

Manuscript for the Theme Issue in *Philosophical Transactions B*

“Risk taking and impulsive behaviour: Fundamental discoveries, theoretical perspectives and clinical implications”

### **Three Gaps and What They May Mean for Risk Preference**

Ralph Hertwig

Max Planck Institute for Human Development, Berlin, Germany

Dirk U. Wulff and Rui Mata

University of Basel, Basel, Switzerland

& Max Planck Institute for Human Development, Berlin, Germany

Abstract: 218 words

Total word count (including main text & references): 8,051 (without Supplementary Material)  
3 figures & 1 table

We have no competing interests.

Author contributions: D.U.W analyzed data; R.H., D.U.W and R.M. wrote the paper.

Key words: risk preference, description–experience gap, revealed preferences, stated preference, convergent validity, temporal stability

#### Author Note

Ralph Hertwig, Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany. Dirk U. Wulff and Rui Mata, University of Basel, Department of Psychology and Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany. We are grateful to Deborah Ain for editing the manuscript. Correspondence 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](mailto:hertwig@mpib-berlin.mpg.de)

### **Abstract**

Risk preference is one of the most important building blocks of choice theories in the behavioural sciences. In economics, it is often conceptualised as preferences concerning variance of monetary payoffs, whereas in psychology risk preference is often thought to capture the propensity to engage in behaviour with the potential for loss or harm. Both concepts are associated with distinct measurement traditions: Economics has traditionally relied on behavioural measures, while psychology has often relied on self-reports. We review three important gaps that have emerged from work stemming from these two measurement traditions: First, a description–experience gap which suggests that behavioural measures do not speak with one voice and can give very different views on an individual’s appetite for risk; second, a behaviour–self-report gap which suggests that different self-report measures, but not behavioural measures, show a high degree of convergent validity; and, third, a temporal stability gap which suggests that self-reports, but not behavioural measures, show considerable temporal stability across periods of years. Risk preference, when measured through self-reported preferences—but not behavioural preferences—appears as a moderately stable psychological trait with both general and domain-specific components. We argue that future work needs to address the gaps which have emerged from the two measurement traditions and test their differential predictive validity for important economic, health and wellbeing outcomes.

The construct of risk preference is one of the most important building blocks of economic and psychological theories of choice. It is often invoked to explain behaviours and interindividual differences therein in domains as diverse as individuals' financial choices, unlawful behaviours (e.g., speeding, tax evasion), health choices (e.g., consuming recreational and possibly illicit drugs), and professional choices (e.g., entrepreneurial initiatives). Risk preference—also termed “risk attitude,” “risk tolerance,” or “sensitivity to risk”—is often understood to represent a personal characteristic. Despite this default view, opinions about how to best conceptualise this construct vary (Barseghyan, Molinari, O'Donoghue, & Teitelbaum, 2018; Stigler & Becker, 1977), including whether it represents a stable individual characteristic (Chuang & Schechter, 2015; Schildberg-Hörisch, 2018), how it relates to other mainstay psychological constructs such as impulsivity (Frey, Pedroni, Mata, Rieskamp, & Hertwig, 2017; Nigg, 2016), and how individual differences in risk preference should be measured (Barseghyan et al., 2018; Beshears, Choi, Laibson, & Madrian, 2008; Mata, Frey, Richter, Schupp, & Hertwig, 2018).

In what follows, we distinguish between two major measurement traditions of investigating individual differences in risk preference. One originates in economics and rests on behavioural measures, such as choice between monetary gambles. Another, originating in psychology, tends to rely on self-reports. As we argue below, understanding the two approaches and their somewhat conflicted relationship is key to assessing the current literature on risk preference. We then review evidence which suggests that the two sets of measures provide different and almost opposing views about the nature of risk preference. Specifically, we introduce three gaps between measurements of risk preferences in economics and psychology that have been identified recently: first, a gap in the behavioural patterns observed for monetary gambles presented in different formats, the *description–experience gap*; second, a gap in the convergent validity of behavioural and self-report measures, the

*behaviour–self-report gap*; and, third, a gap in the observed temporal stability of behavioural and self-report measures, the *temporal stability gap*. We conclude that future work must reconcile the contradictory findings produced by the two measurement traditions. In particular, we propose that a systematic and evidence-based understanding of the relation between different measures will be crucial for making conceptual and empirical progress in the study of risk preferences.

### **What is Risk Preference?**

When economists and psychologists call behaviours “risky” they use the same term but mean different things. In economics and finance, risk preference commonly refers to the tendency to choose an action that involves higher variance in potential monetary outcomes, relative to another option with lower variance of outcomes (but equal expected value). This holds independent of whether these outcomes involve gains or losses (Markowitz, 1952; Pratt, 1964). For example, when offered the choice between a safe option of receiving €500 guaranteed and a risky option of a 50% chance of receiving €1,000 and a 50% chance of receiving nothing, a risk-neutral person would not prefer one option over the other. Expected value maximisation (i.e., multiplying all outcomes per options with their respective probabilities, summing the products, and maximising), which embodies risk neutrality, values both options equally. However, other modelling frameworks make it possible to consider risk-averse and risk-loving individuals as well. For example, according to expected utility theory, a classic model in economics (Bernoulli, 1956/1738), a risk-averse person may be said to possess a concave utility function that leads to a preference for choosing the safe option, implying that gaining €500 contributes more than half the utility of a 50% chance of gaining €1,000, thus accommodating their choice of the sure option. Likewise, a risk-prone person’s utility curve may be said to be convex. There is a long tradition of using such mathematical theories in economics (Schoemaker, 1982), with several competing formulations now

available to capture individuals' risk preferences (O'Donoghue & Somerville, 2018). Let us highlight and clarify that utility-based economic modelling of human behaviour has not only been applied to risky activities such as financial investments but also to 'choices' such as drug use (addiction; Becker & Murphy, 1988) and criminal behaviours: As Becker (1968) put it: "a person commits an offense if the utility to him exceeds the utility he could get by using his time and other resources at other activities" (p. 176).

In psychology, risk preference is often broadly interpreted as the propensity to engage in behaviours or activities that, although rewarding, involve the potential for loss or harm (for oneself or others). Psychologists have shown less interest than have economists in comparing the implications of different mathematical formulations of utility for risk preference (with prominent exceptions; Kahneman & Tversky, 1979), focusing instead on understanding whether latent constructs derived from self-reports of attitudes and behaviours are associated with drug use or daring activities such as speeding, rock climbing, and imprudent online behaviour—all of which may be rewarding but also carry the possibility of physical or psychological harm (e.g., Steinberg, 2013).

Although we adopt the umbrella term "risk preference," we note that a number of distinct but overlapping psychological constructs have been used to account for individual differences in such "risky" behaviours, including impulsivity (Romer, 2010), sensation-seeking (Zuckerman, 2008), novelty-seeking (Kelley, Schochet & Landry, 2004), and impulse control (Steinberg, 2013). Further, a number of psychological theories concerning these constructs do not necessarily propose them as unitary. For example, the tripartite view on impulsivity suggests that it consists of components of reward sensitivity, loss sensitivity, and inhibitory control (see Cross, Copping, & Campbell, 2011). Ultimately, the degree of overlap between these constructs will be largely an empirical question—an issue we discuss in more detail below.

