Journal of Personality and Social Psychology https://doi.org/10.1037/pspp0000287

# Identifying Robust Correlates of Risk Preference: A Systematic Approach Using Specification Curve Analysis

Renato Frey
University of Basel and Max Planck Institute for Human Development, Berlin

David Richter and Jürgen Schupp
German Institute for Economic Research

Ralph Hertwig
Max Planck Institute for Human Development, Berlin

Rui Mata
University of Basel and Max Planck Institute for Human Development, Berlin

People's risk preferences are thought to be central to many consequential real-life decisions, making it important to identify robust correlates of this construct. Various psychological theories have put forth a series of candidate correlates, yet the strength and robustness of their associations remain unclear due to disparate operationalizations of risk preference and analytic limitations in past research. We addressed these issues with a study involving several operationalizations of risk preference (all collected from each participant in a diverse sample of the German population; N = 916), and by adopting an exhaustive modeling approach—specification curve analysis. Our analyses of six candidate correlates (household income, sex, age, fluid intelligence, crystallized intelligence, years of education) suggest that sex and age have robust and consistent associations with risk preference, whereas the other candidate correlates show weaker and more (domain-) specific associations (except for crystallized intelligence, for which there were no robust associations). The results further demonstrate the important role of construct operationalization when assessing people's risk preferences: self-reported propensity measures picked up various associations with the proposed correlates, but (incentivized) behavioral measures largely failed to do so. In short, the associations between the six candidate correlates and risk preference depend mostly on how risk preference is measured, rather than whether and which control variables are included in the model specifications. The present findings inform several theories that have suggested candidate correlates of risk preference, and illustrate how personality research may profit from exhaustive modeling techniques to improve theory and measurement of essential constructs.

*Keywords:* risk preference | risk taking | correlates | specification curve analysis

*Supplemental materials:* *https://doi.org/10.1037/pspp0000287.supp*

Risk preference is a mainstay construct in the behavioral sciences and of fundamental relevance for outcomes in many life domains (Barseghyan et al., 2013; Brailovskaia et al., 2018; Clark & Lisowski, 2017; Corter & Chen, 2005; Dohmen et al., 2011; Mata et al., 2018; Schonberg et al., 2011; Slovic, 1964). To illustrate, whether to try out recreational drugs in adolescence, pursue a career in an uncertain work environment, invest one's income in a volatile pension fund, or undergo critical surgery in old age: all these decisions depend on whether a person prefers highly variable outcomes that may involve large rewards but also significant losses—key aspects of "risk" (Aven, 2012)—over relatively fixed outcomes of rather moderate magnitude.

Unsurprisingly, the behavioral sciences have long been interested in the correlates of risk preference for both theoretical and applied reasons, starting with the role of wealth in the conceptions of utility and risk in the 18th century (Bernoulli, 1738). Alike, in his seminal review of the psychology and economics literature on risky decision making, Ward Edwards proposed that it would be important for future work on risky choice to ask "To what degree do people differ, and can these differences be correlated with environmental, historical, or personality differences?" (Edwards, 1954, p. 403). Since then, various theoretical accounts have put forth predictions concerning how and why specific variables might be associated with people's risk preferences (see Tab. 1 for an overview). Yet, as our brief review below will illustrate, empirical tests of these theories have often produced mixed evidence, rendering it unclear whether variables such as a person's economic situation, sex, age, or cognitive ability are indeed robust correlates of risk preference. Arguably, this inconclusive state of affairs has originated primarily from the following two issues.

**The issue of disparate measurement approaches.** In psychological research, people's actual risk-taking behaviors are rarely assessed in real life—such as in observational studies, which tend to be laborious. Although sometimes people are prompted to self-report their current and past risk-taking behaviors by means of "frequency measures" (e.g., "During the past six months, how often did you drive while being drunk?", cf. Frey et al., 2017), research in psychology and economics has typically assumed that how people deal with risks and uncertainty may largely depend on their *risk preferences* (depending on the subdiscipline, this construct is also referred to as *risk attitude*, *risk-taking propensity*, or as *risk aversion* when framed inversely). For example, in psychology people's risk preferences are thought to drive their engagement in activities that involve rewards but also potential losses, including physical or mental harm (e.g., substance use). Similarly, in economics risk preference is considered to account for whether people prefer variable options that involve (monetary) gains and losses over relatively certain outcomes (cf. Mata et al., 2018). Hence, two main measurement approaches—with numerous associated measures—have been developed to gauge people's risk preferences (Appelt et al., 2011; Frey et al., 2017):

Proponents of the *stated preference approach* rely on self-reports of people's general or domain-specific attitudes towards risk taking (thus, these measures are often called "propensity measures"; e.g., "Are you generally a person who is willing to take risks or do you try to avoid risks?"; Blais & Weber, 2006; Frey et al., 2017; Hanoch et al., 2006; Linnér et al., 2019; Weber et al., 2002). It has been argued that latent attitudes are capable of accounting for substantial variance in actual behavior (cf. *theory of planned behavior;* Ajzen, 1991), and self-reported propensity measures have repeatedly been found to have a series of desirable psychometric properties, such as high convergent validity and test–retest reliability (e.g., Frey et al., 2017; Mata et al., 2018; Steiner & Frey, 2020). A potential reason underlying these observations might be that self-reports of risk-taking propensity are well rooted in people's idiosyncratic past experiences (Steiner et al., 2019).

Yet, self-reported propensity measures might not be entirely free of implicit or explicit reporting-biases, which is partly why proponents of the *revealed preference approach* have proposed eliciting people's risk preferences by means of behavioral tasks (cf. Appelt et al., 2011; Dohmen et al., 2011; Frey et al., 2017; Samuelson, 1938; Schonberg et al., 2011). In principle, such behavioral measures permit inferring people's preferences from observable behavior, and these tasks often involve monetary incentives based on the assumption that such incentives will render the observed behavior more consequential—and thus more representative of risky behaviors in real life. Although a promising idea, behavioral measures have shown some empirical limitations, such as poor convergent validity and weak test–retest reliability (cf. Frey et al., 2017; Mata et al., 2018).

Importantly, it has previously been illustrated that different measures of risk preference (across, but also within measurement approaches) may lead to substantially different conclusions concerning a person's risk preference (e.g., Frey et al., 2017; Slovic, 1964). It is thus problematic that past work has often tested the links between specific candidate correlates and risk preference with single operationalizations, rendering it unclear how generalizable previously found associations are across different measurement approaches.

**The issue of focusing on single theories and correlates.** Even though various theories have proposed multiple candidate correlates of risk preference, most empirical investigations have focused on only one or a small subset of these correlates. In other words, past research has not analyzed large sets of different model specifications involving potentially competing predictor variables. The lack of such analyses within the same datasets renders it difficult not only to compare the relative associations of various candidate correlates with risk preference, but also to uncover potential confounds. For example, to what extent are estimates of age differences in risk preference confounded by differences in income or education that may vary systematically across age groups?

Unfortunately, robustness checks are the exception rather than the rule in the psychological literature on risk preference, and if existent, the number of predictor variables and resulting model specifications has typically been small (e.g., Dohmen et al., 2011; Josef et al., 2016; Mamerow et al., 2016). Either the lack of data or computational limitations might be reasons why thorough robustness checks are not routinely conducted. Furthermore, even robustness checks based on only a few model specifications, and in particu-

Renato Frey, Department of Psychology, University of Basel, and Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin https://orcid.org/0000-0002-3190-3523; David Richter and Jürgen Schupp, German Institute for Economic Research, Berlin; Ralph Hertwig, Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin; Rui Mata, Department of Psychology, University of Basel, and Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin.

We are grateful to Laura Wiles for editing the manuscript. The full dataset and all analysis scripts are available from https://osf.io/b2fej/.

Corresponding author: Renato Frey, Center for Cognitive and Decision Sciences, Department of Psychology, University of Basel, Missionsstrasse 64A, CH-4055 Basel, Switzerland. E-mail: renato.frey@unibas.ch.

**Table 1**

*Theoretical Accounts With Predicted Associations Between Candidate Correlates and Risk Preference*

| Theoretical accounts | Groups of candidate correlates | | | |
|---|---|---|---|---|
| | Income and wealth | Sex (female) | Age | Cognitive ability and education |
| Bernoullian expected utility theory (Bernoulli, 1783) | + | | | |
| Financial-cushioning hypothesis (cf. Dohmen, Falk, Huffman, Sunde, Schupp, Wagner, 2011) | + | | | |
| Risk-sensitivity theory (Stephens, 1981; Mishra, Gregson, & Lalumière, 2011) | - | | - | |
| Gender schema theory (cf. Slovic, 1966; Bem, 1981) | | - | | |
| Parental investment theory (Trivers, 1972) | | - | | |
| Evolutionary signaling hypothesis (cf. Wilson & Daly, 1985; Baker & Maner, 2009) | | - | - | |
| Socio-emotional selectivity theory (Carstensen, Pasupathi, Mayr, & Nesselroade, 2000; Depping & Freund, 2011) | | | - | |
| Social-investment theory (Bleidorn et al., 2013; Roberts, Wood, & Smith, 2005) | | | - | |
| Dopaminergic neuromodulation hypothesis (Düzel et al. 2010) | | | - | |
| Brain maturation / cognitive control hypothesis (Figner & Weber, 2011; Steinberg, 2013; Dohmen, Falk, Huffman, & Sunde, 2018) | | | - | - |
| Ontogenetic co-development hypothesis (cf. Dohmen, Falk, Huffman, & Sunde, 2018) | | | | + |
| Phylogenetic co-evolution hypothesis (cf. Dohmen, Falk, Huffman, & Sunde, 2018) | | | | + |
| Confound hypothesis (Mata et al., 2011; Olschewski, Rieskamp, & Scheibehenne, 2018) | | | + / - | + / - |

Note. Associations predicted to be positive vs. negative are depicted by + and -, respectively. A concurrent + / - indicates that associations are predicted to vary as a function of specific situations or task characteristics.

lar classic approaches to reducing the number of predictors (e.g., stepwise procedures), are known to give misleading results(Raftery, 1995).

