American Psychological Association
Psychological Review

0033-295X/18/$12.00

2018, Vol. 125, No. 5, 844–849
http://dx.doi.org/10.1037/rev0000121

COMMENTARY

The Construct–Behavior Gap and the Description–Experience Gap: Comment on Regenwetter and Robinson (2017)

Ralph Hertwig and Timothy J. Pleskac
Max Planck Institute for Human Development, Berlin, Germany

Regenwetter and Robinson (2017) discuss a challenging construct–behavior gap in psychological research. It can emerge when testing hypotheses that pertain to a theoretical construct (e.g., preferences) on the basis of observed behavior (e.g., actual choices). The problem is that the different heuristic methods that are sometimes used to link overt choices to covert preferences may ignore heterogeneity between and within individuals, rendering inferences drawn from choices to preferences invalid. Regenwetter and Robinson's remedy is to make heterogeneity an explicit part of the theory. They illustrate the problem and a remedy to it with the description–experience gap (D-E gap), the systematic gap in choices based on described versus 'experienced' probabilities. We welcome their sophisticated reanalysis of some early data sets, which, by taking heterogeneity into account, finds strong evidence for a D-E gap in probability weighting. Yet we see three issues with the remedy, which we likewise highlight using the D-E gap. First, the D-E gap cannot be reduced solely to probability weighting but rather unfolds across several different psychological constructs suggesting that part of the construct–behavior gap may stem from trying to reduce multidimensional behavior to a single construct. Second, the authors' modeling of heterogeneity leaves aside the heterogeneity of people's sampled experience in decisions from experience, which highlights the importance of also considering the potential causes of heterogeneity. Third, we identify potential sources of heterogeneity in choice behavior that go beyond probabilistic responses and preferences and advocate for a pluralistic approach to modeling it. Last but not least, we emphasize that, notwithstanding the importance of rigor and logical coherence in scientific theories, simplifications and (false) generalizations are indispensable in the pursuit of scientific knowledge.

Keywords: risky choice, decisions from experience, description–experience gap, heuristics, interindividual variability

First things first. Regenwetter and Robinson's (2017) contribution is impressive in its rigor and principled approach to what they call the construct–behavior gap in psychological research. This gap emerges in tests of hypotheses about a theoretical construct (e.g., people's preferences) via observed behavior (e.g., the choices they make). In particular, the various heuristic methods that are sometimes used to draw conclusions about constructs from behavior, such as preferences from choices, may ignore the potential heterogeneity between and within individuals, potentially rendering those conclusions logically invalid—a construct–behavior gap. Regenwetter and Robinson illustrate this problem in the context of what has become known as the description–experience gap (D-E gap), which refers to the systematic difference in choices people make when they learn about attributes of payoff distributions (henceforth gambles) on the basis of experience as opposed to symbolic descriptions (Hertwig & Erev, 2009). A frequent explanation of the D-E gap has been that people appear to differentially weight probabilities in gambles when making decisions from experience versus description. However, some past tests of this thesis have aggregated choice data across people and/or choice problems, accepting the risk of creating a construct–behavior gap.

Regenwetter and Robinson propose a remedy that explicitly models the heterogeneity between and within individuals, thereby, so they argue, bridging the gap between construct and behavior. In terms of the D-E gap, the implementation of their solution is based on the QTEST methodology for testing theories of binary preferences (e.g., Regenwetter et al., 2014). Applying their method to three original D-E gap data sets, Regenwetter and Robinson show that it is difficult to reach strong conclusions about the nature of probability weighting when description and experience conditions are analyzed separately. When preferences in the two conditions were jointly modeled, however, they found strong evidence for a D-E gap in probability weighting: based on the objective probabilities of the choice options, people overweighted rare events in description and underweighted rare events in experience (consistent with early conclusions about the D-E gap; Barron & Erev,

Ralph Hertwig and Timothy J. Pleskac, Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany.

Correspondence concerning this article should be addressed to Ralph Hertwig, Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: hertwig@mpib-berlin.mpg.de

2003; Hertwig, Barron, Weber, & Erev, 2004; Weber, Shafir, & Blais, 2004).