## Two Measurement Traditions: Revealed and Stated Preferences

One would expect that a construct as important and frequently invoked as risk preference rests on a firm measurement foundation, but this is far from the truth. Indeed, the conceptualisations of risk preference in economics and psychology gave rise to two distinct ways of measuring risk preference (for reviews see Charness, Gneezy & Imas, 2013; Mata et al., 2018; Schonberg, Fox, & Poldrack, 2011)<sup>1</sup>. How these measures relate to each other has, at least until recently, received scant attention. The *revealed-preference tradition* (Samuelson, 1938, 1948) holds that people's utilities and true beliefs are revealed through the (incentivised) choices they make.<sup>2</sup> Consequently, this tradition has relied predominantly on simple monetary gambles (e.g., Ellsberg, 1961; Harrison & Rutström, 2008; Kahneman & Tversky, 1979) or extensions of it, such as the multiple-price-list method (Holt & Laury, 2002). It is unsurprising that choices among monetary gambles have played an outsized role in measuring risk preference in economics: They were midwives of the Enlightenment concept of mathematical expectation (Hacking, 1975); they gave rise, through the St. Petersburg gamble, to what is known today as expected utility theory (Bernoulli, 1738/1954); they were invoked to demonstrate that people's choices are at odds with axioms of expected utility theory (e.g., Allais, 1953; Ellsberg, 1961); they were enlisted to demonstrate the key concepts of prospect theory such as loss aversion and the fourfold pattern of risk preference (Kahneman & Tversky, 1979); and last but not least, one simple gamble type—the choice

---

<sup>1</sup> Admittedly, this binary distinction is a simplification. In our past work, we have further distinguished between self-reported propensity measures (assessing stated preferences), and self-reported frequency measures (tracking specific and observable behaviours; Frey et al., 2017). Others have considered epidemiological data, such as crime or cause-specific mortality (Steinberg, 2013); administrative data, such as arrests (Moffitt et al., 2011); field data, such as property or health insurance data (Barseghyan et al., 2018); and informant reports from relatives or acquaintances (Roberts, Lejuez, Krueger, Richards, & Hill, 2014). Most of the work on risk preference, however, rests on behavioural and self-report measures; our focus here is therefore on them.

<sup>2</sup> Samuelson (1938) thus aimed to overcome what he criticised as the “discrediting of utility as a psychological concept” within cardinal utility theory that cannot explain human behaviour because of its “circular sense, revealing its emptiness as even a construction” (p. 61).

between a safe and a risky option—has frequently been employed to gauge people’s risk preference.

Proponents of the *stated-preference tradition*, in contrast, bank on people’s introspective abilities rather than their observable behaviour. They elicit data concerning risk using relatively general questions (“Are you generally a risk-taking person or do you try to avoid risks?”; Dohmen et al., 2011), specific but hypothetical questions (“How likely would you be to go whitewater rafting at high water in the spring?”; Blais & Weber, 2006), or ask people to report on the frequency of actual risky activities (“How many cigarettes do you smoke per day?”). Self-report measures have been widely used in applied and epidemiological contexts, presumably because they are easy to administer. For example, financial institutions often rely on self-report measures to gauge their clients’ risk preference so as to meet the legal requirements for the sale of financial products (Marinelli & Mazzoli, 2011). Self-report measures of risk preference have commonalities with those used to measure overlapping constructs such as impulsivity (Cross et al., 2011) and sensation-seeking (Zuckerman, Kolin, Price, & Zoob, 1964).

Is the existence of these two measurement traditions consistent with a single latent trait that comes into sight regardless of how it is being probed? Or do the measures “make” the construct—do two distinct constructs surface if these measures are employed simultaneously? Recent years have seen considerable efforts to assess the operation and implications of these measures, including their convergent validity (Frey et al., 2017), temporal stability (Frey et al., 2017; Mata et al., 2018); their associations with personality traits and demographic characteristics such as age, gender, and cognitive abilities (Dohmen, Falk, Golsteyn, Huffman, & Sunde, 2017; Josef et al., 2016); and their genetic basis (Benjamin et al., 2012; Linnér et al., 2018). In what follows, we describe three major findings involving measures stemming from both traditions that each suggest an incongruity to be addressed.

### The Description–Experience Gap

Monetary gambles are first among equals in behavioural measures. Interestingly, there is little variability in their guise. Researchers typically present the options' outcomes and probabilities numerically (e.g., €500 guaranteed vs. €1,000 with .5; €0 with .5) or with a spinner wheel or bar chart (see the meta-analysis from Weber, Blais, & Betz, 2002), and gambles explicitly state all possible outcomes and their probabilities. This nearly invariant choice architecture for measuring people's response to risk is rather odd because "it is hard to think of an important natural decision for which probabilities *are* objectively known" (Camerer & Weber, 1992, p. 325). Indeed, in everyday life, people rarely encounter convenient descriptions of objective probability distributions, with a few exceptions such as the probability of rain (e.g., Gigerenzer, Hertwig, van den Broek, Fasolo, & Katsikopoulos, 2005). Instead, people must turn to whatever experience they may have, making *decisions from experience* rather than the *decisions from description* that are often studied in laboratories (Hertwig, Barron, Weber, & Erev, 2004; Wulff, Mergenthaler Canseco, & Hertwig, 2018).

Decisions from description versus decisions from experience—this simple distinction, which should be understood more as poles on a continuum rather than a dichotomy, raises a new question: Do these two modes of learning about the probabilistic texture of the world (Hertwig, Hogarth, & Lejarraga, 2018) result in the same or systematically different choices? The question has received much attention since three articles in the early 2000s (Barron & Erev, 2003; Hertwig et al., 2004; Weber et al., 2002) demonstrated a systematic discrepancy in description- and experienced-based choices: the *description–experience gap* (for reviews see Hertwig, 2015; Hertwig & Erev, 2009; Rakow & Newell, 2010). These and many subsequent studies presented both gambles in which all outcomes and probabilities are stated (description) and gambles in which the payoff distributions were initially unknown but people could randomly draw from them (experience). Each draw produced one outcome; draw-by-



draw, the properties of the outcome distributions were revealed. Two major experiential paradigms have been employed (although many hybrid variants exist): In the *sampling paradigm*, people first sample as many outcomes as they wish, then decide from which distribution to make a single draw. In the *partial feedback paradigm*, each of a typically large, fixed number of draws contributes to people's earnings and they receive draw-by-draw feedback on the obtained payoffs. The sampling paradigm removes the exploitation–exploration tradeoff (Cohen, McClure & Yu, 2007), whereas the partial-feedback paradigm incorporates it. Assuming that people sample sufficiently and equally across the payoff distribution, the description and experience offer equivalent information. But are the resulting choices equivalent?

### **Manifestations of the Description–Experience gap**

As it turns out, they are often not equivalent. There are several ways to illustrate the description–experience gap (Hertwig & Erev, 2009; Rakow & Newell, 2010). Figure 1 plots the gap in terms of a systematic difference in the observed choice proportions in description and experience, as a function of gamble type (and for a subset of studies that examined the gap rather than its boundary conditions; see Wulff et al., 2018 for how the systematic differences were determined).<sup>3</sup> When a choice involves a risky and a safe option—the choice task often used to behaviourally measure risk preference—the gap is 18.7 percentage points; when a choice involves two risky options the gap is 7 percentage points.

[Figure 1]

---

<sup>3</sup> The literature has adopted different ways of operationalising the description–experience gap. One approach is to count how many of individuals' actual choices in the description and experience condition, respectively, are consistent with the predicted choices, based on cumulative prospect theory's (CPT) parameters (commonly using those derived by Tversky & Kahneman, 1992). These parameters, derived from stated probabilities, embody overweighting of rare events; therefore, choices consistent with the predictions of CPT indicate a tendency to overweight rare events. When this definition is applied, a description–experience gap emerges when systematically fewer experienced-based than description-based choices are correctly predicted.

A second manifestation of the description–experience gap pertains to the maximisation rate. In decisions from description, Wulff et al. (2018) found a median of 55% of choices maximised expected value; in decisions from experience, in contrast, 66% of people who encountered all possible outcomes and 89% of people who experienced some, but not all outcomes (often missing the rare event) maximised the experienced mean return—that is, the “expected value” of the actually experienced sample of outcomes. A third way to demonstrate that choices in description and experience are systematically different and one that brings us straight back to risk preference focuses on the *fourfold pattern of risk preference* (Tversky & Kahneman, 1992). This pattern is shown in table 1. The classic model of decisions under risk, expected utility theory, assumes that individuals are generally risk averse (i.e., a concave utility function). Challenging this view, Tversky and Kahneman showed that people are both risk averse and risk seeking. Consider the choice proportions in gambles with stated probabilities. In the gain domain most people were risk averse, preferring the safe option when the probability of winning was high (4 with .8). When the gamble had the same expected value but a low probability of winning (32 with .1), preference reversed: Most people were risk seeking and chose the risky option. With the same choices but with outcomes in the loss domain, preferences flipped again: Many people proved risk averse when the stated probability of losing was low (-32 with .1) but risk seeking when it was high (-4 with .8). Now consider what happens when people make decisions from experience with the same options: The fourfold pattern reverses, suggesting that in decisions from description, people choose *as if* they tend to overweight rare events, whereas in decisions from experience they choose *as if* they underweight rare events; we return to the implied probability weighting shortly.