## Theories and candidate correlates of risk preference

### *Income and wealth*

**Theory.** Historically speaking, the first accounts theorizing about possible correlates of risk preference focused on the role of a person's economic situation. For example, in expected utility theory (Bernoulli, 1738) "risk aversion" is driven by the fact that one's marginal utility decreases with increasing wealth (i.e., the utility curve is thought to reflect how much an increase of one monetary unit triggers in terms of the associated psychological utility; in this framework, the curvature of the utility function directly reflects the degree of risk aversion). Even though in this framework wealthy and poor people may in principle be described by the exact same concave utility curve—that is, the same degree of general "risk aversion"—they will be located at different positions on the curve, implying different propensities to take or avoid a risk. Specifically, people with larger initial wealth will be located more to the right of the curve, implying that these people will be less willing to pay a premium to avoid a risky choice option. In colloquial terms, wealthy people should thus shy away relatively less from taking a risk, as opposed to poor people (cf. Eeckhoudt et al., 2005). This prediction is in line with the hypothesis that wealth bolsters people against the consequences of a potential setback when taking risks. According to this financial-cushioning hypothesis, it is wealthier people's reserves that reduce their (perceived) financial vulnerability, permitting them to become risk-seeking in the first place (cf. Dohmen et al., 2011). To the extent that the effects of a person's economic situation may extend across domains, this mechanism may also apply beyond the financial domain: for instance, larger income allows buying better health insurance, which in turn may permit taking greater health risks.

Yet, one may derive the opposite prediction based on the view of risk preference as highly need- and state-based: risk-sensitivity theory (Mishra, 2014; Mishra et al., 2011; Stephens, 1981) predicts that people will generally behave risk-averse when being above a critical threshold (i.e., given sufficiently high income and wealth), but start to become risk-seeking when falling below that threshold.

**Empirical evidence.** In line with the various theoretical assumptions predicting that people in a less solid economic situation tend to be risk-seeking, one study relying on a propensity measure of general risk preference reported that in countries with greater hardship (i.e., lower incomes and wealth) people tended to have an increased risk preference (Mata et al., 2016). However, another study using a similar propensity measure of general risk preference found that wealthier people reported a higher risk preference (Dohmen et al., 2011). Yet another study that investigated whether changes in income are associated with changes in risk preference (also relying on self-reported propensity measures) did not observe any effects across general or domain-specific risk preference (Josef et al., 2016). That is, past empirical tests have provided quite mixed evidence concerning the links between a person's economic situation and risk preference.

### *Sex*

**Theory.** To date, a person's sex remains one of the most-frequently theorized candidate correlates of risk preference. On the one hand, some theories have proposed cultural contributions to sex differences in risk preference, by pointing out the pervasive norms and beliefs about how "men should, and do, take greater risks than women" (Slovic, 1966, p. 169), particularly in young age (cf. Moffitt, 1993). These hypotheses were rooted in gender schema theory, which provides a cognitive account of how "sex typing" results from a person's self-concept getting assimilated to gender stereotypes (Bem, 1981).

On the other hand, according to evolutionary theories, reduced levels of risk preference in females might be associated with the smaller potential reproductive rate of females compared to males, and with differential parental investment costs when accounting for specific reproductive strategies (Mishra, 2014; Trivers, 1972). Relatedly, it has been argued that, particularly in men, risk taking serves as a signaling device that permits demonstrating fitness to potential partners (Baker & Maner, 2009; Frankenhuis et al., 2010).

**Empirical evidence.** In line with the relatively univocal predictions stemming from the various theoretical perspectives, there exist a number of meta-analyses that provide support for the idea that men are more risk-seeking than women (Byrnes et al., 1999; Cross et al., 2011; Cross et al., 2013). Yet, the studies considered in these meta-analyses were very heterogeneous in many respects. For example, Byrnes et al. (1999) analyzed 150 studies, which they coded as "hypothetical choice tasks" (N = 39; e.g., hypothetical choice dilemmas), "self-reported behavior" (N = 66; e.g., concerning drinking and drugs), and "observed behavior tasks" (N = 48; e.g., gambling tasks but also skill-based tasks such tossing rings onto pegs). These tasks were further subdivided in multiple content categories. Moreover, the studies were coded to reflect five age categories ranging from "3 to 9 years" to "older than 21 years". It is thus not surprising that sex differences varied considerably across studies and measures, suggesting that age and domain, as well as possibly further variables, may be moderators of sex differences.

To illustrate, sex differences may be domain-specific because they are mediated by the expected enjoyment of specific outcomes (Harris et al., 2006), opportunity sets (Schubert et al., 1999), and knowledge (Dwyer et al., 2002), which can differ systematically between men and women in various domains (e.g., work, social relations). One should note that although these past findings suggest heterogeneity of sex differences across domains, there is a lack of theoretical predictions that would lead to concrete expectations about such domain-specific patterns. In sum, the empirical evidence suggests that whereas sex differences in risk preference are pervasive, they may vary substantially as a function of measures, domains, and other relevant covariates.

### *Age*

**Theory.** There are a number of theoretical accounts that suggest reductions in risk preference across the adult lifespan caused by age-related changes in the need for resource accumulation, motivation, social roles, as well as biological and cognitive factors. For example, the view that risk taking reflects a functional adaption to maximize reproductive success (Mishra, 2014) suggests that risk preference might be elevated in adolescence and young adulthood, a period during which it is crucial to gain access to resources and mating partners. Conversely, in line with risk-sensitivity theory one may argue that older adults have accumulated substantial resources and are thus no longer in a state that requires taking risks (Mishra et al., 2011; Stephens, 1981).

According to motivational theories, older age is increasingly associated with a focus on positive emotions (Carstensen et al., 2000) or the avoidance of losses (Depping & Freund, 2011), with such motivational reorientations resulting in a smaller appetite for highly variable outcomes (i.e., risky options) that are associated with negative emotions and potential losses.

Other theories emphasize the importance of social roles, for example, social-investment theory (Bleidorn et al., 2013; Roberts et al., 2005) suggests that reductions in risk preference across the life span could stem from normative life transitions to adult roles (getting a job, marrying, having children) and associated systematic changes in personality (e.g., higher agreeableness and conscientiousness).

From a more biological perspective, it has been theorized that age-related reductions in dopaminergic function lead to reduced exploration and novelty seeking (Düzel et al., 2010), thus to potentially declining risk preference. Other theories have focused on brain maturation underlying cognitive control to explain differences in risk preferences between adolescents and adults, particularly in "hot" situations involving high emotional arousal or social influence (Figner & Weber, 2011; Steinberg, 2013).

Finally, some have suggested a role for cognitive ability and interaction with task complexity that moderate age differences in risky choice (Frey, Mata, et al., 2015; Mata et al., 2011). Age-related cognitive decline may lead to information processing limitations that reduce learning and lead to the use of simpler, less cognitively-demanding strategies, which will affect choices under risk and uncertainty particularly in situations requiring learning or integration of large amounts of information.

**Empirical evidence.** Studies implementing self-reported propensity measures of general and domain-specific risk preference provide substantial evidence for an increase in risk preference in adolescence and early adulthood (Steinberg, 2013), followed by a decline across the adult lifespan (Dohmen et al., 2017; Josef et al., 2016; Mata et al., 2016)—albeit with potential variations across domains (Josef et al., 2016). Importantly, evidence suggests that these patterns are quasi-universal in that they can be detected across countries (Mata et al., 2016) and in both cross-sectional and longitudinal studies (Josef et al., 2016; Rolison et al., 2014).

In turn, two meta-analyses focusing on adolescence (Defoe et al., 2015) and adult development (Mata et al., 2011) suggest that different behavioral measures produce inconsistent results regarding age differences in risk preference. This points to the issue that age differences assessed through behavioral tasks might be moderated by specific task characteristics, including whether measures tap into fluid intelligence (Frey, Mata, et al., 2015; Mata et al., 2011; Rolison et al., 2012). For example, research on decisions from experience suggests that age differences emerge particularly when many choice options have to be explored (Frey, Mata, et al., 2015). We discuss the role of cognitive ability as a potential correlate of risk preference *and* as a confound of age in more detail in the next section.

### *Cognitive ability and education*

**Theory.** According to Dohmen et al. (2018), there are various hypotheses concerning how cognitive ability may be related to risk preference. First, the *cognitive control hypothesis* predicts a negative association between cognitive ability and risk preference by assuming that people with higher cognitive ability are better able to regulate their behavior, and thus avoid impulsive risk-taking behaviors. This is related to the brain maturation hypothesis presented above that has been advanced to predict heightened risk preference in adolescents because they lack the necessary cognitive control (Steinberg, 2013).

However, Dohmen et al. (2018) also suggest other accounts that predict positive associations between cognitive ability and risk preference: the *ontogenetic co-development hypothesis* suggests that risk-seeking persons could select into particular environments that foster cognitive abilities (e.g., applying at a prestigious school with a low admission rate); the *phylogenetic co-evolution hypothesis* suggests that evolution could favor specific combinations of cognitive ability and risk preference, namely, that reduced risk preference may be adaptive for people with lower cognitive ability.