As elegant and rigorous as Regenwetter and Robinson's (2017) approach is, no method comes without costs. While we acknowledge the serious problem of the construct–behavior gap, we wonder how effective their specific method of modeling heterogeneity will be in closing it given what we have come to learn about the D-E gap. This comment has three parts. First, we argue that in order to make operative their solution to the construct–behavior gap in research on the D-E gap, the authors focus on a single dimension of the gap, namely, whether people weigh probabilities differently when choosing between gambles from description versus experience. Although early conclusions in the D-E literature did the same, later developments have shown that the D-E gap cannot be reduced to probability weighting but rather requires invoking additional cognitive constructs, with each one entering their own sources of heterogeneity. Second, even if the D-E gap were reducible to probability weighting, Regenwetter and Robinson's solution of modeling heterogeneity in preferences neglects the potentially large heterogeneity in experiences that people garner when sampling from the gambles. We show that conditioning the probability weights on objective probabilities rather than on experienced relative frequencies is questionable when the goal is to model heterogeneity of preferences in the D-E gap. Third, we suggest that identifying other potential causes of the D-E gap starts with modeling each individual's experience and its impact on preferences. More generally, we emphasize that no method of modeling human behavior is free of downsides. To close the construct–behavior gap we advocate for the use of multiple methods that focus on identifying the causes of heterogeneity and not to lose sight of the important role heuristics, simplifications, and false models, can and do play in science.

Two Interpretations of the D-E Gap: Choice Behavior and Probability Weighting

The original goal of research on the distinction between decisions from description and decisions from experience was to understand when and why choices deviate from expected value maximization and how the deviations differed in description and experience (Barron & Erev, 2003; Hertwig et al., 2004; Weber et al., 2004). Both the early and recent evidence (Erev, Ert, Plonsky, Cohen, & Cohen, 2017; Wulff, Mergenthaler-Canseco, & Hertwig, 2018) suggests that experience- and description-based choices systematically differ in how they deviate from expected value maximization. For instance, a meta-analysis of more than 40,000 trials examining the average proportion of expected value-maximizing choices under description and experience found that the average gap size (or difference between choice proportions for description and experience) was about 20 percentage points when the choice problems included a risky option and a safe option (for details see Wulff et al., 2018 Figure 3)—the type of problem frequently used to infer individual risk preference. This D-E gap does not hinge on any particular theory of choice, for instance, one that assumes probability weighting. It occurs across different experimental paradigms (see Hertwig & Erev, 2009) and within individuals (Camilleri & Newell, 2009).

One way to interpret this behavioral difference is in terms of prospect theory's probability weighting function, as did early observers of the D-E gap, for instance:

> Observed choices indicated not overweighting of small-probability outcomes (henceforth, *rare events*), but rather the opposite: People made choices as if they underweighted rare events; that is, rare events received less weight than their objective probability of occurrence warranted. (Hertwig et al., 2004, p. 535)

Such statements invited the inference that the D-E gap arises primarily from a reversal of the probability-weighting pattern, which, in turn, prompted numerous investigations of probability weighting in experience versus description (Abdellaoui, L'Haridon, & Paraschiv, 2011; Glöckner, Hilbig, Henninger, & Fiedler, 2016; Kellen, Pachur, & Hertwig, 2016; Ungemach, Chater, & Stewart, 2009).

Focused on this as-if weighting discussion of the D-E gap, Regenwetter and Robinson (2017) used probability weighting as a case study on how to close the construct–behavior gap. In our view, there are several reasons not to reduce the D-E gap to probability weighting. One is that the gap is robust across experimental paradigms. Regenwetter and Robinson focused on what has been called the *sampling paradigm* where people make one incentivized choice but sample from the options as long as they wish. However, the D-E gap has also been reported in the *partial-feedback paradigm* (Hertwig & Erev, 2009), where people make many incentivized choices. This difference is important because repeated experience-based choices have commonly been modeled without assuming probability weighting but rather using reinforcement learning models (Barron & Erev, 2003) or a set of choice strategies (Erev & Barron, 2005; Erev et al., 2017). While this does not exclude the possibility that studies employing the repeated choice paradigm also suffer from lack of modeling heterogeneity, the robustness of the D-E gap across paradigms at least suggests that different or at least additional constructs are at play.