[Table 1]

*What contributes to the description–experience gap?*

The evidence reviewed above suggests there is a description–experience gap in choice proportions, but what causes it? Several determinants have been examined (Hertwig, 2015); here we briefly review two of the most extensively studied explanations. The first is reliance on small samples. Indeed, Wulff et al. (2018) found that across many thousands of trials, the median sample size was 14 across trials with one safe and one risky option, and 22 across trials with two risky options. Relatively modest sampling effort exacts a price. In about one third of trials, people did not experience at least one of the possible outcomes—typically the rare event. Consequently, people relied, on average, on small samples that caused systematically distorted representations of the true probabilities. Yet sampling error is not the sole determinant of the gap. In an analysis of trials in which the experienced frequencies closely tracked the true probabilities, Wulff et al. (2018, figure 7) nevertheless observed a description–experience gap. Sampling error may thus be sufficient but not necessary for the gap to emerge (see also Ludvig & Spetch, 2011).

A second factor that has received much attention is the weighting of stated and experienced probabilities (i.e., relative frequencies). In a recent sophisticated analysis of the weighting of the *objective* probability of choice options, Regenwetter and Robinson (2017) found strong evidence for a gap: People overweighted rare events in choices from description and underweighted rare events in those based on experience (consistent with early conclusions about the gap; see Barron & Erev, 2003; Hertwig et al., 2004). In experience, however, people do not have access to options' objective probabilities. Therefore, many studies analysed the weighting of the actually experienced probabilities. In an exploratory analysis of the meta-analytical dataset, Wulff et al. (2018; figure 10) found that experienced relative frequencies and stated probabilities prompted different weighting functions—more linear weighting in experience versus overweighting of rare events in description—for choices involving a risky and a safe option, but similar overweighting for choices with two risky options.

[Figure 2]

Going beyond search (i.e., small samples) and probability weighting explanations, researchers concerned with the description–experience gap have proposed several new models to account for experiential choice (e.g., Gonzalez & Dutt, 2011; for a review see Hertwig et al., 2015). Together these models suggest that description- and experience-based choices engage different cognitive processes: Unlike in description, experience-based choices are based on sequential search and updating, and involve rudimentary or no explicit use of probabilities (Camilleri & Newell, 2013; Wulff et al., 2018). Figure 2 offers another illustration of the diverging processes in description and experience. Using the meta-analytical data compiled in Wulff et al., we analysed the extent to which 12 different models of choice predict people’s choices in description and experience. Two results are noteworthy: First, in decisions from experience, a simple heuristic—the natural-mean heuristic (see Supplementary Material)—performs as well as or even better than cumulative prospect theory; a Bayesian model; and two mechanistic explanations specific to decisions from experience, round-wise integration (Hills & Hertwig, 2010; Wulff, Hills, & Hertwig, 2015a), and the ExCon model (Hawkins, Camilleri, Heathcote, Newell, & Brown, 2014). The natural-mean heuristic reaches the same choice as expected value theory (applied to the “experienced” data) would but it does so without any multiplication or explicit representation of probabilities. In decisions from description, in contrast, this heuristic falls far behind. Here, however, another simple heuristic, which weighs all distinct outcomes per gamble equally (the equiprobable heuristic; Supplementary Material), performs nearly as well as cumulative prospect theory (Tversky & Kahneman, 1992).

To conclude, people’s choices and revealed risk preferences are systematically different in description and experience. The description–experience gap has an important implication for risk preference: Depending upon how individuals learn about their options they may employ different cognitive processes, thereby arriving at different decisions. In this

regard, the description–experience gap bears similarity to other, classic format effects such as gain-loss framing (Kahneman & Tversky, 1984) or branch-splitting (Allais, 1953; Birnbaum, 2008), except that the description–experience gap has been found to generalise beyond monetary gambles to, for instance, causal reasoning (e.g., Rehder & Waldmann, 2017), consumer choice (Wulff, Hills & Hertwig, 2015a), or Bayesian reasoning (Armstrong & Spaniol, 2017). The description–experience gap has also been observed in nonhuman primates, thus suggesting that it “does not depend on uniquely human cognitive abilities, such as those associated with language,” and supporting the idea that “epistemic influences on risk attitudes are evolutionarily ancient” (Heilbrunner & Hayden, 2016, p. 593). In any case, format effects demonstrate how difficult it is to estimate general risk preferences from behavioural choices between monetary gambles, most likely because different formats tap into different decision strategies and associated cognitive mechanisms, as figure 2 also illustrates. Do other tools, such as self-report measures and behavioural tasks other than monetary gambles, produce similar levels of dissonance? We turn to this question next.

### **The Behaviour–Self-Report Gap**

Psychologists and economists measure risk preference employing either behavioural or self-report measures, rarely relying on both simultaneously. This need not be a problem if the measures converge toward equivalent conclusions about this latent attribute of a person. But do they? Frey et al. (2017) investigated this crucial question in what is likely the most comprehensive study of risk-preference measures so far, involving 1,507 participants who responded to an extensive battery of self-report *and* behavioural measures of risk preference. The authors implemented a psychometric (bifactor) model to gauge how responses to 39 measures of risk preference are interrelated and estimate how much variance across these could be accounted for by a single general factor of risk preference,  $R$  (for a complete list of the measures see table 1 in Frey et al.). In this model,  $R$  directly accounts for somewhat more

than half of the explained variance across all the measures employed. In addition, self-report measures gauging impulsivity and sensation-seeking also loaded on the *R* factor, highlighting the empirical overlap of operationalisations of risk preference and other such constructs. Frey et al.'s analysis is in line with the robust observation that a small set of traits appears to account for large portions of variance in psychological constructs (Caspi et al., 2014; Deary, 2001). In addition, another set of six domain-specific factors representing risk-taking in domains such as health, finances, and recreation accounted for the remaining explained variance. Overall, this analysis contrasts with the idea of risk preference as a purely general or domain-specific construct and suggests instead that risk preference encompasses both general and domain-specific components (see also Highhouse, Nye, Zhang, & Rada, 2016).

Frey et al.'s (2017) comprehensive psychometric analysis of risk-preference measures produced yet another consequential observation. The general *R* factor did not generalise to the behavioural measures of risk preference and, in fact, did not account for any significant amount of variance in those measures. More generally, the eight behavioural measures investigated failed to converge (correlate) not only with the self-report measures but also with each other, suggesting that disparate behavioural measures do not capture the same construct of risk preference. We refer to the different patterns of convergent validity between self-report and behavioural measures as the *behavioural–self-report gap*. Convergent validity refers to the degree to which measures of a psychological construct capture a common underlying characteristic or trait—something that risk preference measurement traditions, in particular the behavioural measures, seem to lack. Overall, the behavioural–self-report gap raises numerous questions for the future; let us consider just three.

First, what exactly do the behavioural measures capture and what causes the high level of inconsistency between them? One possibility is that the behavioural measures, unlike self-report measures, capture *transient* states rather than a *stable* preference. Their inconsistencies may be a result of them engaging, to different degrees, various decision strategies and

cognitive processes such as memory and learning (Mata et al., 2011, as we discussed in our examination of the description–experience gap. Another possibility is that the “preferences revealed are not independent of the procedure (institution) through which they are revealed” (Berg, Dickhaut & McCabe, 2005, p. 4213), and this contingency on the elicitation procedure may be more pronounced for revealed-preference measures than for stated-preference measures. Finally, it may be that behavioural measures lack the necessary reliability for consistency between measures to emerge. Recent analyses, for instance, have demonstrated that at least 200 choices between monetary lotteries— many more than are usually employed—are required in order to reliably identify whether individuals weigh probabilities linearly (Stahl, 2018; see also Pedroni et al., 2017).