Furthermore, whether it pays to take risks typically depends on the statistical properties of the choice environment, such as the magnitude of potential rewards and losses and their associated probabilities (cf. Pleskac & Hertwig, 2014). Risk taking may thus also depend on a person's ability to understand or learn about specific risk–reward structures and reason about them in an expedient fashion (Dohmen et al., 2018; Frey, Mata, et al., 2015; Mata et al., 2018). Consequently, depending on particular task characteristics cognitive ability may systematically lead to risk-seeking or risk-averse behavior—according to such a *confound hypothesis*, cognitive ability may not relate directly to qualitative changes in people's preferences but rather lead to spurious associations due to increased choice inconsistency (i.e., noisier choices; cf. Olschewski et al., 2018) and poorer learning (Mata et al., 2011).

**Empirical evidence.** Some of the past empirical research has found that people with higher cognitive ability were willing to take more financial risks compared to people with lower cognitive ability, based on self-reported propensity measures as well as behavioral measures (e.g., Boyle et al., 2011; Dohmen et al., 2010). Education, which is conceptually and empirically linked to cognitive ability, was also found to be positively associated with higher levels of risk preference, as captured with a self-reported propensity measure (Mata et al., 2016). Yet, a recent meta-analysis investigating the role of cognitive ability in behavioral measures found a weak positive association with risk preference in the gain domain, but there were no such associations in the loss domain or in gambles with mixed outcomes (Lilleholt, 2019). Finally, general intelligence was found to be related to engaging in fewer risky behaviors in the health domain, such as smoking and alcohol consumption (Batty et al., 2007).

Past work further suggests that cognitive ability may play a differential role not only as a function of domain (e.g., gains vs. losses, or finance vs. health) but also depending on other measurement characteristics. Specifically, the various measures of risk preference vary substantially in their complexity, such as the degree of feedback available (Frey, Rieskamp, et al., 2015; Pleskac, 2008; Yechiam & Busemeyer, 2006) or the number of available choice options (Frey, Mata, et al., 2015; Hills et al., 2013), which may represent additional cognitive challenges (Dohmen et al., 2018; Mata et al., 2018). As previewed above, in a risky decision task requiring learning from experience, inter-individual differences in fluid intelligence were associated with differences in exploratory behavior in decisions involving many choice options (Frey, Mata, et al., 2015).

In sum, several open questions concerning the associations between cognitive ability and risk preference persist, including the links between different dimensions of cognitive ability (e.g., fluid vs. crystallized intelligence) and different measures of risk preference. As a result, it is important to ask whether fluid intelligence ($g_f$), crystallized intelligence ($g_c$), and (years of) education—all of which are conceptually and empirically related—are associated similarly to a person's risk preference, as for example assessed by means of self-reported propensity measures, static monetary lotteries, or lotteries requiring repeated learning from experience.

## Current study: Towards a general framework for testing correlates of risk preference

The relatively mixed patterns of results concerning the associations between several candidate correlates and risk preference, as observed in past empirical research, might have emerged largely due to the two reasons outlined above: a lack of systematic operationalization of the construct of risk preference, as well as a lack of simultaneously considering multiple competing predictor variables.

Our approach in this article promises to overcome these limitations by making two central contributions: first, we assembled a comprehensive dataset involving multiple measures of risk preference completed by a large and diverse sample of the German population; second, we implemented an exhaustive modeling technique—specification curve analysis—to systematically assess the robustness of the candidate correlates reviewed above.

**A novel dataset: The risk preference module in the SOEP Innovation Sample.** In order to obtain a large sample of persons who vary substantially in terms of both their risk preferences and the potential correlates thereof, we made use of comprehensive panel data. Specifically, in the German Socio-Economic Panel ("SOEP") self-reported information on general and domain-specific risk preference is routinely collected (Dohmen et al., 2011; Josef et al., 2016). However, as in most similar panels there exist no or only limited data for behavioral measures of risk preference. To complement the original panel data we were granted access to include a risk module in the "Innovation Sample" of the SOEP, which consisted of two major implementations of behavioral risk-preference assessments. The behavioral measures involved substantial monetary incentives of up to 68 EUR per choice and—according to the arguments of the revealed preference approach—should thus be able to elicit people's actual risk preferences.

The different versions of the behavioral elicitation methods varied concerning how information was presented (decisions from description vs. decisions from experience; Hertwig & Erev, 2009), as well as in terms of the number of choice options (two vs. four). These manipulations can be helpful in understanding to what extent different presentation formats and cognitive demands contribute to capturing individual differences in risk preference, including those due to age and cognitive ability (e.g., Frey, Mata, et al., 2015).

The behavioral module was completed by 951 participants, resulting in a large and diverse dataset consisting of (a) a series of self-report measures of risk preference, (b) a series of (incentivized) behavioral measures of risk preference, and (c) an array of person indicators (i.e., the candidate correlates of risk preference, such as household income, sex, age, or cognitive ability).

**A novel analytic approach: Specification curve analysis.** Identifying which of several candidate correlates show robust associations with a variable of interest is a basic analytic problem that is by no means unique to personality research, and a number of solutions to this problem have hence been proposed: for example, approaches such as extreme-bound analysis in econometrics aim to estimate the distribution of coefficients across a subset or the full set of possible specifications, which results from a given set of explanatory variables (Granger & Uhlig, 1990; Leamer, 1983; Sala-i-Martin, 1997).

In the current study we implement specification curve analysis (SCA; Simonsohn et al., 2015), a relatively novel and very promising method. The core idea behind SCA is to exhaustively implement all model specifications that result from (a) systematically varying the operationalization of the dependent variable (i.e., different operationalizations of risk preference), as well as (b) considering all combinations of a set of potentially important independent variables (i.e., the candidate correlates put forth by the various theoretical accounts and, potentially, other covariates). The association between a candidate correlate (e.g., sex) and the construct of interest (e.g., risk preference) can then be visualized across all possible model specifications by means of a "specification curve": that is, the coefficients obtained from all implemented model specifications are ordered by their effect size and plotted accordingly. This empirical specification curve can easily be compared with the distribution of coefficients that is to be expected under the null hypothesis (i.e., that a candidate correlate has no association with risk preference), thus permitting "inference-by-eye". Moreover, a straightforward quantitative estimate can be obtained for the probability that an observed association between a correlate and the construct of interest has merely resulted by chance. SCAs have already provided important insights in the behavioral sciences, such as regarding the role of birth order on several aspects of personality (including risk preference; Lejarraga et al., 2019; Rohrer et al., 2017), or concerning the role of digital technology use in adolescents' well-being (Orben & Przybylski, 2019).

## Summary and goals

The main aim of this article is to identify robust correlates of risk preference. In doing so, we focus on the putative associations of the various candidate correlates with risk preference, as predicted by the reviewed theories. That is, these associations form the combined hypothesis space that we examine empirically using a systematic "meta-test". To this end we rely on a novel dataset comprising various operationalizations of risk preference, and employ a comprehensive modeling technique to systematically examine different model specifications (in terms of the operationalization of the dependent variable, as well as in terms of concurrent independent variables).

This exhaustive and systematic modeling approach is in line with the growing efforts of increasing transparency and reproducibility in the behavioral sciences (see also "multiverse" or "forking path" analyses; Baribault et al., 2018; Steegen et al., 2016; Wacker, 2017). Specifically, exhaustive modeling approaches as implemented in the current study, and multi-model inference more generally (Calcagno & de Mazancourt, 2010; Hoeting et al., 1999), have several advantages over classic analyses due to a reduction in "researchers' degrees of freedom" (Babyak, 2004) and less room for selective reporting (Simonsohn et al., 2015).

## Methods

### Sample characteristics, incentives, and ethical approval

A total of 951 participants of the SOEP Innovation Sample had previously provided self-report data and subsequently also completed the session with the behavioral tasks (in the presence of an interviewer, yet without direct observation). Technical errors in the behavioral tasks prevented a clean matching of the data for 35 participants, resulting in a total sample size of 916. Socio-demographic information for these participants is provided in Table S1. The size of this sample promises sufficient variability in terms of independent and dependent variables. Participants received a performance-contingent bonus depending on two randomly selected decisions (out of eight decisions) in the behavioral tasks. This bonus ranged from 0 to 72 EUR (with a mean of 6.1 EUR). The institutional review board of the Max Planck Institute for Human Development, Berlin, approved the study.

### Independent variables (IVs)

We selected the independent variables based on two criteria. First, the main IVs represented the candidate correlates identified by the theories reviewed above. Second, we also included some additional variables that were available in the SOEP dataset, based on their sheer-face validity. That is, for these covariates no deeper theoretical motivation has been proposed in the literature (unlike for the main candidate correlates investigated here). The distributions of all nine independent variables are shown in Figure S1, and Figure S2 depicts their intercorrelations (see supplemental materials for additional information on the independent variables). All continuous IVs were standardized prior to the modeling analysis.

#### *Selection of candidate correlates*

To represent the four main groups of candidate correlates that we identified above (i.e., based on the review of previous theoretical accounts putting forth potential candidate correlates of risk preference), we included the following variables from the SOEP dataset: first, to reflect a person's economic situation we relied on "household income". Second and third, we included a person's sex and age[1]. Fourth, to reflect a person's cognitive ability and education, we relied on measures of fluid intelligence (obtained from a digit-symbol substitution task) and crystallized intelligence (obtained from a word-lists tasks), as well as on the total number of years of education a person had completed.

#### *Selection of covariates*

As additional covariates, we also included employment status and data on time use in the social domain and in sports. Even though none of the previous theories predicts specific associations of these variables with risk preference, we included these covariates because people's risk preferences in the occupational, social, or recreational domains could emerge partly due to the mere opportunity to take any risks in these settings. For example, people's risk preferences in the occupational domain may differ between employed and unemployed persons, because only the former have an opportunity to take such risks in the first place. We thus included these variables as covariates yet not as explicit candidate correlates of risk preference.