A second, related reason is that the existence of a rare event (and its potential over- or underweighting) is not a *necessary* condition for the D-E gap. For instance, Ludvig and Spetch (2011) demonstrated that a D-E gap can arise when people make choices between a sure thing and a gamble with two equally likely outcomes. In this case, extreme outcomes have a greater impact on people's choices when they make decisions from experience than from description, which again suggests that constructs other than the probability weighting of rare events may rise to the D-E gap. A third reason is that probability weighting is a theoretical abstraction based on the strong assumption that experience-based choice rests on an explicit mental representation of probabilities. Such representations may or may not exist. Indeed, other models of experience-based choice do not assume probability weighting (e.g., Gonzalez & Dutt, 2011). A final reason is that even if probabilities are weighted, making a decision from experience involves much more in addition—for instance, a search process, a rule for stopping search, memory of the events experienced, a process for learning, and so on.

To conclude, reducing the D-E gap to the question of whether or not description and experience result in the same pattern of probability weighting obscures other differences in the learning and choice processes. Moreover, because the D-E gap is multifaceted, there are likely to be more construct–behavior gaps on these interrelated dimensions. As a result, modeling the weighting and utility function alone will not close it. We suggest that any future solution to construct–behavior gaps in research on the D-E gap will have to capture how these different constructs and processes work together to produce the observed choice behavior.

Weights of *Objective* Probabilities Differ in Experience and Description

For the sake of the next argument, let us assume that the D-E gap in the modeler's world is simple insofar as the primary cause of the gap is how people weight probabilities. Now the modeler faces another issue when seeking to close the construct–behavior gap by modeling heterogeneity. When people are masters of their own sampling efforts in decisions from experience, their very experience of the gambles in question will be heterogeneous. As we argue next, this reality requires attention in any attempt to close the construct–behavior gap.

Regenwetter and Robinson (2017) show that reliance on aggregate data could cause a gap between the construct of interest (e.g., individual's preference) and behavior (e.g., choices). For instance, treating the majority preference as indicative of individual preference can be problematic because it does not necessarily reflect any one person's preference: majority preference can satisfy the predictions of prospect theory regardless of whether any one person satisfies prospect theory (Estes, 1956; Regenwetter, Grofman, Marley, & Tsetlin, 2006). Thus, they point out, one should not conclude that a theory that captures aggregate behavior also captures the behavior of some individuals. Regenwetter and Robinson's solution to this problem is twofold. First, they model preferences across choice problems. Second, they account for heterogeneity in preferences between and within people by taking a set of problems and specifying a set of binary preference patterns across them that are consistent with the theory or hypothesis in question (e.g., rare events are overweighted in description and underweighted in experience). Heterogeneity between and within people is handled by permitting each pairwise preference to be probabilistic, and the prediction takes the form of a probability distribution across the set of binary preference patterns.

To implement this solution, one must therefore spell out all possible binary preference patterns across a set of problems and identify which of these patterns is consistent with a specific theory or hypothesis. For example, Hertwig et al. (2004) compared choices from description with choices from experience using six commonly used choice problems. For the D-E gap where preferences in the 'described' problems are compared to preferences in the 'experienced' problems, there are a total of $2^{12} = 4{,}096$ possible joint preference patterns. Only a subset of those preference patterns is consistent with a particular hypothesis (e.g., preferences in description- and experience-based decision making show the same degree of overweighting of rare events). This set of preference patterns forms a system of joint inequality constraints on binary choice probabilities. Regenwetter and Robinson tested if this system better accounted for the data than the null hypothesis assuming no constraint on the choice probabilities.

Table 1 summarizes what we understand to be the key results from Regenwetter and Robinson's (2017) analysis of probability weighting and the D-E gap. It reports several Bayes factors (BFs), each of which summarizes the weight of evidence in favor of one hypothesis relative to the null hypothesis (Kass & Raftery, 1995). A BF of more than 10 is typically interpreted as strong evidence for the hypothesis in question. We take away three important observations from Table 1.