Second, is the extensive empirical overlap between self-reports the product of valid or artefactual (e.g., biased) reporting? Individuals do not have perfect insight into their own personalities and behaviours (Vazire & Carlson, 2010) and some of the consistency across self-report ratings of different measures could be associated with recall biases (e.g., the type of events recalled) or response biases (e.g., how rating scales are used; Schwarz, 1999). Overall, the psychological literature suggests that self-reports are mostly valid, albeit imperfect, measures of behaviour (Haefffel & Howard, 2010; Vazire & Carlson, 2010). However, more work needs to be conducted to assess how some conclusions about the psychometric structure of risk preference (Frey et al., 2017) can be replicated using other measures, such as informant reports or other, objective measures. For example, field studies suggest some but limited consistency across measures of risk preference estimated from different domains (e.g., car and house insurance; see Barsegyahn et al., 2018, for an overview); it would be important to document how the convergent validity of such risk preference indices map onto the convergent validity observed for self-report measures.

A third question is whether this striking behaviour–self-report gap is unique to the measurement of risk preference or if it generalises to other constructs that are also typically

gauged via both behavioural and self-report measures such as self-control and impulsivity (e.g., Duckworth & Kern, 2011) or social preferences (e.g., Galizzi & Navarro-Martinez, 2018). Our reading of the literature suggests that such gaps may be pervasive in the behavioural sciences (e.g., Duckworth & Kern, 2011; Haefel & Howard, 2010) but it would be important to systematically assess how the ontology of risk preference and related constructs differs as a function of measurement choices (Eisenberg et al., 2017).

### **The Temporal Stability Gap**

To the extent that risk preference is considered to be an enduring attribute of a person—tantamount to personality traits (Chuang & Schechter, 2015) or akin to enduring tastes in the classic economic view (Benjamin et al., 2012; Stigler & Becker, 1977)—a pressing issue is whether different measures support the existence of considerable temporal stability of risk preference, or whether they suggest a more capricious attribute that varies substantially with time and resembles more transitory states such as emotions. Systematic variation of these states may, for instance, be a function of the organism's current metabolic needs (as proposed in risk-sensitive foraging theory; McNamara & Houston, 1992; Mishra, 2014; Stephens, 1981) or aspirations. One measure often used to quantify temporal stability is the test–retest reliability of an individual's risk preferences; this makes it possible to quantify the extent to which the same rank-ordering of individuals is preserved across two measurement time points. In a recent meta-analysis of studies reporting test–retest correlations of risk-preference measures, Mata et al. (2018) found substantial temporal divergence of behavioural and self-report measures. For choices between monetary gambles, for which no data were available with retest intervals longer than five years, correlations of about .2 were observed (with considerable variation around this estimate; figure 1A in Mata et al.). For self-report measures, in contrast, the corresponding test–retest correlations were around .5; in addition, this substantially higher level of stability in the self-report measures



appears to not decline further across a 10-year period. These results suggest a third gap between behavioural and self-report measures—this time at the level of temporal stability.

Let us emphasise that the considerable stability of risk preference found for self-report measures by no means precludes intraindividual change across the lifespan. Josef et al. (2016) analysed different notions of stability in a large sample of individuals over a period of 10 years, including *differential stability*—consistency in the rank ordering of individuals' risk preferences over time (as in the meta-analysis by Mata et al., 2018, described above)—and *mean-level stability*—consistency in the respondents' average risk preference over time. Figure 3 plots the findings for these two notions of stability. The results suggest that risk preference—based on a self-report measure—can be understood as a trait with moderate rank-order stability showing correlations of about .5 across measures of up to 10 years (left panel); however, there may be important life span differences in this stability. Specifically, the inverted U-shape pattern in figure 3 suggests that there may be significantly more changes in early adulthood and old age relative to middle age, for example, due to the shifting nature of social roles and challenges during these phases of life. Similarly, there are reliable mean-level differences across the lifespan, with risk-taking preference typically decreasing across adulthood (right panel). Several other results using both cross-sectional and longitudinal analysis support such patterns (Bonsang & Dohmen, 2015; Dohmen et al., 2017).

[Figure 3]

All in all, the temporal-stability gap suggests that self-report and behavioural measures of risk preference show considerable differences in temporal stability, and that self-report measures show important patterns of differential (rank-order) stability and mean-level change that beg investigations of the short- and long-term sources of individual differences in such measures, including biological mechanisms (Benjamin et al., 2012; Kurath & Mata, 2018), cultural origins (Mata, Josef & Hertwig, 2016), and the role of specific life events and momentary challenges (Schildberg-Hörisch, 2018). We look forward to seeing more work

tackling these issues and establishing the convergence or uniqueness of such patterns relative to those for related constructs such as sensation-seeking and impulsivity (Steinberg, 2013).

One obvious but important point is that differential stability is a prerequisite for convergent validity. Only measures that are reliable across time can be expected to converge and be predictive of other relevant constructs or outcomes of interest. Consequently, the temporal stability gap between self-report and behavioural measures suggests that it could be important to develop more reliable behavioural measures before embarking on projects that compare the predictive power of measure types.

### **Conclusions**

To the extent that an individual's risk preference is measurable, one could harness it to, for instance, diagnose and mitigate risk taking that is harmful to the individual and others (e.g., Schonberg et al., 2011; van den Boos & Hertwig, 2017). Yet, despite the risk preference construct's time-honoured pedigree (Bernoulli, 1738/1954) and prominent role in economic models of choice (O'Donoghue & Somerville, 2018; Schoemaker, 1982), it is troubled by unresolved and longstanding conceptual and measurement issues. We have reviewed several of these issues and outlined the challenges ahead.

First, focusing on the classic behavioural measure of choice between monetary gambles, we have highlighted the description–experience gap. It suggests that different forms of learning about and representations of the probabilistic options in an environment (Hertwig et al., 2018) can engage varying cognitive processes that, in turn, give rise to systematically varied choices and implied risk attitudes (Wulff et al., 2018). This work is part of a larger stock of findings according to which different presentation formats trigger specific choice regularities, including systematic preference reversals (Lichtenstein & Slovic, 1971). Second, we have discussed a gap between behavioural and self-report measures of risk preference in terms of their convergent validity. Behavioural measures do not relate to self-report measures

and, to add insult to injury, fail to converge with each other; self-report measures, in contrast, show substantial convergent validity. This means that quite different self-report measures, associated with distinct but overlapping constructs (e.g., risk preference, impulsivity, and sensation-seeking) and domains (e.g., health, finance, and recreation) share large portions of variance. Finally, there is also a gap in the temporal stability of behavioural and self-report measures: Only self-report measures show medium to large rank-order stability across years. This is in stark contrast to the argument—typically advanced by researchers who focused primarily on empirical results from monetary gambles (e.g., Chater, Johansson, & Hall, 2011; Lichtenstein & Slovic, 1971)—that risk preference does not represent a stable and trait-like construct.

Although less is known about the predictive validity of risk-preference measures, self-reported measures appear to show significant predictive validity for notable economic and health outcomes (e.g., teenage pregnancy, drug use; see Hermansson, 2018; Moffitt et al., 2011; cf. Mata et al, 2018 for an overview), thus dispelling the notion of self-reports being little more than “cheap talk.” At the same time, the picture that emerges from studies with revealed (behavioural) preference measures is less promising (e.g., Coppola, 2014; Friedman, Isaac, James, & Sunder, 2014), possibly because they tap into more transient states of risk preference and, furthermore, capture additional cognitive processes such as learning, memory, or numeracy skills (Dohmen et al., 2017; Mata et al., 2018). All in all, it is clear that the field still needs much more work linking different self-report and behavioural measures to field outcomes (cf. Barseghyan et al., 2018).

Theoretical and empirical research on risk preference is more exciting than ever. New questions abound: How and to what degree do the related constructs of, for instance, risk preference and impulsivity, and their measures overlap? Behavioural measures do not appear to be capable of measuring the trait-like characteristics of risk preference, but perhaps they can simulate and predict people’s responses to the specific incentive structure and choice

architecture of a real-world context. If so, conceptualising and testing what is being mustered with each behavioural measure is crucial. And, of course, we need cross-sectional and longitudinal studies that include many of the extant and novel measures in order to quantify their (in)ability to predict important life outcomes such as investment, insurance, and health decisions. Risk preference is a central construct shared across the behavioural sciences but current empirical results present a number of puzzling gaps that must be addressed. We propose that it is important to constructively challenge disciplinary preconceptions and measurement traditions to reveal the still enigmatic persona of risk preference.