### Dependent variables (DVs)

To obtain a broad set of different operationalizations of risk preference, we relied on all self-reported propensity measures of risk preference available in the SOEP. Additionally, we implemented a range of behavioral measures in the risk module, and also extracted a series of different summary measures. The distributions of all 17 dependent variables are shown in Figure S1, and Figure S3 depicts their intercorrelations. All continuous DVs were standardized prior to the modeling analysis.

#### *Propensity measures*

As outlined above, self-report information on risk-taking propensity is routinely collected in the SOEP. Specifically, all participants reported their willingness to take risks "in general", as well as in each of the following six domains: driving, investment, recreational, occupational, health, and social (see supplemental materials for the exact wording). Participants expressed their risk-taking propensity on a scale ranging from 0 to 10.

#### *Behavioral measures*

All participants completed two behavioral tasks (in randomized order): decisions from description (DFD), in which the monetary outcomes and their associated probabilities of each choice option were explicitly stated (i.e., a traditional lottery task), and decisions from experience (DFE), in which participants had to learn about the outcomes and their probabilities by means of sampling with replacement. We included

[1]For age there might be a curvilinear relationship with risk preference, due to an increase in risk preference in adolescence and early adulthood, followed by a continuous decline. We thus ran a pre-analysis implementing a quadratic effect of age, but found that such a model increased the model fit only marginally (see Fig. S4).

these two tasks and different choice set sizes (two vs. four choice options) because previous research has suggested that these potentially pick up differential associations with certain candidate correlates. To illustrate, in DFD participants' risk perceptions may directly depend on the readily available information, whereas in DFE—like in many risky real-life decisions—additional attentional or cognitive effort has to be invested to get an accurate perception of the risks involved in the available choice options. In DFE, exploration of the choice options was free (i.e., the sampled payoffs did not count towards the final bonus) and only the last choice between the choice options resulted in a final draw that counted for the bonus. In both tasks, participants played two trials with two choice options and two trials with four choice options (resulting in a total number of eight trials). Details of the decision problems are provided in Table S2, and screenshots of the choice tasks are presented in Figure S5.

As main dependent variables we considered the proportions of risky choices (defined as a choice of the option with the objectively higher variance; in trials with four options: one of the two choice options with higher variance), separately for DFD and DFE, and separately for the two choice set sizes. In the DFE task, we also included the mean (inverted) "sample size" per participant as a tentative operationalization of risk preference (cf. van den Bos & Hertwig, 2017), separately for the two different choice set sizes. Specifically, the more a person explores before making a final choice (i.e., a larger sample size), the better a person's risk perception should be aligned with the objective risks involved in the choice problems. Thus, extensive exploration may render the final choice less ambiguous and the final payoff will thus not necessarily come at a total surprise. For more information on the implemented measures see https://www.diw.de/en/diw_01.c.511801.en/soep_is_innovative_modules.html and https://renatofrey.net/demos/ for an online demonstration (in German).

### *Summary measures*

Previous research has suggested that the construct of risk preference is best modeled by summarizing multiple measures with both a broad, general factor ($R$), as well as domain-specific components (akin to the general factor of intelligence and its various "facets"; Frey et al., 2017; Highhouse et al., 2016). These findings suggest that testing candidate correlates of risk preference with aggregate measures thereof may lead to more robust conclusions, as opposed to merely relying on a single measure of risk preference. To this end we employed two approaches to aggregate the individual measures.

First, we implemented a psychometric model that makes specific assumptions about the structure of the psychological construct of risk preference. Specifically, we estimated a bifactor model according to which a general factor of risk preference ($R$) extracts the common variance across all measures, whereas several orthogonal factors account for common variance that is specific to certain domains or types of measures. The factors of this model were identified in a preceding exploratory factor analysis with bifactor rotation. In the confirmatory factor analysis, only variables that loaded at least .2 on any of the factors were included. For this modeling analysis, missing data points were imputed by means of Gibbs sampling using the R-package *mice* (van Buuren & Groothuis-Oudshoorn, 2011). As in previous research (Frey et al., 2017), this model revealed that a general factor captured common variance mostly across the different propensity measures but not across behavioral measures (Fig. S6).

Second, we implemented three statistical models to extract aggregate indicators of the two main types of operationalizations of risk preference (self-reported propensity measures vs. behavioral measures), as well as their combination. The rationale for including these statistical aggregates was to assess whether the two measurement approaches as a whole differ systematically in their links to candidate correlates, which could have important implications for future work that aims to use reliable indicators of risk preference. Specifically, we estimated three linear models with random intercepts for participants yet no other independent variables, and as DV we used participants' risk preferences as measured by the various DVs introduced above. That is, each participant's intercept was permitted to vary and thus reflects his or her mean-level risk preference across the implemented measures of risk preference. In the first of these three statistical models, we included all operationalizations of risk preference ("Statist. mod.: All'). In the second statistical model ("Statist. mod.: Prop."), we only included the self-reported propensity measures to estimate a person's risk preference, and finally, in the third statistical model ("Statist. mod.: Beh.") we only included the behavioral measures to estimate a person's risk preference.

## Composition of model specifications

For each of the six candidate correlates of risk preference we implemented the model specifications as follows: first, we generated all possible combinations of the nine IVs (note that we specified only linear additive effects and no interaction terms; see discussion). As each IV can either be in- or excluded in each specification, this amounted to $2^9 = 512$ combinations. Second, for the analysis of each of the six candidate correlates we only retained those specifications in which the current candidate correlate (e.g., "household income" in the first of the six SCAs) was always included as an IV. By definition, this is the case in half of the combinations, that is, in 256 specifications. Third and finally, these 256 specifications were paired with all 17 mutually exclusive operationalizations of the dependent variable "risk preference" (e.g., see Fig. 1). In sum, this resulted in $256 * 17 = 4,352$ model specifications per candidate correlate.

## Estimation methods: Bayesian regressions vs. OLS regressions

We first implemented all model specifications using Bayesian estimation techniques, which in principle have two main advantages over traditional ordinary least-squares (OLS) methods: first, implementing weakly informative priors for the effects to be estimated provides some statistical regularization and thus guards against overfitting the data. Second, the 95% highest density-intervals (HDI; obtained from the estimated effects' posterior distributions) can be interpreted intuitively—as opposed to confidence intervals obtained from frequentist approaches, which tend to be robustly misinterpreted (Hoekstra et al., 2014). We ran the models using the R-package *rstanarm* (Stan Development Team, 2016) and adopted the weakly informative default priors: $\mathcal{N}(0,10)$ for the intercept and $\mathcal{N}(0,2.5)$ for the independent variables. Three chains with 2000 iterations were run per model.

Implementing a large number of Bayesian regression models can be computationally very intensive (i.e., in particular for the simulation analyses to generate the null distribution described below). We thus also implemented all models using a traditional OLS approach, and as can be seen in Figures 1-6 (overlaid black-and-white lines), the correlations between the coefficients obtained from the two different estimation methods were very high (average $r > .99$). Furthermore, the number of credible (Bayesian estimation) and significant (OLS estimation) effects correlated with $r = .91$ across the six candidate correlates. We thus relied on the more parsimonious OLS estimation approach for the simulation analyses described below.

## Simulation analyses to generate the null distributions

We conducted a series of simulation analyses to examine whether the empirical (i.e., observed) specification curves deviated systematically from the null distribution of effect sizes (i.e., the expected distribution of false-positive effect sizes assuming that there exists no systematic association between a candidate correlate and risk preference).

First, for a given candidate correlate (e.g., "household income") and each of the 4,352 model specifications (see above), we estimated the empirical effect size of the main IV, that is, the current candidate correlate (e.g., "household income") using the formula $y = b_0+b_1*x_1+b_2*x_2+...+b_n*x_n$, where the DV, $y$, was the operationalization of risk preference considered in the current model specification (e.g., "self-reported general risk preference"), $b_0$ was the intercept, $b_1$ was the estimated effect of the main IV $x_1$ (e.g., "household income"), and $x_2$ to $x_n$ corresponded to any other IVs that were part of the current model specification (e.g., age, sex, fluid intelligence).

Second, to simulate that the current candidate correlate has no unique effect on the implemented DV (i.e., the null hypothesis), we generated a modified DV, $y_{null}$, by subtracting the estimated effect of $x_1$ on $y$, such that $y_{null} = y - b_1 * x_1$. Thus, $y_{null}$ is merely a function of systematic variance related to the other IVs present in the current model specification (i.e., $x_2$ to $x_n$ as well as residual variance, but no variance uniquely related to $x_1$.

Third, for each model specification we ran 50 case-bootstrapping iterations, during which we sampled data with replacement from the original dataset. In each of these bootstrapped samples, we estimated the effect of $x_1$ on $y_{null}$ using exactly the same formula as introduced above, except replacing $y$ with $y_{null}$. Thus, the formula was $y_{null} = b_0+b_1*x_1+b_2*x_2 + ... + b_n * x_n$. As any unique effect of $x_1$ on $y_{null}$ has been removed in the previous step, all estimates of $b_1$ that differ from 0 are false-positive estimates that reflect either residual variance or variance related to the other IVs.

The resulting distributions of effect sizes (i.e., under the null hypothesis) are shown in Figures 1-6, depicted as gray band in the background of the empirical specification curve. The upper bound of the gray band depicts the 97.5th percentile of the effect sizes that were estimated under the null hypothesis (i.e., across the 50 simulation runs). Conversely, the lower bound of the gray band depicts the 2.5th percentile of these effect sizes. The estimates forming the upper and lower bounds were ordered separately from each other. In short, the gray band indicates the range of false-positive effect sizes to be expected under the null hypothesis and when implementing the 4,352 model specifications. Most importantly, the boundaries of the gray band indicate how strongly the main IVs of interest (i.e., the candidate correlate) is estimated to be related to the DV even if its unique relationship has been removed. That is, if some specifications have a high upper bound in the gray area, this implies that the candidate correlate was erroneously estimated to be related with the DV, for example, because it was correlated with other IVs (and started to capture the respective variance).

