect behavior (Levin, Schneider, & Gaeth, 1998). One frame is usually positive and one is negative, and people provide more favorable ratings when presented with the positive frame. A medical treatment, for instance, is viewed more positively when it is described as leading to a "90% survival rate" rather than a "10% mortality rate". Attribute framing effects have traditionally been explained by an associative account, in which positive (negative) frames

evoke positive (negative) associations, which in turn color the event or object being framed (Levin, 1987; Levin et al. 1998).

It is now well-established, however, that a speaker's *choice of frame* can provide a choice-relevant signal (e.g., Sher & McKenzie, 2006). Speakers do not select frames at random, but tend to describe something in terms of how X it is if they perceive it as relatively X (i.e., as having a higher level of X than it previously did or typically would have). For example, a 4 oz. container with liquid at the 2 oz. line is more likely to be described as "half empty" if it used to be full than if it used to be empty (McKenzie & Nelson, 2003). Similarly, a new medical treatment that leads to a higher survival rate than the traditional treatment is more likely to be described in terms of its survival rather than mortality rate. In this case, a speaker's choice of frame provides information regarding the efficacy of the treatment relative to relevant alternatives.

The fact that a speaker's choice of frame is a choice-relevant cue helps explain why attribute framing effects occur -- and suggests that at least some framing effects are not irrational (for a rational account of risky choice framing, see Mandel, 2014). "Listeners", who receive frames, are justified in responding differently to different frames when those frames "leak" relevant information. Indeed, there is evidence that listeners often draw inferences from a speaker's choice of frame: a medical treatment is believed to be relatively efficacious when its outcome is framed in terms of survival rather than mortality rate (McKenzie & Nelson, 2003), a project is behind schedule when framed in terms of time spent rather than time left (Teigen & Karevold, 2005), and a basketball player is above average when his performance is described in terms of the percentage of shots made rather than shots missed (Leong et al., 2017). Recently, Leong et al. (2017) showed that listener inferences are sufficient to generate attribute framing

effects, and that blocking these inferences attenuates the effects. Inferences are thus not mere epiphenomena, but play a causal role in attribute framing effects.

In summary, attribute frames are informative cues, and their typical behavioral effects are dynamically coherent. It is not in general irrational to respond differently to different frames. Indeed, to the extent that a speaker's choice of frame leaks choice-relevant information, it would be irrational *not* to respond differently.

Cue 2: The Default

In some situations, a particular option must be imposed as a default if people fail to make a choice. People in the US, for example, are not considered organ donors unless they actively choose to be, or opt in. In some other countries, being an organ donor is the default and people must actively opt out if they wish not to be donors. Johnson and Goldstein (2003) showed that countries with "organ donor" as the default had much higher effective consent rates than countries with "not an organ donor" as the default. Default effects have been documented in many choice domains. For instance, employees are much more likely to participate in a retirement plan when participation is the default (Madrian & Shea, 2001).

Why are people so strongly influenced by defaults? People are undoubtedly affected by the cost of opting in or out. But that seems unlikely to be the whole story, since the costs are typically small (and sometimes trivial), and default effects are found even when holding switching costs constant (Johnson & Goldstein, 2003). While a number of factors may contribute to default effects, there is evidence that DMs perceive a policymaker's *choice of default* to be informative (McKenzie et al., 2006; see also Tannenbaum, Valasek, Knowles, & Ditto, 2013). In particular, DMs perceive defaults as *implicit recommendations*. In one study, some participants were told that policymakers had decided to make "organ donor" the default,

while others were told that policymakers had decided to make "not an organ donor" the default. Participants in the "organ donor" condition were more likely to infer that the policymakers thought it was a good idea to be an organ donor and that the policymakers were organ donors themselves (McKenzie et al., 2006).

Why do people treat defaults as implicit recommendations? In another experiment, McKenzie et al. (2006) first asked participants if they were willing to be organ donors and whether they thought people ought to be organ donors (answering both questions with yes, no, or unsure). They were then told to imagine that they would be making the organ donor policy for their state and would have to decide on a default. Participants were much more likely to select the policy with the organ donor default if they were willing to be an organ donor or thought that others ought to be organ donors. In other words, the "policymakers'" preferences were leaked by their choice of default; their defaults were, in effect, signaling a recommendation.

In short, much as a speaker's choice of frame supplies a relevant cue to the DM, so too can a policymaker's choice of default. In both cases, the cue's effect on preferences matches its effect on choice-relevant beliefs.

Cue 3: The Menu

A common observation in the decision making literature is that evaluations of single options, or choices between pairs of options, are influenced by other alternatives in the choice menu. These findings are widely regarded as counter-normative because it is assumed that the rational DM must have a pre-existing preference ranking, which can be probed, but should not be influenced, by the menu of options. However, when the DM's knowledge of the choice space is imperfect (i.e., in most cases), options do not merely supply possible avenues of action. They also provide potentially relevant information about the larger space of options from which they

were sampled – and DMs are highly sensitive to this information. As we show next, inferences from options can generate a broad range of apparent coherence violations, including joint-separate reversals, attraction and compromise effects, and intransitive choice cycles.

Joint-separate reversals. In forming evaluations and making decisions, we sometimes encounter a single option in isolation, while on other occasions we confront two or more options of the same type. Hsee and colleagues have argued that people process attribute information differently in these two settings, leading to systematic preference reversals between joint and separate evaluation (Hsee, Loewenstein, Blount, & Bazerman, 1999; Hsee & Zhang, 2010). In particular, attributes that are unfamiliar and "difficult-to-evaluate" are said to receive greater weight in joint evaluation. For example, participants in one study (Hsee, 1996) evaluated one or two candidates for a programming position involving a fictitious programming language (for similar stimuli, see Table 1). One candidate had a higher GPA (a highly familiar attribute), and received a higher mean salary when evaluated in isolation. The other candidate had written a greater number of programs in the language (a highly relevant but also highly unfamiliar attribute), and received a higher salary when the candidates were evaluated jointly. These "preference reversals" are seen as counter-normative violations of procedure invariance.