First, after accounting for heterogeneity in preferences, there is strong evidence for a D-E gap in probability weighting, with rare objective probabilities being overweighted in description and underweighted in experience. Second, the evidence for a D-E gap in the weights of the objective probabilities might be even stronger than Table 1 suggests. With BFs, the strength of the evidence for a given hypothesis is captured relative to the hypothesis against which it is compared. The BFs in Table 1 estimate the strength of the evidence for a particular hypothesis over the null hypothesis that assumes no systematic relationship in preferences between choice problems. However, if one assumes that the D-E gap is better captured by comparing a reversed weighing function (Hypothesis III) to a more precise hypothesis, say, an isomorphic weighing hypothesis (Hypothesis I), then the evidence for the D-E gap is even stronger (e.g., $300/10^{-8} > 10{,}000$, based on the Hertwig et al., 2004, study).

A third observation pertains to the change in the BFs for a D-E gap across the three studies. Table 1 shows that the BFs for the D-E gap *decrease* across studies. This decrease is no coincidence. It reflects an attempt to investigate an important aspect of decisions from experience, namely, that when making them, people do not know the

Table 1
Summary of the Bayes Factors in Regenwetter and Robinson's (2017) Analysis of Three Hypotheses on the D-E Gap, Accounting for Variability in Preferences Between Participants

	Hertwig et al. (2004)	Hau et al. (2008)	Ungemach et al. (2009)
Hypothesis I: Preferences in description- and experience-based decision making show the same degree of overweighting.	10^{-8}	Unknown	.01
Hypothesis II: Preferences in description- and experience-based decision making are different, but both show overweighting.	.002	.06	.11
Hypothesis III: Preferences in description- and experience-based decision making are different, with overweighting in description and underweighting in experience.	~300	>100	>10

Note. The Bayes factors are relative to a null or saturated model (H_0) that places no constraints on the binary choice probabilities. One value is unknown because computation time exceeded a month (Regenwetter & Robinson, 2017, p. 545). For Hypothesis III, due to the computational demands of the calculation, the value for Hertwig et al. (2004) is an approximation and the values for Hau et al. (2008) and Ungemach et al. (2009) are estimated lower bounds.

possible outcomes and their probabilities. Instead, they learn from experience by sampling from the options and determine for themselves when to stop. Because some people will sample more and others less, their representations of a choice problem's possible payoffs and probabilities can differ substantially (and systematically) both from the actual payoffs and the objective probabilities and from one another's representations. That is, in decisions from experience, an important source of heterogeneity in people's preferences is the heterogeneity in their samples of experience and how those samples were generated (Wulff et al., 2018). The studies by Hau, Pleskac, Kiefer, and Hertwig (2008) and Ungemach et al. (2009) tried to reduce this heterogeneity by increasing the sample sizes people took: Hau et al. (2008) by increasing the stakes of the gambles (Experiment 2) or requiring participants to sample 100 times (Experiment 3) and Ungemach et al. (2009) by requiring participants to sample 40 times. It is in these studies that Regenwetter and Robinson (2017) find *lower* (though still credible) evidence for a D-E gap in probability weighting (see Table 1). Without acknowledging participants' very different amounts of experience, it is difficult to interpret this finding

To be very clear: Regenwetter and Robinson (2017) are fully transparent about their decision to focus on objective probabilities, and their deductive analysis is coherent as a result. Yet when the goal is to test the existence of a D-E gap in probability weighting while taking heterogeneity in preferences into account, leaving aside people's individual experience of outcomes—a defining feature of decisions from experience—strikes us as a debatable choice. Moreover, as the decreasing BFs across the D-E gaps studies illustrate, this choice impacts the authors' analyses and interpretation.

We are not sure how easily Regenwetter and Robinson's solution of modeling heterogeneity via the joint preference patterns across problems can be expanded to include this source of heterogeneity because, in decisions from experience, people's individual experience with the problem shapes their representation of it. Thus, what appears to be *underweighting* in terms of *objective* probabilities could actually be *overweighting* in terms of the *experienced* relative frequencies, or vice versa. More recent research on the D-E gap, which acknowledges the heterogeneity of sample experience, has focused on the experienced relative frequencies when analyzing probability weighting (Fox & Hadar, 2006; Glöckner et al., 2016; Kellen et al., 2016) and when modeling the direct construction of preferences from experience via a memory process (Gonzalez & Dutt, 2011) or via a process of evidence accumulation (Markant, Pleskac, Diederich, Pachur, & Hertwig, 2015).