Based on the simulated null distributions of effect sizes, we computed exact (i.e., nonparametric) p-values for observing a specification curve under the null hypothesis with at least as many significant coefficients as the specification curve observed empirically. In other words, these p-values express the probability of observing a specification curve with a particular number of significant effects, knowing that there is no true association between a candidate correlate and risk preference (i.e., a false-positive result). These exact p-values were obtained by (a) dividing the number of simulation samples with a greater proportion of significant effects than in the empirical sample by the number of simulation runs (i.e., 50) and then (b) dividing the resulting quotient by 2 (Rohrer et al., 2017; Simonsohn et al., 2015). The number of significant effects in the bootstrapped samples was very con-

**Table 2**
*Results of the Specification Curve Analyses*

| | Number of specifications | Median posterior effect size across all specifications | Number (proportion) of credible positive effects | Number (proportion) of credible negative effects | Bootstrap samples with larger proportion of significant effects | Exact p-value of bootstrap test |
|---|---|---|---|---|---|---|
| Household income | 4352 | 0 | 783 (18.0%) | 272 (6.2%) | 1 / 50 | 0.01 |
| Sex (female) | 4352 | -0.14 | 0 (0.0%) | 2201 (50.6%) | 1 / 50 | 0.01 |
| Age | 4352 | -0.09 | 134 (3.1%) | 2229 (51.2%) | 1 / 50 | 0.01 |
| Fluid intelligence | 4352 | 0.01 | 1131 (26.0%) | 745 (17.1%) | 1 / 50 | 0.01 |
| Cryst. intelligence | 4352 | -0.04 | 0 (0.0%) | 1155 (26.5%) | 33 / 50 | 0.33 |
| Years of education | 4352 | 0.01 | 798 (18.3%) | 847 (19.5%) | 1 / 50 | 0.01 |

sistent across the simulation iterations (average SD = 0.04), indicating that 50 simulation runs were sufficient to obtain robust results.

### Open research practices

The full dataset and all analysis scripts are available from https://osf.io/b2fej/.

## Results

### General findings

As Table 2 illustrates, the number of specifications with credible effects differed substantially across the six candidate correlates. Across the 4,352 model specifications, sex and age resulted in the largest total number of credible effects (N = 2201 and N = 2363, respectively). For the other candidate correlates, only less than half of the specifications resulted in credible effects: there were 1876 credible effects for fluid intelligence, 1645 credible effects for years of education, 1155 credible effects for crystallized intelligence, and 1055 credible effects for household income.

These patterns can also be seen in the *specification curve panels* of Figures 1-6, where sex and age clearly have the largest proportions of credible effects, depicted by green (positive effects) and orange (negative effects) 95% highest density-intervals (HDIs) that exclude 0. Conversely, the other correlates have a substantially larger proportion of noncredible effects, depicted by the blue 95% HDIs that include 0.

The simulation analyses indicated that for all candidate correlates except for crystallized intelligence, the probability of observing the respective numbers of effects under the null hypothesis was very small ($p$ = 0.01). For crystallized intelligence in 33 out of 50 simulation runs—each implementing a null distribution of effects using a bootstrap sample—the number of significant effects[2] was larger than the number observed in the empirical specification curve. This means that under the null hypothesis there was a probability of .33 for observing a specification curve as extreme as the one observed empirically, suggesting that the estimated effects for crystallized intelligence likely reflect false-positives (type-I errors).

When ignoring patterns unique to certain model specifications (see next section below) and aggregating across all specifications, the effect of sex on risk preference was strongest (being female equated to a reduction of 0.14 standard deviations [SDs] in risk preference), followed by the effect of age (an increase of one SD in age equated to a reduction of 0.09 SDs in risk preference), crystallized intelligence (an increase of one SD in crystallized intelligence equated to a reduction of 0.04 SDs in risk preference), fluid intelligence (an increase of one SD in fluid intelligence equated to an increase of 0.01 SDs in risk preference), years of education (an increase of one SD in years of education equated to an increase of 0.01 SDs in risk preference), and household income (an increase of one SD in household income equated to a change of smaller than 0.00 SDs in risk preference; see Tab. 2 for all summary statistics).

Again, these patterns can also be observed in Figures 1-6, where sex and age tend to have the specification curves with the most pronounced effects. Conversely, the specification curve for the candidate correlate with the smallest median effect size across all specifications (i.e., household income) had hardly any elevation. Note, however, that the candidate correlates with median effect sizes close to 0 (i.e., household income, fluid intelligence, and years of education) are not necessarily flat. Rather, some of the model specifications produced credible positive effects, whereas other model specifications produced credible negative effects. In the next section, we turn in detail to such patterns that are distinct to particular model specifications.

### Variability across specifications

The red tick marks in the *specification panels* of Figures 1-6 depict the different model specifications: for each model specification, the tick marks indicate (a) which independent variables (IVs) were used as predictors and (b) which dependent variable was implemented to operationalize risk preference. In the extreme case and assuming that it is irrelevant whether and which additional independent variables are considered for estimating the effect of a candidate correlate on risk preference, as well as that the various operationalizations of risk preference all result in the same estimate of a candidate correlate's effect, the red tick marks should be

[2]"Significant" is used here because for efficiency reasons, the simulation analyses were conducted using OLS estimation; see Methods section.

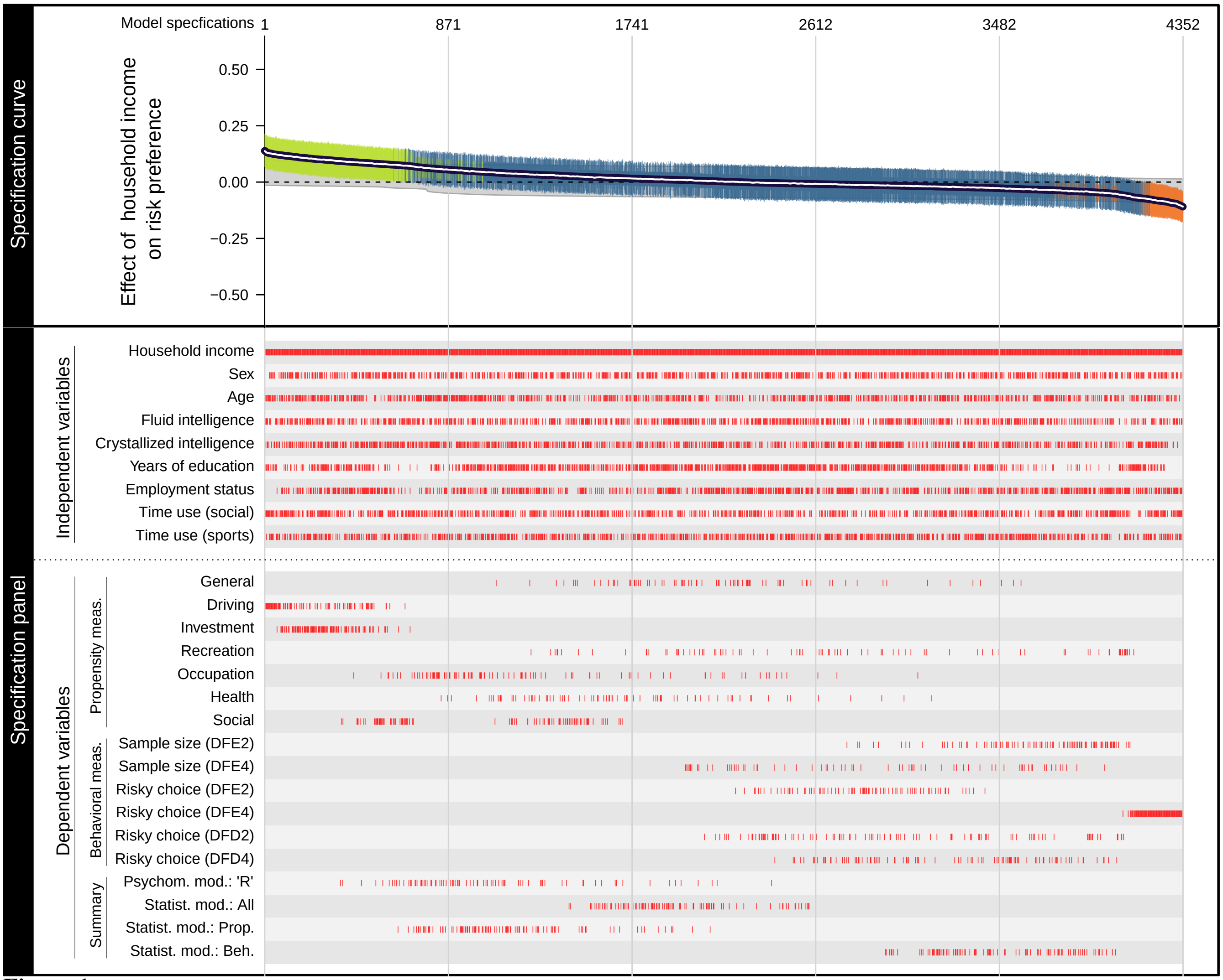


**Figure 1**

*Specification curve analysis for the association of* household income *with risk preference. The* ***specification curve panel*** *shows the association of the candidate correlate (i.e., household income) with risk preference across all 4,352 implemented model specifications (sorted by effect size). The black line depicts the effect sizes of household income obtained from Bayesian estimation methods (i.e., means of the posterior distributions), and the thinner white line the effect sizes obtained from ordinary least-squares regressions (p. 9). Continuous variables were z-transformed; the coefficients thus represent changes in terms of standard deviations (i.e., standardized effects). The colored vertical lines in the specification curve depict 95% highest-density intervals (green for credible positive effects, blue for effects not credibly different from 0, and orange for credible negative effects). The gray band in the background depicts the distribution of effect sizes to be expected under the null hypothesis (i.e., the false-positive estimates to be expected even when any systematic association between the candidate correlate and risk preference has been removed; p. 9). The* ***specification panel*** *depicts which DV and which IVs were used in each of the model specifications: a red tick mark indicates that the variable of the current row was included in the specification of the tick mark's position on the x-axis. A single specification may have any combination of the 9 IVs, but the main IV of interest (in this case: household income) is always included (resulting in a steady row of red tick marks for the main IV). Furthermore, each specification implements exactly one of the 17 different DVs (p. 8). To illustrate, specification 1 includes household income, age, fluid intelligence, years of education, and time use (social) as IVs, and driving as DV.* ***Abbreviations:*** *"DFD" = decisions from description. "DFE" = decisions from experience. The numbers behind DFD and DFE stand for the number of options participants needed to choose between (p. 7). "Psychom. mod.: 'R"' = The general factor* R *extracted from a psychometric model. "Statist. mod.: All" = A statistical model summarizing all measures. "Statist. mod.: Prop." = A statistical model summarizing the propensity measures. "Statist. mod.: Beh." = A statistical model summarizing the behavioral measures. See p. 8 for a description of these summary indicators.*