Sher and McKenzie (2014) argued that joint-separate "preference reversals" are not in fact preference reversals, nor are they counter-normative. Instead, they are parsimoniously explained by the reasonable inferences that participants draw from different evaluation sets. In the joint-separate paradigm, one attribute is unfamiliar by design. The typical subject will have little, if any, prior knowledge of this attribute's distribution in the natural environment. As a consequence, the evaluation set likely does double duty, serving both as the menu of options and as a source of information – a sample drawn from a poorly known attribute distribution.

Rational choice will then proceed via a two-stage process of belief- and preference-updating: First, the DM will update her internal model of the attribute distribution on the basis of the new information contained in the sample. Second, the DM will evaluate the option(s) in light of this updated model. This two-stage process can readily generate joint-separate reversals when a single attribute is poorly known.

For example, Sher and McKenzie found that, in the programmer problem in Table 1, undergraduate participants who received different evaluation sets (the high-GPA candidate alone, the experienced programmer alone, or both candidates together) drew markedly different inferences about the distribution (estimated low, average, and high values) of the unfamiliar attribute (programming experience) in the pool of candidates competing for the job. Reasonably enough, estimated population parameters were drawn towards the statistics of the particular option-sample the participant had seen (Figure 2, left panel). To examine the effect of these distributional inferences on evaluations, the estimates were later provided as background information to a group of yoked participants, all of whom evaluated a *single* candidate. The resulting pattern of evaluations (Figure 2, right panel) reproduced the joint-separate reversal (middle panel). That is, the inferences that were drawn from different evaluation sets had the same effect as the evaluation sets themselves. Joint-separate reversals thus appear to reflect different inferences from different option-samples, rather than inconsistent attribute weighting in joint vs. separate evaluation. They illustrate the close link between belief-updating and preference construction when prior knowledge is limited and the sample of options is informative.

Effects of "irrelevant" alternatives. Two well-known examples of how menu composition can trigger violations of classic coherence principles are the attraction and

compromise effects. Both paradigms contrast choices from a "core" choice menu, with two consumer products that trade off on two attribute dimensions, with choices from an "expanded" menu that includes the two core products along with a third option. In the attraction effect, the addition of a decoy that is clearly inferior to one of the core products but not the other increases the choice share of that core product which dominates it. In the compromise effect, a core product is more likely to be chosen when the addition of a more extreme third option makes it intermediate on both attribute dimensions. Both effects violate the principle of independence of irrelevant alternatives, and are commonly regarded as counter-normative.

Early research on the attraction effect indicated that it can be moderated by stimulus meaningfulness and product familiarity (Ratneshwar, Shocker, & Stewart, 1987). This observation suggested that contextual inferences could play an important role in explaining the effect. Wernerfelt (1995) subsequently developed an inferential model of both attraction and compromise effects. In this model, rational DMs are aware of how their tastes compare to other consumers' tastes, but they are unsure about the distribution of products that are available in the market. These DMs attempt to infer their correct choice from the offerings on the menu, based on the assumption that the offerings reflect the underlying market distribution. This inferential process can give rise to rational attraction and compromise effects.

Inspired by Wernerfelt's (1995) model, Prelec, Wernerfelt, and Zettelmeyer (1997) report a study that attempts to quantify the inferential component of attraction and compromise effects. For each of several product categories, they first asked participants to assess their tastes relative to other consumers. Afterwards, they presented participants with a set of products for each category, manipulating the set composition between participants. Each participant saw a 2-product "core" set for some product categories, and a 3-product "expanded" set for others (where

some expanded sets included an attraction-type decoy while others featured a compromise-type flanker). Participants did not yet make choices at this stage; instead they were asked to estimate the quality of each product in the set, relative to all products in its category available in the market. As predicted, participants' relative quality estimates were sensitive to the composition of the product set. Moreover, combining these relative quality estimates with participants' assessments of their own relative tastes allowed Prelec et al. (1997) to compute the size of a predicted, rational effect of the third alternative for each product category. They then compared these predicted effects with the empirical menu effects in actual choices, which they elicited one week later in a second stage of the study. While the empirical effects were somewhat larger than those predicted by the rational model, Prelec et al. (1997) concluded that inferences largely accounted for the compromise and attraction effects they observed. Thus, even "irrelevant" (i.e., unchosen) alternatives in the choice menu may serve as relevant cues. Inferences from these cues appear to be major contributors to standard attraction and compromise effects.

Intransitive choice cycles. Perhaps the best-known coherence violation in the behavioral literature is intransitivity of preference. The transitivity axiom – a core assumption of most models of rational choice – states that, for any alternatives *a*, *b*, *c*, if *a* is preferred to *b* and *b* is preferred to *c*, then *a* must be preferred to *c*. Tversky (1969), as well as others building on his work, reported data that appear to violate transitivity. Empirical tests of transitivity are far from straightforward, however, since the axiom applies to preferences, but researchers can only observe choices. Because choice behavior is stochastic – the same person may prefer *a* over *b* at one time and *b* over *a* at another – inference from choice to preference inevitably requires ancillary assumptions, which are often subtle and contentious (see, e.g., Regenwetter, Dana, & Davis-Stober, 2010).