### References

- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine [Rational man's behavior in the presence of risk: Critique of the postulates and axioms of the American school]. *Econometrica*, *21*(4), 503–546.
- Armstrong, B., & Spaniol, J. (2017). Experienced probabilities increase understanding of diagnostic test results in younger and older adults. *Medical Decision Making*, *37*(6), 670–679. <http://dx.doi.org/10.1177/0272989x17691954>
- Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*(3), 215–233. <https://doi.org/10.1002/bdm.443>
- Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2018). Estimating risk preferences in the field. *Journal of Economic Literature*, *56*(2), 501–564. <http://doi.org/10.1257/jel.20161148>
- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, *76*(2), 169–217.
- Becker, G. S., & Murphy, K. M. (1988). A theory of rational addiction. *Journal of Political Economy*, *96*(4), 675–700.
- Benjamin, D. J., Cesarini, D., van der Loos, M. J., Dawes, C. T., Koellinger, P. D., Magnusson, P. K., ... & Visscher, P. M. (2012). The genetic architecture of economic and political preferences. *Proceedings of the National Academy of Sciences*, *109*(21), 8026–8031. <https://doi.org/10.1073/pnas.1120666109>
- Berg, J., Dickhaut, J., & McCabe, K. (2005). Risk preference instability across institutions: A dilemma. *Proceedings of the National Academy of Sciences*, *102*(11), 4209–4214. <https://doi.org/10.1073/pnas.0500333102>

- Bernoulli, D. (1738/1954). Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1), 23–36.
- Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). How are preferences revealed? *Journal of Public Economics*, 92(8–9), 1787–94.  
<https://doi.org/10.1016/j.jpubeco.2008.04.010>
- Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, 115(2), 463–501. <http://dx.doi.org/10.1037/0033-295X.115.2.463>
- Blais, A.-R., & Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, 1(1), 33–47.
- Bonsang, E., & Dohmen, T. (2015). Risk attitude and cognitive aging. *Journal of Economic Behavior and Organization*, 112, 112–126. <http://doi.org/10.1016/j.jebo.2015.01.004>
- Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, 5(4), 325–370.  
<https://doi.org/10.1007/BF00122575>
- Camilleri, A. R., & Newell, B. R. (2013). Mind the gap? Description, experience, and the continuum of uncertainty in risky choice. *Progress in Brain Research*, 202, 55–71.  
<https://doi.org/10.1016/B978-0-444-62604-2.00004-6>
- Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., ... & Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2(2), 119–137.  
<https://doi.org/10.1177/2167702613497473>
- Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior and Organization*, 87, 43–51.  
<https://doi.org/10.1016/j.jebo.2012.12.023>
- Chater, N., Johansson, P., & Hall, L. (2011). The non-existence of risk attitude. *Frontiers in Psychology*, 2, 303.

- Chuang, Y., & Schechter, L. (2015). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics*, *117*, 151–170. <http://doi.org/10.1016/j.jdeveco.2015.07.008>
- Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1481), 933–942. <http://doi.org/10.1098/rstb.2007.2098>
- Coppola, M. (2014). Eliciting risk-preferences in socio-economic surveys: How do different measures perform? *Journal of Socio-Economics*, *48*, 1–10. <http://doi.org/10.1016/j.socec.2013.08.010>
- Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex differences in impulsivity: A meta-analysis. *Psychological Bulletin*, *137*(1), 97–130. <http://doi.org/10.1037/a0021591.suppl>
- Deary, I. J. (2001). Human intelligence differences: A recent history. *Trends in Cognitive Sciences*, *5*(3), 127–30. [https://doi.org/10.1016/S1364-6613\(00\)01621-1](https://doi.org/10.1016/S1364-6613(00)01621-1)
- Dohmen, T., Falk, A., Golsteyn, B. H. H., Huffman, D., & Sunde, U. (2017). Risk attitudes across the life course. *The Economic Journal*, *127*(605), F95–F116. <http://doi.org/10.1111/eoj.12322>
- Dohmen, T., Huffman, D., Schupp, J., Falk, A., Sunde, U., & Wagner, G. G. (2011). Individual risk attitudes: measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550.
- Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*(3), 259–268. <https://doi.org/10.1016/j.jrp.2011.02.004>
- Eisenberg, I. W., Bissett, P. G., Canning, J. R., Dallery, J., Enkavi, A. Z., Whitfield-Gabrieli, S., ... Poldrack, R. A. (2017). Applying novel technologies and methods to inform the

- ontology of self-regulation. *Behavior Research and Therapy*, *101*, 46–57.  
<https://doi.org/10.1016/j.brat.2017.09.014>
- Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, *75*(4), 643–669.
- Erev, I., Roth, A. E., Slonim, R. L., & Barron, G. (2002). *Combining a theoretical prediction with experimental evidence to yield a new prediction: An experimental design with a random sample of tasks*. Unpublished manuscript, Columbia University and Faculty of Industrial Engineering and Management, Technion, Haifa, Israel.
- Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances*, *3*(10), e1701381. <http://www.doi.org/10.1126/sciadv.1701381>
- Friedman, D. R., Isaac, R. M., James, D., & Sunder, S. (2014). *Risky curves: On the empirical failure of expected utility*. New York, NY: Routledge.
- Galizzi, M. M., & Navarro-Martinez, D. (2018). On the external validity of social preference games: A systematic lab–field study. *Management Science*. Advance online publication. <https://doi.org/10.1287/mnsc.2017.2908>
- Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). “A 30% chance of rain tomorrow”: How does the public understand probabilistic weather forecasts? *Risk Analysis*, *25*(3), 623–629. <https://doi.org/10.1111/j.1539-6924.2005.00608.x>
- Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological Review*, *118*(4), 523–551.  
<http://dx.doi.org/10.1037/a0024558>
- Hacking, I. (1975). *The emergence of probability*. Cambridge, United Kingdom: Cambridge University Press.



- Harrison, G. W., & Rutström, E. E. (2008). Risk aversion in the laboratory. In J. C. Cox & G. W. Harrison (Eds.), *Risk aversion in experiments, research in experimental economics* (Vol. 12, pp. 41–196). Bingley, United Kingdom: Emerald.  
[http://doi.org/10.1016/S0193-2306\(08\)00003-3](http://doi.org/10.1016/S0193-2306(08)00003-3)
- Haefffel, G. J., & Howard, G. S. (2010). Self-report: Psychology's four-letter word. *American Journal of Psychology*, *123*(2), 181–188.
- Hawkins, G., Camilleri, A., Heathcote, A., Newell, B. R., & Brown, S. D. (2014). Modeling probability knowledge and choice in decisions from experience. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*. Retrieved from <https://pub-jschol-prd.escholarship.org/uc/item/4mv5028f>
- Heilbronner, S. R., & Hayden, B. Y. (2016). The description–experience gap in risky choice in nonhuman primates. *Psychonomic Bulletin & Review*, *23*(2), 593–600.  
<http://doi.org/10.3758/s13423-015-0924-2>
- Hermansson, C. (2018). Can self-assessed financial risk measures explain and predict bank customers' objective financial risk? *Journal of Economic Behavior and Organization*, *148*, 226–240. <https://doi.org/10.1016/j.jebo.2018.02.018>
- Hertwig, R. (2012). The psychology and rationality of decisions from experience. *Synthese*, *187*(1), 269–292. <https://doi.org/10.1007/s11229-011-0024-4>
- Hertwig, R. (2015). Decisions from experience. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (Vol. 1, pp. 239–267). Chichester, United Kingdom: Wiley Blackwell.
- 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*(8), 534–539.  
<https://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*(12), 517–523. <https://doi.org/10.1016/j.tics.2009.09.004>