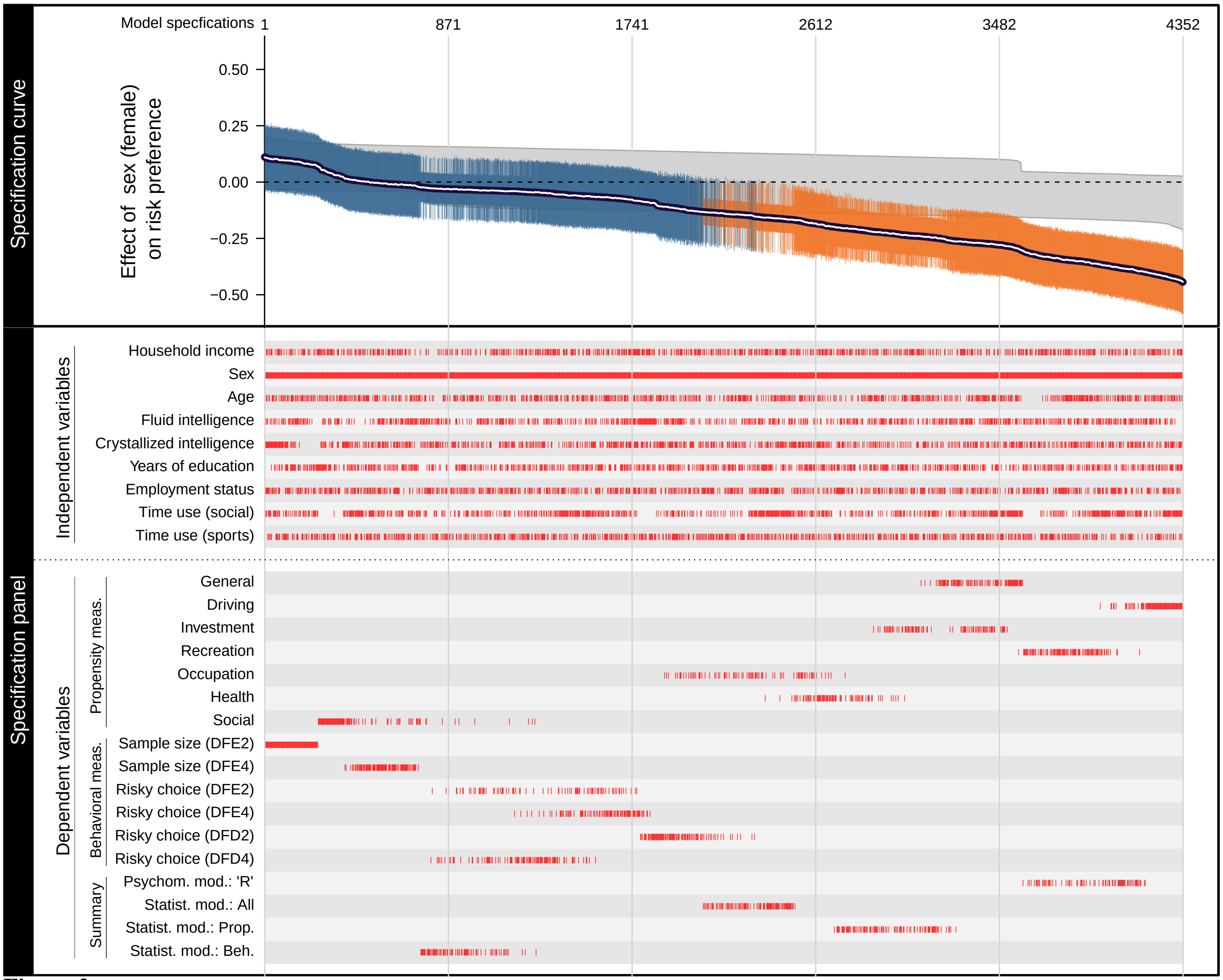


**Figure 2**

*Specification curve analysis for the association of* sex *with risk preference. See Figure 1 for a detailed description of how to interpret the different elements in this figure.*

randomly distributed on the horizontal axis (i.e., across the various model specifications) and there should be no evident clusters of tick marks. Furthermore, in this extreme case, the specification curve should be a flat horizontal line; that is, all model specifications should produce the exact same estimate. Evidently, this was not the case for any of the candidate correlates.

As a general pattern across all candidate correlates, the tick marks in the specification panels reflecting the different IVs were relatively more dispersed, as opposed to the tick marks reflecting the different DVs. This suggests that all else being equal, whether a positive or negative effect was estimated (or, for the candidate correlates with almost only negative estimates: whether a weak or a strong negative effect was estimated) depended less on whether or which additional independent variables were included in the model specifications, but rather on how risk preference was operationalized. This pattern was particularly pronounced for the candidate correlate sex (Fig. 2), where there were marked clusters for the specifications with different DVs, whereas with only few exceptions there were no such clusters for the different IVs. In what follows, we discuss the most prominent clusters for each candidate correlate.

### *Household income*

As Figure 1 shows, there was a series of credible positive effects for household income, and the specification panel suggests that these effects were observed when risk preference was measured in a domain-specific way, namely, concerning investment and driving. The first finding in particu-

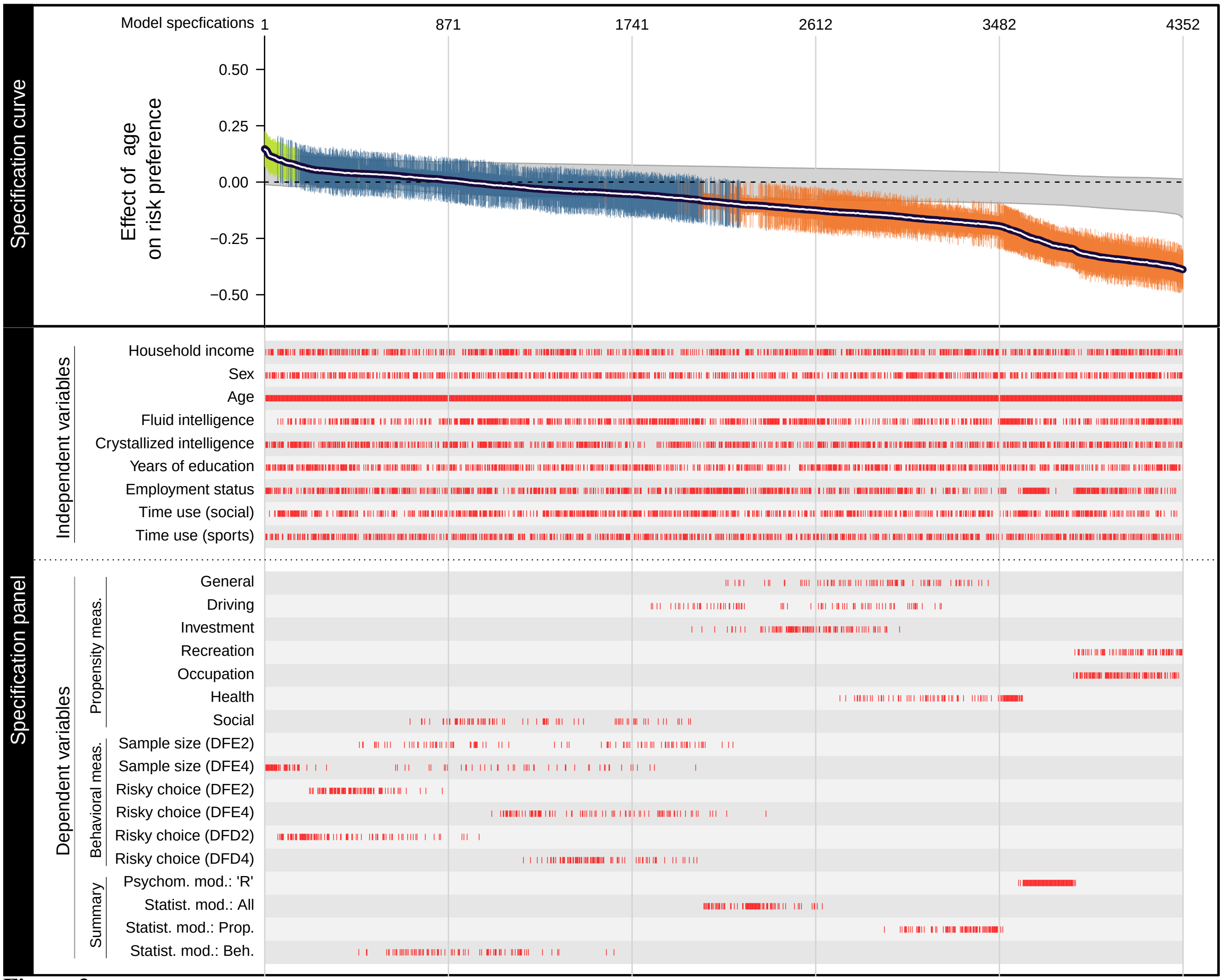


**Figure 3**

*Specification curve analysis for the association of* age *with risk preference. See Figure 1 for a detailed description of how to interpret the different elements in this figure.*

lar is in line with the theoretical perspective that larger wealth increases people's risk preference, based on the rationale that wealthier people are "cushioned" in case of a setback when taking (financial) risks. The second observation does not directly speak to any of the reviewed theories. Yet, a more speculative interpretation of this association is that it reflects a confound in the operationalization of risk preference in the driving context, because it is mostly wealthier persons who may be able to afford expensive sport cars that may encourage them to take risks.