The coherence principle derived from the transitivity axiom with the most traction in recent empirical work is a set of conditions known as the *triangle inequalities*. Letting P_{xy} denote the probability of selecting *x* over *y* in pairwise choice, these inequalities state that, for any triple *a*, *b*, *c*, $P_{ab} + P_{bc} - P_{ac} \leq 1$. The triangle inequalities are more appealing than other operationalizations of transitivity (including the principle of "weak stochastic transitivity" that Tversky tested) because they allow a DM's preferences to independently vary across time. Recent advances in order-constrained statistical inference have enabled decision researchers to test them empirically, leading to a reassessment of evidence for intransitive preference. Reviewing the literature to date, Regenwetter, Dana, and Davis-Stober (2011) concluded that no clear violations of the triangle inequalities had been demonstrated.

But are the triangle inequalities a necessary condition for transitive preference? Müller-Trede, Sher, and McKenzie (2015) argued that, from a normative perspective, the triangle inequalities can be too restrictive – because they assume, unreasonably, that preference variability must be independent of the choice menu and other characteristics of the context. When the DM's prior knowledge is limited, different pairwise choice contexts may lead to different inferences, and these different inferences may in some cases *systematically* induce different transitive preference orders. Müller-Trede et al. (2015) further argued that, in some choice problems (May, 1954; Kivetz & Simonson, 2000), reasonable inferences from pairwise contexts can generate strongly intransitive choice patterns that violate the triangle inequalities, provided that memory for previously sampled contexts is limited.

Consider the master choice set in Table 2, for example, with three sound systems defined on three fictitious attributes (Harmonic Range, Sound Depth, and Acoustic Power) for which, by design, no prior distributional knowledge is available. Confronting a pair of options drawn from

this set, the best the DM can do is to draw inferences about the likely distribution of the attribute from the particular values that are sampled. Figure 3 illustrates inferences that might be drawn upon receipt of the sample {*A*, *B*}. In a normative analysis, Müller-Trede et al. (2015) showed that, if similar weights are placed on the three attributes, sampling the choice set {*A*, *B*} may trigger reasonable inferences that induce the transitive preference order $C \succ A \succ B$; sampling {*B*, *C*} triggers inferences inducing the order $A \succ B \succ C$; while sampling {*A*, *C*} triggers inferences inducing $B \succ C \succ A$. If memory for past options is limited, these normative responses will generate robustly intransitive cyclical choices, with *A* chosen {*A*, *B*}, *B* from {*B*, *C*}, and *C* from {*A*, *C*}.

In two repeated-choice experiments, triangle inequality violations predicted by the normative analysis were clearly demonstrated in numerous participants (Müller-Trede et al., 2015). One third of participants exhibited individually significant triangle inequality violations in pairwise choices drawn from Table 2, and more than half of participants exhibited individually significant violations predicted by the normative model in a second choice task in which some attribute values were missing. This study turns the usual view of the relation between transitivity and rationality on its head. Traditionally, *intransitive choices* are seen as evidence of *intransitive preference*, and hence of *irrationality*. Here, a *rational* analysis of context-dependent *transitive preference* guides the search for empirical conditions under which strongly *intransitive choices* are most likely to found. Transitivity is not a structural requirement on inferences from pairwise choice contexts: One may rationally infer that *a* is better than *b* from one pairwise sample, that *b* is better than *c* from a second, and that *c* is better than *a* from a third. Accordingly, transitivity is not a structural requirement on the choices informed by such contextual inferences.

Summary

While preference construction is often contrasted with rational choice, there is nothing irrational about preference construction *per se*. When prior knowledge is incomplete, frames, defaults, and sampled options may supply choice-relevant information that *should* impact beliefs and preferences. The normative question is thus not whether, but how, preferences are constructed. Is belief-updating appropriately sensitive to implicit cues in the choice environment? The evidence reviewed in this section illustrates the surprising descriptive power of a normative analysis of preference construction. In typical experiments, rational DMs with incomplete knowledge would be expected to exhibit attribute framing effects, default effects, joint-separate reversals, attraction and compromise effects, and (if memory for past contexts is limited) intransitive choices. Like striking illusions in visual perception, these apparent "anomalies" in decision making illustrate not malfunctions of the cognitive system, but its adaptive sensitivity to ecologically relevant cues.

It is important to note that the normative framework outlined here offers a parsimonious account of multiple "biases" that otherwise appear unrelated. It also aligns with other rational analyses of broad classes of behavior -- beyond visual perception -- such as causal inference, concept learning, language acquisition, memory retrieval, problem solving, and reasoning (e.g., Anderson, 1991; Tenenbaum, Kemp, & Griffiths, 2011; see also Chater & Oaksford, 1999; Oaksford & Chater, 2007; Tenenbaum, Griffiths, & Kemp, 2006; Tenenbaum & Griffiths, 2001). Together, these approaches hold out promise for a perspective on decision making that is better-integrated on three critical levels: in developing a more unified account of diverse context effects; in identifying shared principles across different cognitive domains; and in substantively accounting for the successes, not just the failures, of human judgment and action.

Implications for Choice Architecture

Before we can improve people's decision making behavior, we must first understand it. Just as visual illusions do not necessarily reveal shortcomings in the visual system, coherence violations do not necessarily reveal shortcomings in decision making (cf. Arkes, Gigerenzer, & Hertwig, 2016). If prescriptions are based on misguided interpretations of rational behavior as incoherent, they may have the potential to do more harm than good. For example, if frames and defaults are signals, they may not have the intended effect, or even backfire, if DMs are skeptical about the motives of the source of the frame or default (Altmann et al., 2018; Arad & Rubenstein, 2017; Keren, 2007). More generally, different models of preference construction will have different implications both for what aspects of decision making are in need of improvement, and for how improvements can be most effectively brought about.