Admittedly, each of these approaches also comes with costs, for instance, by ultimately making stronger parametric assumptions than Regenwetter and Robinson's (2017) approach or ignoring the variability between individuals. Our more general point is that even if one could reduce the multidimensionality of the D-E gap to one psychological construct—probability weighting, say—there will be heterogeneity in the sampled experience (in the sampling paradigm), which, in turn, must be considered when trying to build a bridge from construct to behavior. Otherwise, it is hard to imagine how we will ever make sense of the strong evidence that Regenwetter and Robinson find for reversed probability weighting of the objective probabilities in description and experience and the quite different (and mixed) results of the weighting of experienced probabilities (see Table 9 in Wulff et al., 2018).

A Pluralism of Theories of Heterogeneity

According to Regenwetter and Robinson (2017), behavioral decision research needs models of heterogeneity to bridge constructs and behavior. In both theory and practice, we agree with them about the importance of modeling individual behavior and differences. Yet we also advocate for pluralism in this endeavor. Models of heterogeneity address the causes of heterogeneity and seek ways and how to model them. The *hypotheses* Regenwetter and Robinson examined reflect the default approach in psychology and economics, which is to capture heterogeneity in people's response or their underlying preferences in terms of variability in how they weight probabilities or value outcomes. But it is unclear to what extent interindividual differences in response consistency or in the parameter values of, say, the probability-weighting and utility functions predict meaningful differences between people. While some work suggests that meaningful differences exist (e.g., Pleskac, 2008), other work casts doubt on the power of parameters to predict, for instance, risk or social preferences (e.g., Blanco, Engelmann, & Normann, 2011; Friedman, Isaac, James, & Sunder, 2014).

There are other ways to model heterogeneity. One that we already mentioned is to consider how people's experiences shape their preferences. This link between search and choice has long been recognized—for instance, by March (1996; Denrell, 2007), who showed that experiential learning alone can give rise to systematic risk preferences. Since, in decisions from experience, individuals partly control the experience they gather, modeling heterogeneity in preferences would seem to require modeling differences in how they search and learn before making a choice (including, for instance, differences in working memory capacity; Rakow, Demes, & Newell, 2008). Sequential sampling models offer one starting point (Busemeyer & Diederich, 2002; Zeigenfuse, Pleskac, & Liu, 2014) by describing processes such as how a person switches between options, stops search, and integrates experience into a preference (Markant et al., 2015).

Heterogeneity may also arise from different choice strategies. For instance, if different people might apply different strategies or heuristics to the same choice problem (Erev & Barron, 2005; Payne, Bettman, & Johnson, 1993), individual differences in the shapes of prospect theory's parametric functions could stem, at least in part, from their use of different heuristics (Pachur, Suter, and Hertwig (2017). Furthermore, heterogeneity could be modeled in terms of the probabilistic nature of choice. Instead of invoking the framework of classical probability, however, researchers could turn to quantum probability models (Busemeyer & Bruza, 2012), which trace the heterogeneity in preferences to their indefinite nature (e.g., Kvam, Pleskac, Yu, & Busemeyer, 2015).

Finally, let us emphasize a risk in the detailed modeling of modeling heterogeneity. Regenwetter and Robinson (2017) suggest that all the fluctuations in a person's choice behavior "must be taken extremely seriously in decision theory" (p. 534). A prediction model that aims to capture every fluctuation in a person's behavior must be flexible. That is a strength when the goal is to perfectly describe all the idiosyncrasies in, say, a choice pattern. It may be a liability, however, when the goal is to predict a person's future choice. Capturing all fluctuations in a finite, noisy sample of choices runs the risk of overfitting known data, that is, compromising the ability to predict the same person's as-yet-unknown behavior under new conditions. Thus, when making inferences from a finite sample of choice data, which are commonly noisy and limited, it may pay to bet on a simpler model that poses a lower risk of overfitting (a problem also known as the bias/variance dilemma; e.g., Geman, Bienenstock, & Doursat, 1992).