- Hertwig, R., Hogarth, R. M., & Lejarraga, T. (2018). Experience and description: Exploring two paths to knowledge. *Current Directions in Psychological Science*, *27*(2), 123–128. <https://doi.org/10.1177/0963721417740645>
- Highhouse, S., Nye, C. D., Zhang, D. C., & Rada, T. B. (2016). Structure of the Dospert: Is there evidence for a general risk factor? *Journal of Behavioral Decision Making*, *30*(2), 400–406. <https://doi.org/10.1002/bdm.1953>
- Hills, T. T., & Hertwig, R. (2010). Information search in decisions from experience: Do our patterns of sampling foreshadow our decisions? *Psychological Science*, *21*(12), 1787–1792. <http://doi.org/10.1177/0956797610387443>
- Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review*, *92*(5), 1644–1655.
- Josef, A. K., Richter, D., Samanez-Larkin, G. R., Wagner, G. G., Hertwig, R., & Mata, R. (2016). Stability and change in risk-taking propensity across the adult lifespan. *Journal of Personality and Social Psychology*, *111*(3), 430–450. <http://dx.doi.org/10.1037/pspp0000090>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341–350. [https://doi.org/10.1142/9789814417358\\_0016](https://doi.org/10.1142/9789814417358_0016)
- Kelley, A. E., Schochet, T., & Landry, C. F. (2004). Risk taking and novelty seeking in adolescence: Introduction to part I. *Annals of the New York Academy of Sciences*, *1021*(1), 27–32. <https://doi.org/10.1196/annals.1308.003>
- Kurath, J., & Mata, R. (2018). Individual differences in risk taking and endogenous levels of testosterone, estradiol, and cortisol: A systematic literature search and three independent meta-analyses. *Neuroscience and Biobehavioral Reviews*, *90*, 428–446. <http://doi.org/10.1016/j.neubiorev.2018.05.003>

- Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*(1), 46–55.  
<http://dx.doi.org/10.1037/h0031207>
- Linnér, R. K., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., ...  
Beauchamp, J. P. (2018). Genome-wide study identifies 611 loci associated with risk tolerance and risky behaviors. bioRxiv, 261081. <http://dx.doi.org/10.1101/261081>
- Lopes, L. L., & Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology*, *43*(2), 286–313. <https://doi.org/10.1006/jmps.1999.1259>
- 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.  
<https://doi.org/10.1371/journal.pone.0020262>
- Marinelli, N., & Mazzoli, C. (2011). An insight into suitability practice: Is a standard questionnaire the answer? In P. Molyneux (Ed.), *Bank strategy, governance and ratings* (pp. 217–245). London, United Kingdom: Palgrave Macmillan.  
[https://doi.org/10.1057/9780230313866\\_11](https://doi.org/10.1057/9780230313866_11)
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, *7*(1), 77–91.  
<https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- Mata, R., Frey, R., Richter, D., Schupp, J., & Hertwig, R. (2018). Risk preference: A view from psychology. *Journal of Economic Perspectives*, *32*(2), 155–172.
- Mata, R., Josef, A. K., Samanez-Larkin, G. R., & Hertwig, R. (2011). Age differences in risky choice: A meta-analysis. *Annals of the New York Academy of Sciences*, *1235*, 18–29.  
<https://doi.org/10.1111/j.1749-6632.2011.06200.x>
- Mata, R., Josef, A. K., & Hertwig, R. (2016). Propensity for risk taking across the life span and around the globe. *Psychological Science*, *27*, 231–243.  
[doi:10.1177/0956797615617811](https://doi.org/10.1177/0956797615617811)

- McNamara, J. M., & Houston, A. I. (1992). Risk-sensitive foraging: a review of the theory. *Bulletin of Mathematical Biology*, *54*(2/3), 355–378.
- Mishra, S. (2014). Decision-making under risk: Integrating perspectives from biology, economics, and psychology. *Personality and Social Psychology Review*, *18*(3), 280–307. <https://doi.org/10.1177/1088868314530517>
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 2693–2698. <http://doi.org/10.1073/pnas.1010076108>
- Nigg, J. T. (2016). Attention and impulsivity. In D. Cicchetti (Ed.), *Developmental psychopathology: Vol. 1. Maladaptation and psychopathology* (3<sup>rd</sup> ed., pp. 1–56). Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781119125556.devpsy314>
- O'Donoghue, T., & Somerville, J. (2018). Modeling risk aversion in economics. *Journal of Economic Perspectives*, *32*(2), 91–114. <http://doi.org/10.1257/jep.32.2.91>
- Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour*, *1*(11), 803–809.
- Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, *32*(1–2), 122–136.
- Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making*, *23*(1), 1–14. <https://doi.org/10.1002/bdm.681>
- Regenwetter, M., Robinson, M. M. (2017). The construct–behavior gap in behavioral decision research: A challenge beyond replicability. *Psychological Review*, *124*(5), 533–550. <http://dx.doi.org/10.1037/rev0000067>

- Rehder, B., & Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & Cognition*, *45*(2), 245–260. <http://www.doi.org/10.3758/s13421-016-0662-3>
- Roberts, B. W., Lejuez, C., Krueger, R. F., Richards, J. M., & Hill, P. L. (2014). What is conscientiousness and how can it be assessed? *Developmental Psychology*, *50*(5), 1315–1330. <http://dx.doi.org/10.1037/a0031109>
- Romer, D. (2010). Adolescent risk taking, impulsivity, and brain development: Implications for prevention. *Developmental Psychobiology*, *52*(3), 263–276. <https://doi.org/10.1002/dev.20442>
- Samuelson, P. A. (1938). A note on the pure theory of consumer's behavior. *Economica*, *5*(17), 61–71.
- Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica*, *15*(60), 243–253.
- Schildberg-Hörisch, H. (2018). Are risk preferences stable? *Journal of Economic Perspectives*, *32*(2), 135–154. <http://doi.org/10.1257/jep.32.2.135>
- Schoemaker, P. J. H. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature*, *20*(2), 529–563.
- Schonberg, T., Fox, C. R., & Poldrack, R. A. (2011). Mind the gap: Bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends in Cognitive Sciences*, *15*(1), 11–19. <https://doi.org/10.1016/j.tics.2010.10.002>
- Schwarz, N. (1999). Self-reports: how the questions shape the answers. *American Psychologist*, *54*(2), 93–105.
- Stahl, D. O. (2018). Assessing the forecasting performance of models of choice. *Journal of Behavioral and Experimental Economics*, *73*, 86–92. <https://doi.org/10.1016/j.socec.2018.02.006>

- Stephens, D.W. (1981). The logic of risk-sensitive foraging preferences. *Animal Behaviour*, 29(2), 628–629. [http://dx.doi.org/10.1016/S0003-3472\(81\)80128-5](http://dx.doi.org/10.1016/S0003-3472(81)80128-5)
- Steinberg, L. (2013). The influence of neuroscience on US Supreme Court decisions about adolescents' criminal culpability. *Nature*, 14(7), 513–518.  
<http://doi.org/10.1038/nrn3509>
- Stigler, G. J., & Becker, G. S. (1977). De gustibus non est disputandum. *The American Economic Review*, 67(2), 76–90.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.  
<https://doi.org/10.1007/BF00122574>
- van den Bos, W., & Hertwig, R. (2017). Adolescents display distinctive tolerance to ambiguity and to uncertainty during risky decision making. *Scientific Reports*, 7, 40962.
- Vazire, S., & Carlson, E. N. (2010). Self-knowledge of personality: Do people know themselves? *Social and Personality Psychology Compass*, 4(8), 605–620.  
<http://doi.org/10.1111/j.1751-9004.2010.00280.x>
- Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15(4), 263–290. <https://doi.org/10.1002/bdm.414>
- Wulff, D. U., Hills, T. T., & Hertwig, R. (2015a). Online product reviews and the description–experience gap. *Journal of Behavioral Decision Making*, 28, 214–223.  
<https://doi.org/10.1002/bdm.1841>
- Wulff, D. U., Hills, T. T., & Hertwig, R. (2015b). How short- and long-run aspirations impact search and choice in decisions from experience. *Cognition*, 144, 29–37.  
<https://doi.org/10.1016/j.cognition.2015.07.006>

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*(2), 140–176. <http://dx.doi.org/10.1037/bul0000115>

<http://dx.doi.org/10.1037/bul0000115>

Zuckerman, M. (2008). Sensation seeking. In W. Donsbach (Ed.), *The international encyclopedia of communication*. Hoboken, NJ: Wiley.

<https://doi.org/10.1002/9781405186407.wbiecs029>

Zuckerman, M., Kolin, E. A., Price, L., & Zoob, I. (1964). Development of a sensation-seeking scale. *Journal of Consulting Psychology*, *28*(6), 477–482.