In light of its descriptive utility in accounting for behavioral "anomalies," a rational analysis of preference construction may also have prescriptive utility for choice architecture. Accordingly, in the remainder of this article, we pose, and sketch a partial answer to, a novel question: What choice architectures *would* be optimal if preference construction were assumed to be rational? Though we do not claim that preference construction is rational in all respects, and it is always subject to basic structural limitations of information processing capacity, the rational model's ability to (at least qualitatively) explain robust effects of frame, default, and choice menu suggest that its implications for choice architecture merit careful investigation. Moreover, a choice architecture strategy that expressly seeks to identify and build on rational aspects of human decision making has an important virtue: In deliberately capitalizing on the DM's potential for rationality, recommendations derived from this strategy are more likely to respect her dignity as an agent – and thus, in the long term, to preserve her trust.

We begin with a brief discussion of pre-choice interventions, which can circumvent the drawbacks of over-reliance on contextual cues. Next, we turn to interventions at the choice point that strive to maximize cue accuracy, relevance, and accessibility. Finally, we contrast two broad agendas for choice architecture – process *versus* outcome facilitation. We compare the behavioral interventions these approaches recommend as well as the ethical questions they raise.

Pre-Choice Interventions

As we have seen, a number of apparent coherence violations can be parsimoniously explained by DMs' inferences from features of the choice environment. In these cases, DMs exploit contextual cues to learn about options that they are either unfamiliar with or for which they lack clear preferences. Is the treatment that is framed as having an "80% success rate" a relatively good one? DMs are not informed explicitly about competing options in attribute framing tasks, but draw inferences about relative efficacy based on the frame. In the case of default effects, DMs may simply be uncertain about their preferences. McKenzie et al. (2006) found that almost one-quarter of undergraduate participants were unsure whether they would enroll in a retirement plan when they got a job after graduation, and one-third were unsure if they were willing to be organ donors. DMs may then look to the default for an implicit recommendation. For some effects of choice menu, including joint-separate reversals, it is crucial that one or more attributes are unfamiliar to DMs, who then update their beliefs about attribute distributions based on the presented option(s).

While contextual cues can be useful in filling out the DM's incomplete model of the choice problem, it is important to note their shortcomings. First, though the information such cues supply is relevant, it is also limited in scope, sketching only a fragmentary picture of the choice space. Second, the information such cues do supply is probabilistic, and hence will be

misleading some of the time. Accordingly, the quality of decisions can often be improved by systematically learning about options *before* being presented with the choice, thus reducing reliance on contextual cues. Such pre-choice education will be desirable when the risks of error outweigh the costs of learning. If one has already reached the conclusion that saving for retirement is important before filling out the paperwork for the new job, then it will not matter much whether plan participation is the default (see Löfgren et al., 2012). If one knows that the standard medical treatment leads to an 85% survival rate, then it will not matter much whether the new treatment is framed as leading to an 80% survival rate or a 20% mortality rate (see Leong et al., 2017). Encouraging, or even helping, DMs to be more informed prior to the choice point would maximize DMs' autonomy and reduce the need for, and the effects of, nudges.

Assistive Cueing

Nevertheless, there will be many situations in which a DM facing a choice lacks clear preferences or sufficient knowledge about options and attributes, and looks to the context for relevant cues. A benevolent choice architect who ascribes rationality to the DM then has a straightforward mission: to craft a choice environment that facilitates accurate choice-relevant inferences. By providing inputs that are as clear, accurate, relevant, and useable as possible, such "assistive cueing" strives to capitalize on the DM's potential for rationality. This objective can be achieved in at least two ways – by constructing representative choice environments (e.g., contexts that preserve the statistical structure of the relevant natural environment) and by substituting accurate and precise explicit information for potentially misleading or ambiguous implicit information. We consider these two strategies in turn.

Even optimal inferences from the choice context will be systematically inaccurate if, unbeknownst to the DM, that context is systematically unrepresentative. For example, consider

judgments made in joint and separate evaluation. The standard view that joint and separate evaluation sets trigger different evaluation processes (Hsee et al., 1999) has led a number of researchers (e.g., Shafir, 2002) to ask whether joint versus separate evaluation mode leads to better decisions in particular domains. A rational analysis of joint-separate reversals, according to which apparent inconsistencies in joint and separate evaluation reflect the distinct inferences that are drawn from different evaluation sets (Sher & McKenzie, 2014), suggests a very different formulation of the prescriptive question. From the normative perspective, the critical question is simply how informative or misleading the information in the evaluation set is (i.e., how closely does the model of the attribute distribution inferred from the option-sample resemble the true distribution?). The joint evaluation context, which provides more information and thus may spark richer inferences, will presumably be preferable when evaluation sets are representative -- but not when they are misleading.

Alternatively, a choice architect who believes that frames, defaults, choice menus, and other task elements operate by conveying implicit information may consider making that information explicit, converting a covert "nudge" into an overt message. (This is in contrast to simply disclosing the intended effects of the nudge; see Loewenstein, Bryce, Hagmann, & Rajpal, 2015; Steffel, Williams, & Pogacar, 2016.) Explicit messages – provided that they are salient and easily processed – have some obvious advantages over implicit cues.

First, because the primary function of the choice menu is to gather rather than transmit information, the architect may have limited control over the contents of the menu. When the choice menu of interest *is* unrepresentative, explicit information can serve to correct the misleading inferences it would otherwise induce. For example, in evaluating a student applicant with a score on an unfamiliar standardized test, learning the test's mean score would enable the

evaluator to know if the student's score is above or below average (cf. Hsee et al., 1999), without relying on best guesses from contextual comparisons.