All these methods of modeling heterogeneity likely face the same tradeoff between capturing the idiosyncrasies in observed behavior and predicting future behavior. Regenwetter and Robinson's (2017) remedy aims to guard against overfitting by using a Bayesian model comparison approach and penalizing more complex models (e.g., Myung & Pitt, 1997), but no approach to penalizing models for complexity is fail-safe. The Bayesian model comparison approach focuses on accounting for the data at hand, seeking the best way to reduce it into a simpler form. Implied in this reductionist approach is the assumption that the most parsimonious model will predict well. What we miss in Regenwetter and Robinson's (2017) analysis is a more explicit examination of how well models that embrace heterogeneity quantitatively *predict* new behavior under new conditions (vs. describing known behavior; see Busemeyer & Wang, 2000; Erev et al., 2017).

To conclude, heterogeneity in choice likely has various causes and can be modeled in various ways. When variability is solely due to measurement error, it seems sufficient to include an error term in the model. When it is due to other differences, such as in the cognitive system (e.g., working memory, age; Lindenberger & Mayr, 2014) or in choice strategies, however, more than response error is required. In that case, the default approach of reducing heterogeneity in people's behavior to response error or differences in the subjective functions of outcomes and probabilities—as Regenwetter and Robinson's (2017) hypotheses do but not necessarily their larger modeling framework—has limited ability to shed light on the underlying causes of the variability or to predict people's choices under new conditions.

Conclusion

Researchers seeking scientific progress will almost by definition make bold and unjustified claims, some of which will prove to be mistaken. For illustration, take prospect theory, which Regenwetter and Robinson (2017) invoked to analyze the D-E gap. Prospect theory has its roots in tests involving aggregate data (modal choice) and a small set of carefully constructed choice problems. On this basis, Kahneman and Tversky (1979) concluded that "utility theory, as it is commonly interpreted and applied, is not an adequate descriptive model" and instead proposed "an alternative account of choice under risk" (p. 263). If we understand Regenwetter and Robinson correctly, this conclusion represents a sweeping generalization and is based on what Regenwetter and Robinson call a fallacy of composition. Yet until today prospect theory is *the* descriptive theory of risky choice.

In the pursuit of scientific knowledge, researchers need to work hard to be logically coherent and propose precise theories. Yet even the best among them cannot help but simplify, generalize and propose biased and false models (Wimsatt, 2007). Indeed, Regenwetter and Robinson's (2017) sophisticated analysis would not work without an important simplification, namely, the weighting by the objective probabilities rather than the probabilities actually experienced. Our theories and hypotheses, all too simple and ultimately false, inform the design of experiments, the selection of experimental stimuli, and the statistical analyses employed. They are rarely derived from surefire logico-deductive methods but rather are often based on heuristic processes. It is hard not to read Regenwetter and Robinson's position as equating the use of heuristics in science with undermining progress in science. In our view, this equation would misconstrue the important role of heuristics in the discovery, development, and testing of theories (Gigerenzer, 1991).

References

Abdellaoui, M., L'Haridon, O., & Paraschiv, C. (2011). Experienced vs. described uncertainty: Do we need two prospect theory specifications? *Management Science, 57,* 1879–1895. http://dx.doi.org/10.1287/mnsc.1110.1368

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16,* 215–233. http://dx.doi.org/10.1002/bdm.443

Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior, 72,* 321–338. http://dx.doi.org/10.1016/j.geb.2010.09.008

Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511997716

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences, 43,* 345–370. http://dx.doi.org/10.1016/S0165-4896(02)00016-1

Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44,* 171–189. http://dx.doi.org/10.1006/jmps.1999.1282

Camilleri, A. R., & Newell, B. R. (2009). Within-subject preference reversals in description- and experience-based choice. In N. Taatgen & H. V. Rijn (Eds.), In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 449–454). Austin, TX: Cognitive Science Society. Retrieved from https://escholarship.org/uc/item/4h4328gv

Denrell, J. (2007). Adaptive learning and risk taking. *Psychological Review, 114,* 177–187. http://dx.doi.org/10.1037/0033-295X.114.1.177

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review, 112,* 912–931. http://dx.doi.org/10.1037/0033-295X.112.4.912

Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological Review, 124,* 369–409. http://dx.doi.org/10.1037/rev0000062

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53,* 134–140. http://dx.doi.org/10.1037/h0045156

Fox, C. R., & Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making, 1,* 159–161.