<http://dx.doi.org/10.1037/h0040995>

Table 1

*The Fourfold Pattern of Risk Attitudes in Decisions From Description and its Reversal in Decisions From Experience (based on Hertwig et al., 2004 and Hertwig, 2012)*

Probability	Description		Experience	
	Gain domain	Loss domain	Gain domain	Loss domain
Low	32, .1 <sup>a</sup> vs. 3, 1.0	-32, .1 vs. -3, 1.0	32, .1 vs. 3, 1.0	-32, .1 vs. -3, 1.0
	Rare event: 32, .1	Rare event: -32, .1	Rare event: 32, .1	Rare event: -32, .1
	Risk seeking 48% <sup>b</sup>	Risk averse 36%	Risk averse 20%	Risk seeking 72%
High	4, .8 vs. 3, 1.0	-4, .8 vs. -3, 1.0	4, .8 vs. 3, 1.0	-4, .8 vs. -3, 1.0
	Rare event: 0, .2	Rare event: 0, .2	Rare event: 0, .2	Rare event: 0, .2
	Risk averse 36%	Risk seeking 72%	Risk seeking 88%	Risk averse 44%

*Note.* <sup>a</sup> The alternative outcome (0 otherwise) has been omitted for all risky options. <sup>b</sup> Proportion of risky choices. In past studies, this proportion has been found to be greater than 50% (e.g., Tversky & Kahneman, 1992).



## Figure legends

*Figure 1. The magnitude of the description–experience gap.* The difference in the proportions of choices consistent with discrete underweighting in decisions from experience minus those in decisions from description, from a variety of published and unpublished studies. Results are based on the data compiled in Wulff et al. (2018) and are restricted to studies that aimed to measure the gap rather than study its boundary conditions. The results are shown for choice problems involving a risky and safe option (blue) versus two risky options (grey). The references behind the acronyms can be found in Wulff et al. (2018) or received from the authors of this paper. Adapted from “A meta-analytic review of two modes of learning and the description–experience gap,” by D. U. Wulff, M. Mergenthaler-Canseco, & R. Hertwig, 2018, *Psychological Bulletin*, 144(2), 140–176.

*Figure 2. Predicting decisions from experience and description.* The plot shows the predictive accuracy of 12 choice models tested using choice problems with a risky and a safe option. In decisions from experience, the choice proportion is calculated as a function of three categories of sample sizes (0–10, 11–30, >30). Data include 15,054 experience- and 10,239 description-based choices compiled in Wulff et al.’s (2018) meta-analysis. The model competition includes six heuristics (NM, P, LEX, EP, MIN, MAX), three variants of cumulative prospect theory (CPT<sub>lo</sub>, CPT<sub>tk</sub>, CPT<sub>e</sub>), a Bayesian updating model (BM), round-wise integration (RW), and the exemplar confusion model (ExCon). Details of the models are described in the Supplementary Material.

*Figure 3. Stability and change in risk preference.* The effect of age on (left panel) temporal stability of self-reported risk preference and (right panel) cross-sectional mean-level change (“Are you generally a person who is willing to take risks or do you try to avoid taking risks? Please tick a box on the scale, where the value 0 means not at all willing to take risks and the

value 10 means very willing to take risks.”). In the left panel, the black line represents the weighted average from test–retest correlations between all pairs of measurement waves for males and females combined. In the right panel, red and blue lines represent the results for females and males, respectively. All data stem from the German Socio-Economic Panel and are plotted on a kernel density plot in which darker red indicates a higher density of responses. Adapted from “Stability and change in risk-taking propensity across the adult lifespan,” by A. K. Josef, D. Richter, G. R. Samanez-Larkin, G. G. Wagner, R. Hertwig, & R. & Mata, 2016, *Journal of Personality and Social Psychology*, 111(3), 430–450.