Second, implicit cues are frequently ambiguous, conveying unfocused information about multiple correlated variables. Explicit messages, by contrast, can be deliberately focused to resolve such ambiguity. For example, the choice architect's selection of a default may plausibly reflect injunctive norms (norms specifying what people should do) or descriptive norms (norms specifying what most people in fact do; cf. Cialdini, 2003). It may thus indiscriminately leak information about both (McKenzie et al., 2006). The explicit statement that a relevant authority recommends the selection of an option does not suffer from this ambiguity; the intended injunctive status of the explicit recommendation is unmistakable. Similarly, Sher and McKenzie (2006) showed that a speaker's choice of attribute frame not only leaks information about relative frequency (because, as noted above, speakers tend to describe options in terms of relatively abundant attributes) but may, like defaults, also leak implicit recommendations (because speakers tend to frame favorably perceived options in positive terms). While both kinds of information are broadly choice-relevant, they may be relevant in different ways. An explicit message – in this case, either that the attribute is relatively abundant, or instead that the speaker endorses the option – can uniquely signal one of the many confounded cues conveyed by a speaker's choice of frame.

Third, when messages are explicit, the communicative and persuasive intent of the choice architect is transparent to the DM. This preserves the DM's autonomy in deciding whether to accept or reject the architect's attempted influence (a nudge can't be resisted if it isn't detected), and may, in the long term, lessen skepticism about covert manipulation. In some cases, the clearer signals of persuasive intent in explicit messages, as well as the DM's greater awareness

and control in reacting to them, may diminish their impact. In particular, in circumstances where DMs would regard any deliberate top-down attempt to engineer choice as inappropriate, recommendations are likely to be less effective, and may even backfire, when they are explicit. But if we respect the DM's dignity, we should want the influence attempt to be less effective under these circumstances.

The effectiveness of overt communication, relative to covert nudges, is ultimately an empirical question. While the rational analysis reviewed here – which would suggest that carefully selected overt messages can effectively substitute for many covert nudges – has substantial explanatory power, we do not propose that it explains everything. In some cases, it may turn out empirically that highly effective covert nudges have no comparably effective overt counterpart. In these cases, the conscientious choice architect must seriously grapple with whether the proposed nudge, whatever its efficacy, is consistent with respect for the DM's dignity. In contrast, insofar as context effects stem from reasonable implicit inferences, this delicate dilemma need not arise. The choice architect can then resort to "rational nudges" which are only likely to strengthen with transparency.

Even where procedures of preference construction are assumed to be rational, they will be subject to the structural capacity limits that constrain human cognition more broadly. Accordingly, the choice architect must recognize that the DM is limited, not only in information, but also in attention, energy, and time. The provision of explicit or implicit information will only be effective if it respects these structural boundaries. Thus, even the choice architect who aims to mine the DM's potential for rationality will be guided by two simple but indispensable principles: Calibrate information load to attentional capacity; and, in selecting and organizing information, calibrate salience to relevance.

The Architect's Agenda: Process or Outcome Facilitation

Choice architects may differ in two ways – they may pursue different ends, and they may adopt different means for pursuing a given end. Especially in the policy domain, much of the controversy concerning nudges revolves around disagreements about the proper ends of choice architecture. We believe that prescriptive debates about choice architecture can be clarified by distinguishing between two kinds of ends – *process facilitation* and *outcome facilitation*.

Every choice is the outcome of a decision process. Choice architects can aim to broadly aid the process, or they can try to engineer specific outcomes that they believe are best for the DM. Because people can make good choices for bad reasons, the distinct ends of process and outcome facilitation will often (but not always) recommend the selection of different means. The central motivation of the original nudge program (Thaler & Sunstein, 2008) is outcome facilitation. The architect of nudges, having identified what seems to be a systematically suboptimal choice outcome (e.g., people save too little or eat too much), seeks to steer the decision process towards a better outcome. In some cases, this will involve redirecting apparent biases so that, like the broken clock that reads the right time twice a day, they point in the architect's favored direction. If a nudge of this kind is successful, more DMs will have made the right choice for the wrong reasons. The strategies of choice architecture we have explored above seek instead to improve processes, by conveying the most relevant and accurate information in the most transparent and chooser-friendly way.

Figure 4 summarizes these approaches to choice architecture, and relates them to a third – the "boosts" recently proposed by Hertwig and Grüne-Yanoff (2017). These approaches can be distinguished by their broad aims and by the specific decision stages they target. Cueing and boosting both aim at process facilitation, but they focus on different means to this shared end.

Note that two factors jointly determine the quality of a decision process: (1) the quality of the information fed to the process and (2) the appropriateness of the operations that transform information into option-selections. Cueing targets the first half of this equation. The goal is to improve the existing choice process by enhancing the accuracy, relevance, and convenience of its inputs. Such cueing is, we believe, the implicit guiding principle in many existing behavioral interventions – for example, automatically providing salient alerts to health care providers when they are about to refer a patient to a service with a long waiting time (Behavioural Insights Team, 2017, pp. 14-15). The boost strategy targets the second half of the equation in addition to the first. That is, rather than merely trying to enhance the inputs to a potentially rational decision process, a boost may seek to improve the decision process itself, such as by training DMs in skills that will allow them to both build on and go beyond existing competencies. Such interventions have the potential to empower DMs, though they may also require stronger assumptions about the drawbacks of the intuitive decision process, and about the costs and benefits of training. Finally, nudging simply aims to steer the DM towards a better outcome, by any effective non-coercive means – including, at times, by rerouting existing biases for the supposed benefit of the DM.