Friedman, D., Isaac, R. M., James, D., & Sunder, S. (2014). *Risky curves: On the empirical failure of expected utility*. London, UK: Routledge.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation, 4,* 1–58. http://dx.doi.org/10.1162/neco.1992.4.1.1

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review, 98,* 254–267. http://dx.doi.org/10.1037/0033-295X.98.2.254

Glöckner, A., Hilbig, B. E., Henninger, F., & Fiedler, S. (2016). The reversed description–experience gap: Disentangling sources of presentation format effects in risky choice. *Journal of Experimental Psychology: General, 145,* 486–508. http://dx.doi.org/10.1037/a0040103

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological Review, 118,* 523–551. http://dx.doi.org/10.1037/a0024558

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21,* 493–518. http://dx.doi.org/10.1002/bdm.598

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15,* 534–539. http://dx.doi.org/10.1111/j.0956-7976.2004.00715.x

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

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291. http://dx.doi.org/10.2307/1914185

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572

Kellen, D., Pachur, T., & Hertwig, R. (2016). How (in)variant are subjective representations of described and experienced risk and rewards? *Cognition, 157,* 126–138. http://dx.doi.org/10.1016/j.cognition.2016.08.020

Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. *Proceedings of the National Academy of Sciences of the United States of America, 112,* 10645–10650. http://dx.doi.org/10.1073/pnas.1500688112

Lindenberger, U., & Mayr, U. (2014). Cognitive aging: Is there a dark side to environmental support? *Trends in Cognitive Sciences, 18,* 7–15. http://dx.doi.org/10.1016/j.tics.2013.10.006

Ludvig, E. A., & Spetch, M. L. (2011). Of black swans and tossed coins: Is the description-experience gap in risky choice limited to rare events? *PLoS ONE, 6*(6), e20262. http://dx.doi.org/10.1371/journal.pone.0020262

March, J. G. (1996). Learning to be risk averse. *Psychological Review, 103,* 309–319. http://dx.doi.org/10.1037/0033-295X.103.2.309

Markant, D., Pleskac, T. J., Diederich, A., Pachur, T., & Hertwig, R. (2015). Modeling choice and search in decisions from experience: A sequential sampling approach. In R. Dale, . . . (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1512–1517). Austin, TX: Cognitive Science Society.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4,* 79–95. http://dx.doi.org/10.3758/BF03210778

Pachur, T., Suter, R. S., & Hertwig, R. (2017). How the twain can meet: Prospect theory and models of heuristics in risky choice. *Cognitive Psychology, 93,* 44–73. http://dx.doi.org/10.1016/j.cogpsych.2017.01.001

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139173933

Pleskac, T. J. (2008). Decision making and learning while taking sequential risks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 167–185. http://dx.doi.org/10.1037/0278-7393.34.1.167

Rakow, T., Demes, K. A., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes, 106,* 168–179. http://dx.doi.org/10.1016/j.obhdp.2008.02.001

Regenwetter, M., Davis-Stober, C. P., Lim, S. H., Guo, Y., Popova, A., Zwilling, C., . . . Messner, W. (2014). QTEST: Quantitative testing of theories of binary choice. *Decision, 1,* 2–34. http://dx.doi.org/10.1037/dec0000007

Regenwetter, M., Grofman, B., Marley, A. A. J., & Tsetlin, I. M. (2006). *Behavioral social choice: Probabilistic models, statistical inference, and applications*. New York, NY: Cambridge University Press.

Regenwetter, M., & Robinson, M. M. (2017). The construct–behavior gap in behavioral decision research: A challenge beyond replicability. *Psychological Review, 124,* 533–550. http://dx.doi.org/10.1037/rev0000067

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science, 20,* 473–479. http://dx.doi.org/10.1111/j.1467-9280.2009.02319.x

Weber, E. U., Shafir, S., & Blais, A. R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review, 111,* 430–445. http://dx.doi.org/10.1037/0033-295X.111.2.430

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard University Press.

Wulff, D. U., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin, 144,* 140–176. http://dx.doi.org/10.1037/bul0000115

Zeigenfuse, M. D., Pleskac, T. J., & Liu, T. (2014). Rapid decisions from experience. *Cognition, 131,* 181–194. http://dx.doi.org/10.1016/j.cognition.2013.12.012

Received April 21, 2017
Revision received May 13, 2018
Accepted June 15, 2018 ■