Figure 1

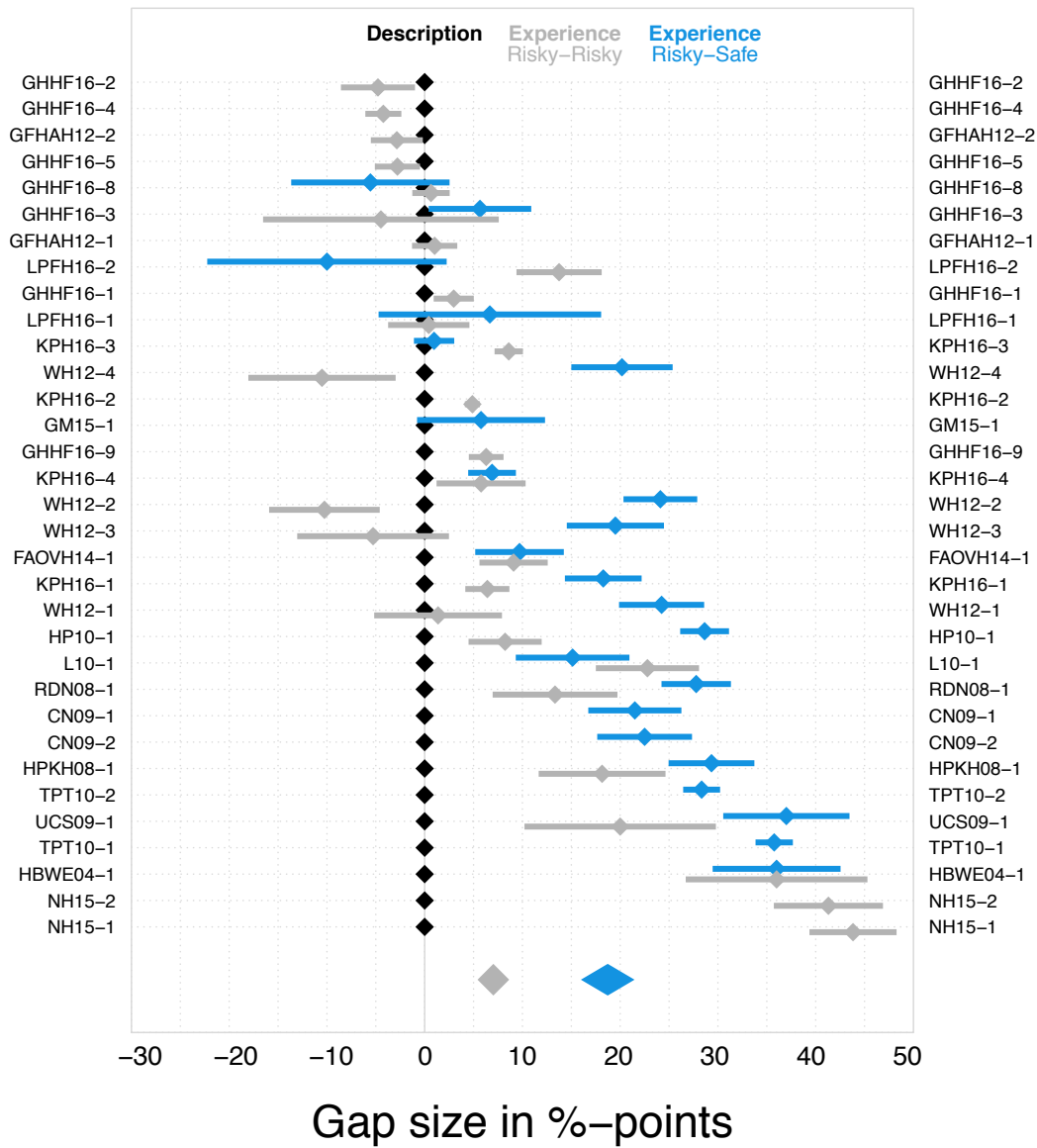


Figure 2

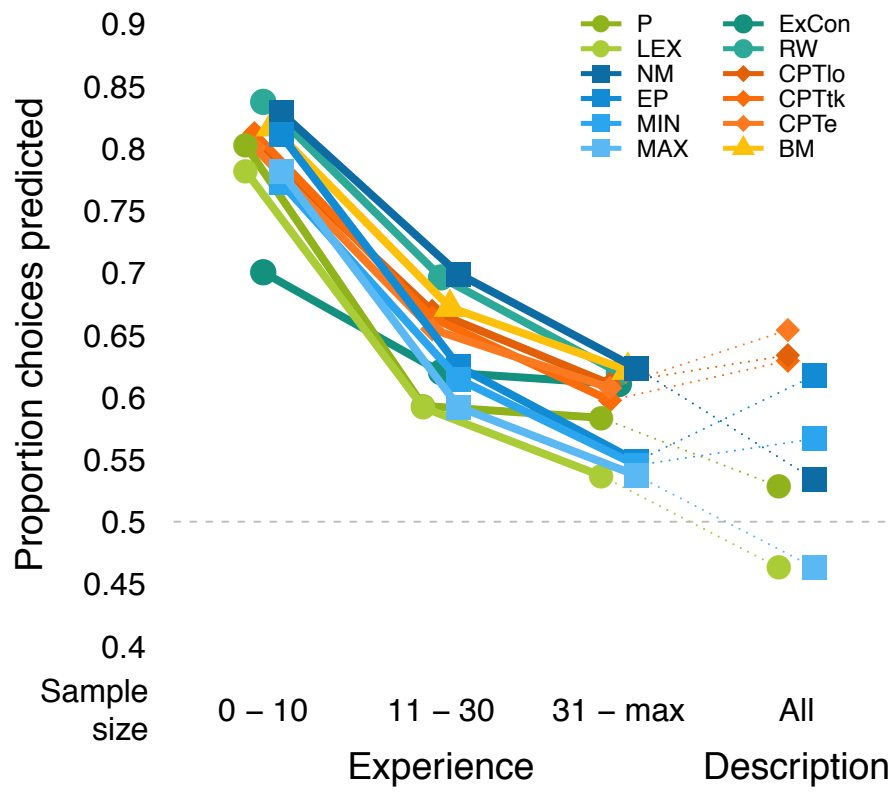
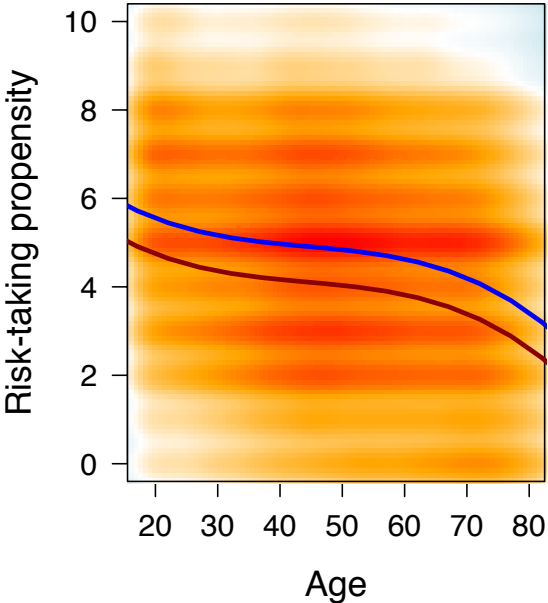
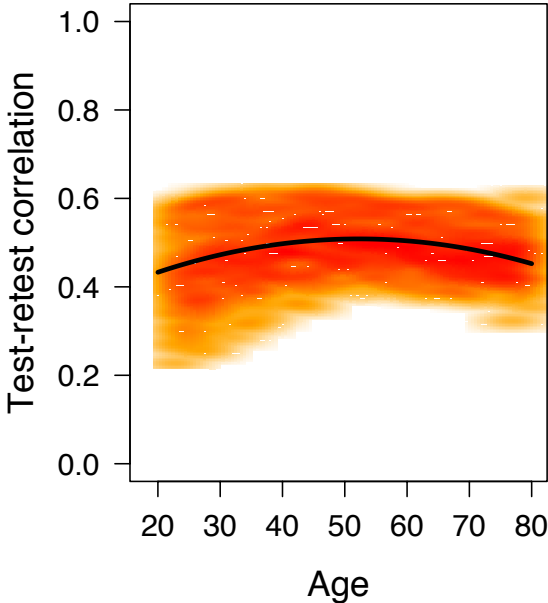


Figure 3



**Supplementary Material: Model competition (see figure 2).**

The heuristics can be separated into two categories: those that use only outcome information and ignore probabilities altogether (outcome heuristics) and those that use at least rudimentary probabilities (dual heuristics). In order to explain the policy of the heuristics, we borrow the token–type distinction from linguistics. In the present context, “types” are all distinct monetary outcomes per gamble. For instance, in the gamble offering -32 with a chance of 10%, the outcomes -32 and 0 are two tokens. Types, in contrast, refer to the actual instantiation of these tokens in a sequence of, say, six draws: e.g., -32, 0, 0, 0, 0, -32.

Heuristics	Decision policy
	<i>Outcome heuristics</i>
Maximax (MAX)	Choose the gamble with the highest maximum type.
Minimax (MIN)	Choose the gamble with the highest minimum type.
Equiprobable (EP)	Calculate the sum of all types within both gambles. Choose the gamble with the highest monetary sum.
Natural Mean (NM)	Calculate the natural mean of experienced tokens in both gambles by summing, separately for each one, all $n$ experienced types and then dividing by $n$ . Choose the gamble with the larger natural mean (i.e., the gamble with the best average outcome in the sampling phase).
	<i>Dual heuristics</i>
Lexicographic (LEX)	Determine the most likely type of each gamble and their respective payoffs. Then select the gamble with the highest, most likely type. If both payoffs are equal, determine the second most likely type of each gamble, and select the gamble

with the highest (second most likely) type. Proceed until a decision is reached.

Probable (P)

Determine the most likely type of each gamble. Then select the gamble with the highest, most likely type. If both types are equal, determine the second most likely type of each gamble, and select the gamble with the highest (second most likely) type. Proceed until a decision is reached.

### *Bayesian mean model (BM)*

The Bayesian mean model chooses the option with the higher average payoff, with outcomes per gamble weighted by the subjective probability as determined by Bayesian updating.

### *Cumulative prospect theory (CPT)*

Cumulative prospect theory (Tversky & Kahneman, 1992) attaches decision weights to cumulated rather than single probabilities. The theory uses five adjustable parameters. Three parameters fit the shape of the value function; the other two fit the shape of the probability weighting function. The value function is

$$\begin{aligned} v(x) &= x^{\alpha^+} \\ v(y) &= -\lambda(-y)^{\alpha^-} \end{aligned} \quad (1)$$

where  $\alpha^+$  and  $\alpha^-$  reflect the sensitivity to differences in positive (gains) and negative (losses) outcomes, respectively. The parameter  $\lambda$  in Equation 1 indicates the relative weight of gains and losses. It is assumed that a loss of a certain magnitude has a greater psychological impact on a choice than does a gain of the same magnitude.

Further, the probability *weighting function* is

$$w^+(p) = \frac{p^{\gamma^+}}{(p^{\gamma^+} + (1-p)^{\gamma^+})^{1/\gamma^+}}, \text{ and}$$

$$w^-(p) = \frac{p^{\gamma^-}}{(p^{\gamma^-} + (1-p)^{\gamma^-})^{1/\gamma^-}}, \quad (2)$$

where the parameters  $\gamma^+$  and  $\gamma^-$  govern the function's curvature in the gain and loss domains, respectively, and indicate how sensitive choices are to differences in probability. Rather than fitting the CPT's parameter to the data (giving it an advantage over the heuristics with no free parameters), we used three sets of parameter estimates from Erev, Roth, Slonim, & Barron (2002); Lopes and Oden (1999); and Tversky and Kahneman (1992). Based on these three sets of parameters, we arrived at three sets of choice predictions.

#### *Round-wise integration (RW)*

Round-wise integration (Hills & Hertwig, 2010; Wulff, Hills, & Hertwig, 2015b) assumes that individuals compare the options across switches in the sequence of samples and tally the number of superior sample subsets for each option. According to RW, decision makers prefer the option with the higher tally.

#### *Exemplar confusion model (ExCon)*

The exemplar confusion model (Hawkins, Camilleri, Heathcote, Newell, & Brown, 2014) assumes that with probability  $(1-p)$  each sample is accurately recorded in option-specific, limitless memory stores. However, with probability  $p$  a confusion process is activated that replaces the current sample with a value drawn at random from the unique values currently present in the option's memory store. According to ExCon, decision makers prefer the option whose memory store has the highest average utility assuming a risk-averse utility function based on Lopes and Oden (1999):



$$U = \frac{1}{N} \sum_{i \in \{1, \dots, N\}} x_i^{551}.$$