As broad motivations for choice architecture, outcome and process facilitation share some commonalities, but diverge in important ways. The "nudging" architect who seeks to facilitate outcomes may, at times, opt for a cue or a boost. Insofar as better information and improved skills are effective means to the end of better outcomes, the architect can use them to direct DMs towards the outcome she believes to be best for them (cf. "educative" nudges; Sunstein, 2018). But these process aids are only optional means to the choice architect's end, and interventions that bypass rationality or exploit irrationality may be just as suited to her

objective – which is to non-coercively engineer a foreordained outcome. Thus, for example, information salience will be strategically controlled in both approaches. But while the facilitator of process will strive to calibrate salience to relevance, the facilitator of outcomes will simply calibrate salience to desired effects. If clearly conveying the most relevant information happens to be the straightest path to a desired effect, the two architects will converge on a similar architecture. But if less relevant information provides a more potent nudge, their interventions are likely to differ.

Importantly, it is the choice architect's ultimate end, not her specific means of attaining it, that makes outcome facilitation paternalistic. The architect, like a parent nudging a child, presumes to know better than the DM what option will serve the DM's interests. Process facilitation, by contrast, is not inherently paternalistic. Its objective is to help the DM clearly understand and think about the problem, without prejudging the solution at which the DM should arrive. The two approaches, accordingly, have contrasting (dis)advantages. Outcome facilitation – helping people make the right choices – has the obvious advantage of greater directness. This may sometimes make for greater effectiveness, provided that the choice architect truly knows which outcome is best for the DM. Process facilitation – helping people make rational choices – has the advantage of respecting the DM's potential for rationality, while also recognizing the choice architect's own potential for error.

Conclusion

The analysis of constructed preference developed here has normative, descriptive, and prescriptive implications. Normatively, we have argued that rational decision making, like optimal perception, is inherently constructive and context-dependent. When prior knowledge is

limited, choice contexts may be informative, and the appropriate test of coherence is dynamic rather than static. Descriptively, we have reviewed evidence that three components of context – the frame, the default, and the menu – often serve as choice-relevant cues. Some well-known effects of these components, though traditionally taken as evidence of irrationality, turn out to pass the test of dynamic coherence. Preference changes across contexts are coordinated with context-based belief-updating.

Prescriptively, we have offered tentative answers to the question of how to assist DMs who are adaptively attuned to contextual cues, but have limited knowledge and cognitive resources. The approach explored here differs from the prototypical nudge in seeking to facilitate decision *processes* (e.g., by enhancing the inputs provided to them) rather than by trying to engineer foreordained *outcomes*. Finally, we have highlighted two advantages of choice architectures that build on the DM's existing potential for rationality: They are more respectful of individual dignity, and they are better-poised to preserve trust.

References

Altmann, S., Falk, A., & Grunewald, A. (2018). *Incentives and information as driving forces of default effects*. Unpublished manuscript.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences, 14*, 471-517.

Arad, A., & Rubinstein, A. (2017). *The people's perspective on libertarian-paternalistic policies*. Unpublished manuscript.

Ariely, D. (2009). *Predictably irrational: The hidden forces that shape our decisions.* Harper Collins.

Arkes, H. R., Gigerenzer, G., & Hertwig, R. (2016). How bad is incoherence? *Decision, 3*, 20-39.

Behavioural Insights Team. (2017). *The behavioural insights team update report 2016-17.*

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences, 3*, 57-65.

Cialdini, R. B. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science, 12*, 105-109.

Fine, I., Wade, A. R., Brewer, A. A., May, M. G., Goodman, D. F., Boynton, G. M., Wandell, B. A., & MacLeod, D. I. A. (2003). Long-term deprivation affects visual perception and cortex. *Nature Neuroscience, 6*, 915-916.

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

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Glaeser, E. L. (2006). Paternalism and psychology. *University of Chicago Law Review, 73*, 133-156.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review, 95*, 384–394.

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science, 12*, 973-986.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67*, 247–257.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin, 125*, 576-590.

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science, 5*, 343-355.

Iverson, G., & Falmagne, J.-C. (1985). Statistical issues in measurement. *Mathematical Social Sciences, 10*, 131-153.

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science, 302*, 1338–1339.

Kahneman, D. (2011). Thinking, fast and slow. New York: Farrar, Straus, and Giroux.

Kahneman, D. Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98*, 1325-1348.

Kahneman, D., & Tversky, A. (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press.

Keren, G. (2007). Framing, intentions, and trust–choice incompatibility. *Organizational Behavior and Human Decision Processes, 103,* 238-255.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55,* 271-304.

Kivetz, R., & Simonson, I. (2000). The effects of incomplete information on consumer choice. *Journal of Marketing Research, 37,* 427-448.

Klassen, A. C., & Klassen, D. K. (1996). Who are the donors in organ donation? The family's perspective in mandated choice. *Annals of Internal Medicine, 125,* 70–73.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211-228.

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge University Press.

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences, 27,* 313-376.

Leong, L. M., McKenzie, C. R. M., Sher, S., & Mueller-Trede, J. (2017). The role of inference in attribute framing effects. *Journal of Behavioral Decision Making, 30,* 1147-1156.

Levin, I. P. (1987). Associative effects of information framing. *Bulletin of the Psychonomic Society, 25,* 85–86.

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,* 149–188.

List, J. A. (2002). Preference reversals of a different kind: The "more is less" phenomenon. *American Economic Review, 92*, 1636-1643.

Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2015). Warning: You are about to be nudged. *Behavioral Science & Policy*, *1*(1), 35–42.

Löfgren, Å, Martinsson, P., Hennlock, M., & Sterner, T. (2012). Are experienced people affected by a pre-set default option—Results from a field experiment. *Journal of Environmental Economics and Management, 63*, 66-72.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Mandel, D. R. (2014). Do framing effects reveal irrational choice? *Journal of Experimental Psychology: General, 143*, 1185-198.

Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review, 104*, 277-290.

May, K. O. (1954). Intransitivity, utility, and the aggregation of preference patterns. *Econometrica, 22*, 1-13.

Mazar, N., Amir, O. and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*, 633–644.

McKenzie, C. R. M., Liersch, M. K., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science, 17*, 414–420.

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review, 10*, 596–602.

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.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608-631.

Oaksford, M. & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Prelec, D., Wernerfelt, B., & Zettelmeyer, F. (1997). The role of inference in context effects: Inferring what you want from what is available. *Journal of Consumer Research, 24,* 118-125.

Ratneshwar, S., Shocker, A. D., & Stewart, D. W. (1987). Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of Consumer Research*, 13 (4), 520-533.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2010). Testing transitivity of preferences on two-alternative forced choice data. *Frontiers in Psychology, 1* (148), 1-14.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review, 118*(1), 42–56.

Samuelson, W. & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1*, 7-59.

Shafir, E. (2002). Cognitoin, intuition and policy guidelines. In Gowda, M. V. & Fox, J. C., (Eds), *Judgments, decisions, and public policy* (pp. 71-90). Cambridge

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology, 53*, 491-517.

Shepard, R. N. (1990). *Mind sights: Original visual illusions, ambiguities, and other anomalies*. New York: Freeman.

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition, 101*, 467–494.

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.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research, 16,* 158-174.

Slovic, P. (1995). The construction of preference. *American Psychologist, 50,* 364-371.

Steffel, M., Williams, E. F., & Pogacar, R. (2016). Ethically deployed defaults: Transparency and consumer protection through disclosure and preference articulation. *Journal of Marketing Research, 53*(5), 865-880.

Sunstein, C. R. (2018). Misconceptions about nudges. *Journal of Behavioral Economics for Policy, 2*, 61-67.

Tannenbaum, D., Fox, C. R., & Rogers, T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour, 1,* 0130.

Tannenbaum, D., Valasek, C. J., Knowles, E. D., & Ditto, P. H. (2013). Incentivizing wellness in the workplace: Sticks (not carrots) send stigmatizing signals. *Psychological Science, 24*, 1512-1522.

Teigen, K. H., & Karevold, K. I. (2005). Looking back versus looking ahead: Framing of time and work at different stages of a project. *Journal of Behavioral Decision Making, 18,* 229-246.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629-640.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10,* 309-318.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331,* 1279-1285.

Thaler, R., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth and happiness*. New York, NY: Simon & Schuster.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76,* 31-48.

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

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* S251–278.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty, 5, 297-323.

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *American Economic Review, 80,* 204-217.

Author Note

An earlier draft of this article was presented at the “Behavioral Economics and New Paternalism” symposium, hosted by the Classical Liberal Institute, at the NYU School of Law on April 13 and 14, 2018. Shlomi Sher was supported by a Scholar Award from the James S. McDonnell Foundation. Correspondence regarding this article should be sent to the first author. Email: cmckenzie@ucsd.edu .

Table 1. Programmer problem from Sher & McKenzie (2014) (adapted with minor changes from Hsee, 1996).

	Candidate A	Candidate B
Education	BS in computer science from UCSD	BS in computer science from UCSD
GPA	3.8	3.1
Experience with YT	has written 10 YT programs	has written 70 YT programs

Table 2. Fictitious-attribute choice problem from Müller-Trede et al. (2015).

	Sound system A	***Sound system B***	***Sound system C***
Harmonic Range	10 mill	26 mill	16 mill
Sound Depth	17 sones	12 sones	29 sones
Acoustic Power	3.4 phons	2.0 phons	1.4 phons

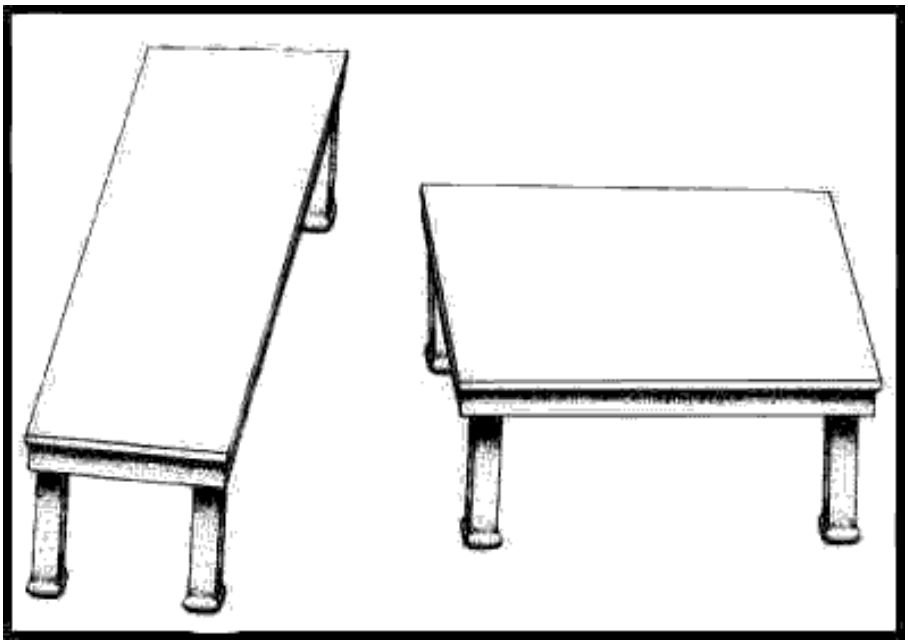

Figure 1. Shepard's (1990) tables.