At the same time, in line with the prediction that wealthier people are more risk averse, there were a number of credible negative effects of household income. One should note that this effect was almost exclusively observed for a single behavioral DV (i.e., the proportion of risky choices in decisions from experience with four choice options) and the observed effects were relatively weak, suggesting that this finding needs to be interpreted cautiously.

### *Sex*

In line with the various theoretical predictions, all of the credible effects of sex (female) on risk preference were negative (Fig. 2) yet the strength of this effect varied substantially as a function of the implemented operationalization of risk preference. Specifically, being female was associated with the strongest effect on risk preference in the driving context, followed by risk preference in the recreational and investment domains. Being female was also associated relatively strongly with a reduction in general risk preference (as captured with the general factor *R* extracted from the psychome-

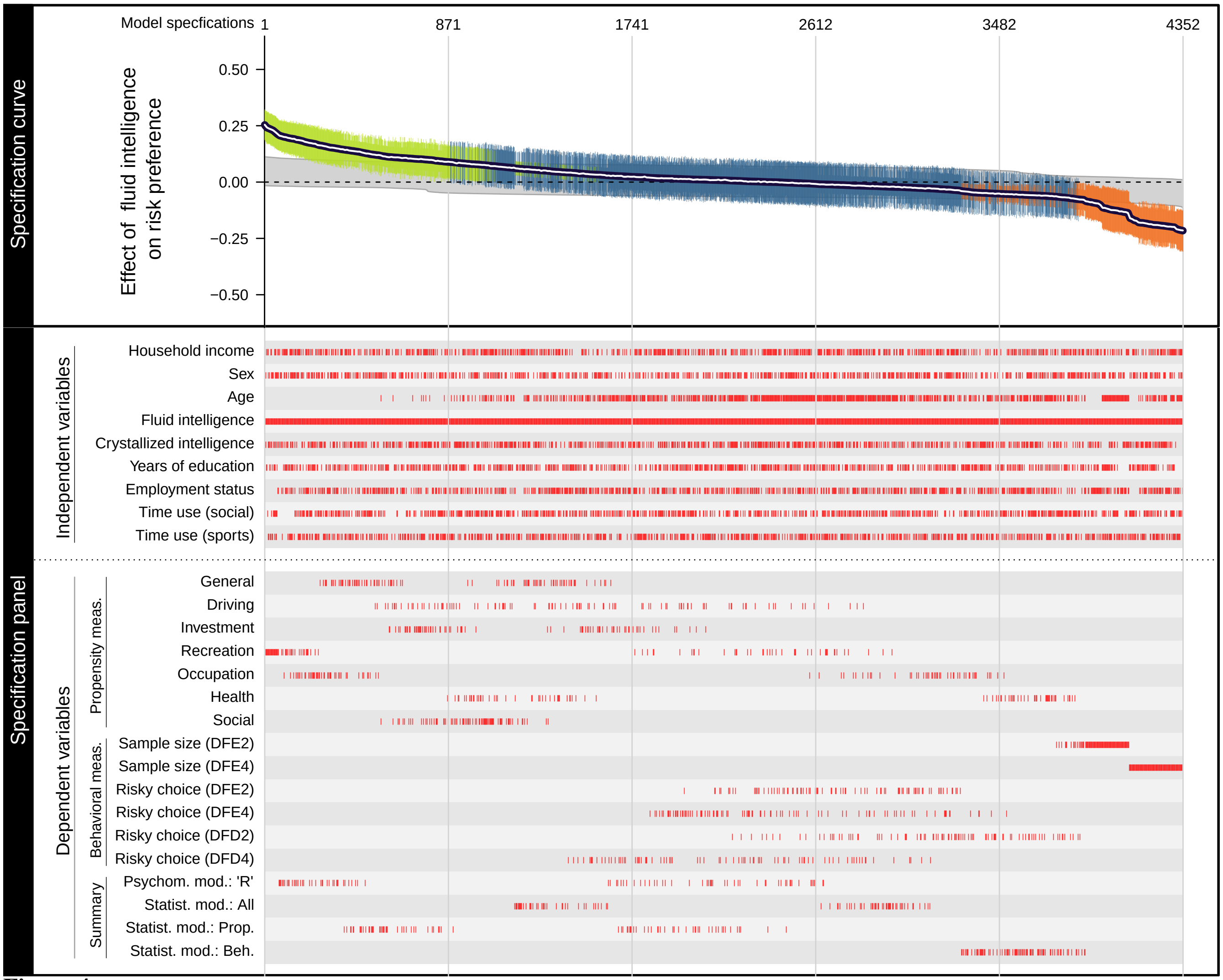


**Figure 4**

*Specification curve analysis for the association of* fluid intelligence *with risk preference. See Figure 1 for a detailed description of how to interpret the different elements in this figure.*

tric model; as well as by the propensity measure capturing general risk preference and by the statistical model summarizing all propensity measures). Finally, there were also negative associations with risk preference in the health and occupational domains. These findings largely corroborate the various cultural and evolutionary theories on the role of sex in people's risk preferences. Yet, when risk preference was measured in the social domain or by any of the various behavioral measures, no credible effects emerged.

### *Age*

The association of age with risk preference (Fig. 3) was strongest concerning occupational and recreational risk preference, followed by general risk preference as measured by the psychometric model. Older age was also associated with a reduced risk preference concerning health and investment, as well as in the driving context. These findings corroborate the various theories focusing on socio-emotional and biological contributions to changes in risk preference across the adult lifespan.

However, like for sex, there were again no credible associations between age and risk preference when the latter was assessed by the behavioral measures, with one exception: when risk preference was operationalized as the amount of pre-decisional exploration in the DFE task with four choice options, older adults searched less extensively before making a final choice. These effects only emerged when not controlling for fluid intelligence, thus illustrating a potential confound and speaking to the view that reductions in fluid intelligence, which occur across the adult lifespan, may relate

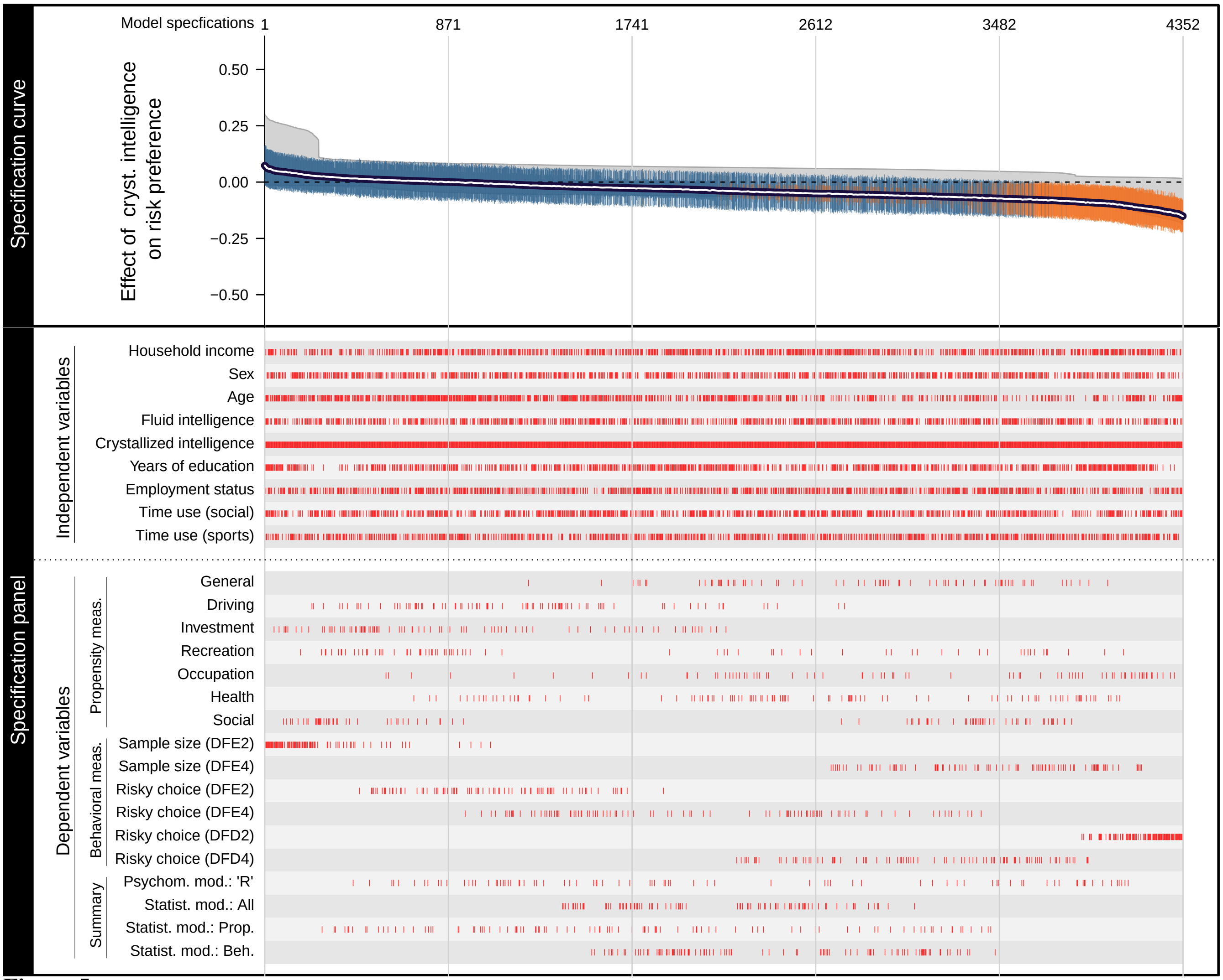


**Figure 5**

*Specification curve analysis for the association of* crystallized intelligence *with risk preference. See Figure 1 for a detailed description of how to interpret the different elements in this figure.*

to more stochastic and noisier choice behavior in cognitively demanding situations.