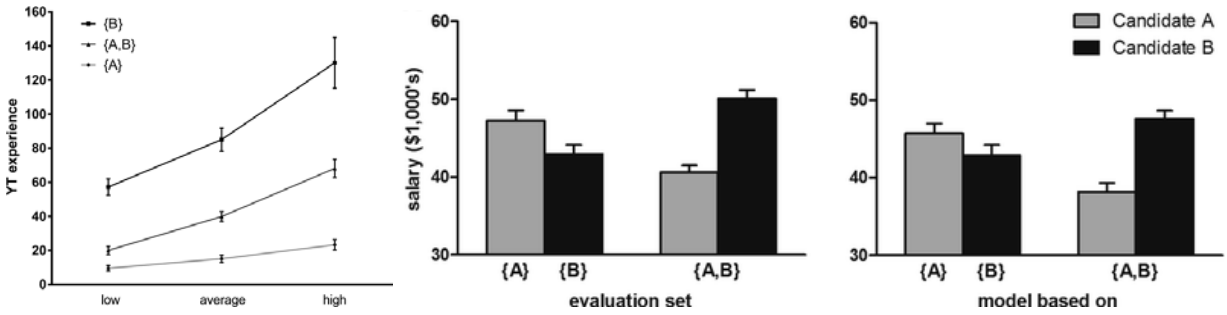

Figure 2. Inferences and evaluations in the joint-separate paradigm. The left panel shows mean low, average, and high values of YT experience estimated by participants who saw and evaluated one or both of the candidates ({A}, {B}, or {A, B}) from Table 1. The middle panel shows the mean salaries assigned by participants who saw one or both candidates. The right panel shows mean salaries assigned by other participants who evaluated a single candidate (A or B), based on the model of the YT experience distribution inferred by a yoked participant who had seen one of the three evaluation sets. Reprinted from Sher & McKenzie (2014).

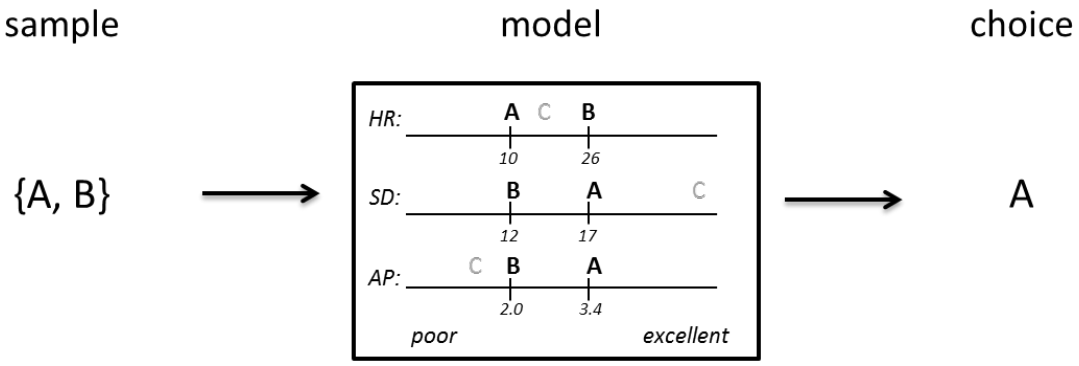

Figure 3. Options-as-information analysis of the choice problem in Table 2 (reprinted from Müller-Trede et al., 2015). Without prior knowledge of attribute distributions, the receipt of sample {*A*, *B*} in the choice set provides support for the inference that *A* is relatively good on two dimensions while *B* is relatively good on the third. If the weights attached to the three attributes are similar, this will lead to the selection of *A* over *B*. The standing of the non-sampled option *C* relative to the models of the attribute distributions inferred from sampling {*A*, *B*} is indicated in grey. HR = Harmonic Range; SD = Sound Depth; AP = Acoustic Power.

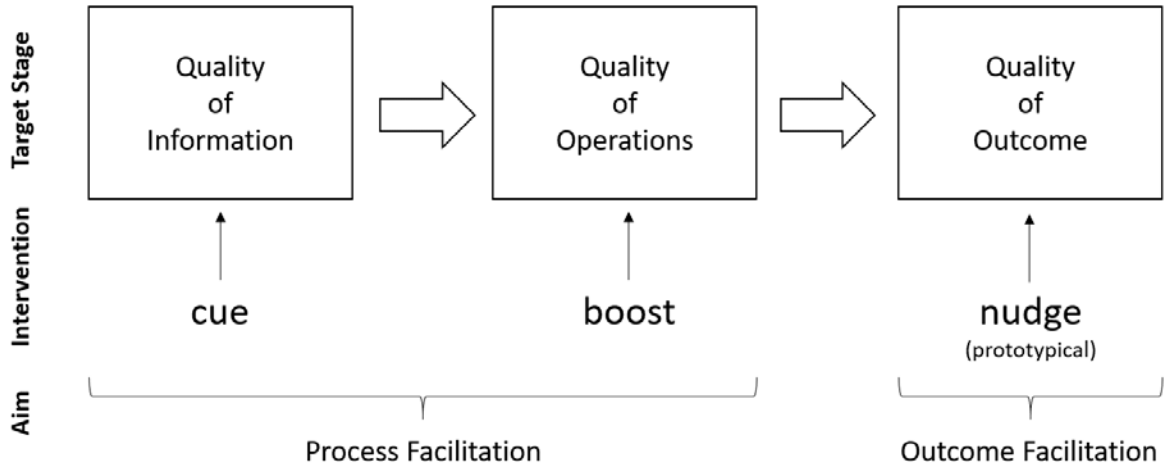

Figure 4. Two broad aims (process vs. outcome facilitation) motivating choice architecture interventions (cue, boost, nudge) that aim to improve the quality of distinct target stages in the choice process.