### *Fluid intelligence*

Fluid intelligence was a clear example for how the in-/exclusion of specific independent variables may systematically affect the estimated associations with risk preference: the relatively large proportion of specifications with credible positive effects of fluid intelligence on risk preference occurred almost exclusively when age was not included as a covariate (Fig. 4). In other words, these specifications evidently picked up the almost identical age effects as described in the section above.

There were also two marked clusters with credible effects for specific operationalizations of risk preference: when risk preference was measured in terms of pre-decisional information search in the DFE task, people with higher fluid intelligence explored substantially more, which arguably indicates a reduced risk preference (i.e., the final choice will be made with more complete knowledge about the payoff distribution as opposed to when the choice options are explored only briefly). Strikingly, this effect was picked up even more strongly in the choice options with four options, where higher fluid intelligence may be particularly instrumental in helping decision-makers process the larger amounts of information.

### *Crystallized intelligence*

There were a number of specifications with credible effects suggesting that higher crystallized intelligence is asso-

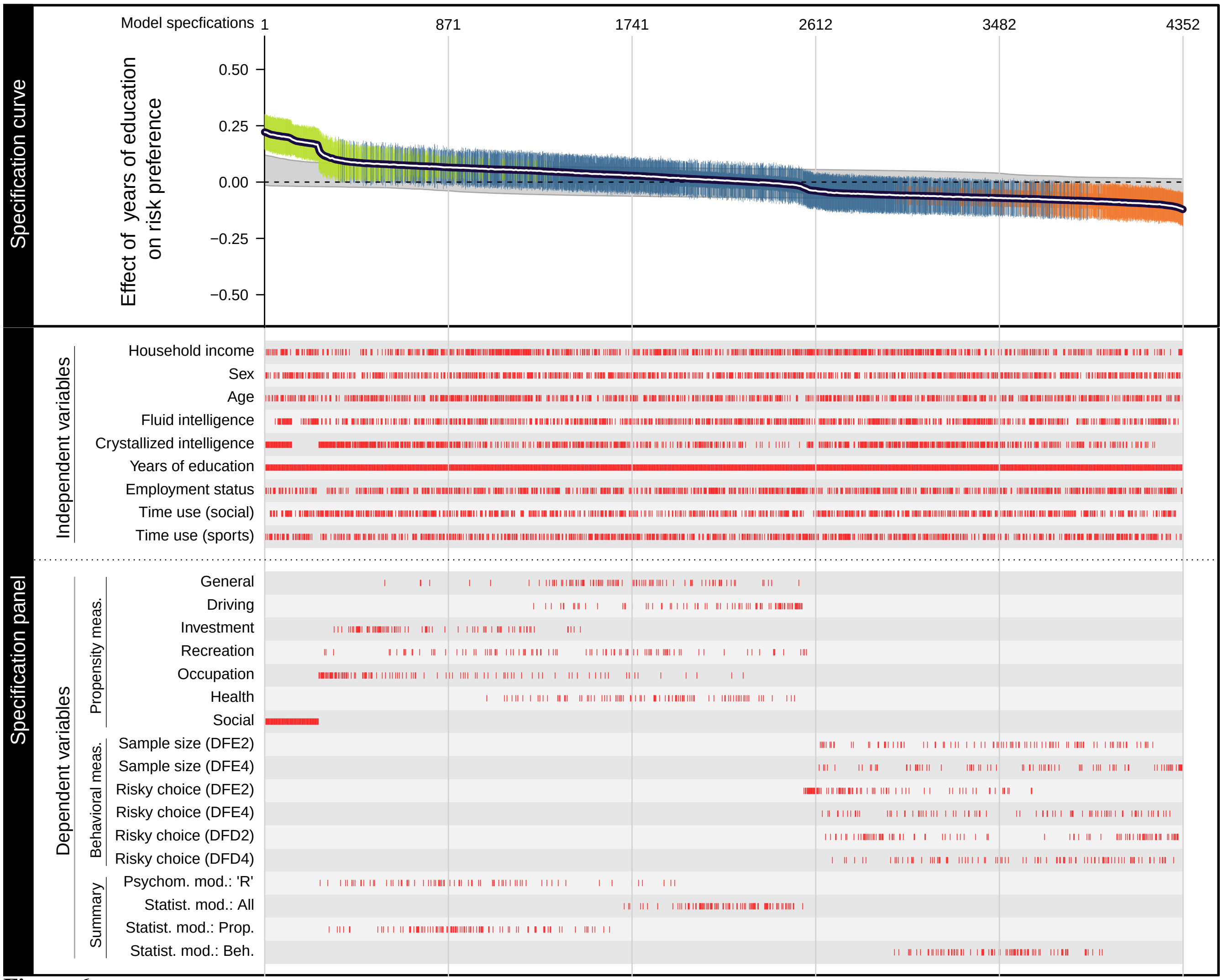


**Figure 6**

*Specification curve analysis for the association of* years of eduction *with risk preference. See Figure 1 for a detailed description of how to interpret the different elements in this figure.*

ciated with a reduced risk preference (Fig. 5), yet the simulation analyses illustrated that this finding may largely reflect a false-positive result (see previous section on the general results). If at all, these effects emerged in a single behavioral measure, namely, the proportion of risky choices in the DFD task with two options.

Furthermore, given that most estimated effects were close to 0, the tick marks for crystallized intelligence for the different IVs not surprisingly appear rather randomly dispersed compared to those for the other candidate correlates.

### *Years of education*

Finally, two interesting patterns emerged for the effect of years of education on risk preference (Fig. 6): on the one hand, a series of model specifications produced credible positive effects. Virtually all of these specifications used the operationalization of risk preference in the social domain. To our knowledge, no theoretical account predicts this particular association; however, one may speculate that higher education fosters competences and situations that may lead persons to take more social risks (e.g., public speaking; publicly challenging others' opinions; taking a leadership role).

On the other hand, all model specifications resulting in negative estimates occurred when one of the various behavioral measures was used. This is somewhat surprising given that more educated persons should, in principle, be in a better position to assess expected values, which in our tasks could lead to increased risk taking (the expected values of the riskier options were higher than that of the safer options).

## Discussion

Identifying correlates that are robustly associated with people's risk preferences is important for both theoretical and applied reasons. First, the role of some individual characteristics is directly linked to theories of risk preference. Consider again Bernoulli's seminal work on risk preference that aimed to capture how the same monetary amount could have more utility to a "pauper than a rich man" (Bernoulli, 1738): A corollary of Bernoulli's theory is a positive association between wealth and risk preference,[3] which stands in contrast to the negative association suggested by risk-sensitivity theory (Mishra, 2014). Estimating the association between risk preference and income can thus be seen as a test of specific and even competing theories. Second, with regard to policy implications solid knowledge about the correlates of risk preference promises to help identify individuals who are overly prone to take or to avoid risks (Steinberg, 2007), a first step towards providing personality-targeted prevention programs (Conrod et al., 2013).

Our main contribution in this article consists of conducting a comprehensive "meta-test" of a number of candidate correlates of risk preference, such as household income, sex, age, and cognitive ability, which have previously been put forth by various and largely independent theoretical accounts (see Tab. 1). Crucially, our analytic approach is unique in considering the links between the candidate correlates and several disparate measurement approaches to assessing risk preference, including general and domain-specific "stated preferences", as well as different types of "revealed preferences" (decisions from description and decisions from experience; Hertwig & Erev, 2009). Finally, our test of candidate correlates took into account the potentially competing correlates and possible confounds by relying on an exhaustive analytic approach, namely, specification curve analysis (Simonsohn et al., 2015).

### Main findings

Our analyses suggest that a person's economic situation, sex, age, as well as fluid intelligence and education are robustly associated with a person's risk preference. This general conclusion rests on an exhaustive analysis of a large number of model specifications, including a comparison of the number of observed effects against the number of effects expected by chance.

Beyond this general finding, the reported SCAs also clearly illustrate that the associations between the candidate correlates and risk preference are heterogeneous, with differences emerging both between domains (e.g., recreational vs. investment) and between measurement approaches (self-reported propensity measures vs. behavioral measures of risk preference). To illustrate, for sex and age, we found consistently negative associations between self-reported propensity measures and being female and older, but the strength of these associations varied across domains of life and, by and large, did not emerge for behavioral measures of risk preference. One exception to this pattern concerns the negative correlation between age and pre-decisional search in decisions from experience for larger choice set sizes.

Conversely, household income, fluid intelligence, and years of education were either positively *or* negatively associated with risk preference, depending on how risk preference was operationalized. Such specific variations across operationalizations of risk preference are not predicted by current theories.

### Implications for theories of risk preference

Some of our findings can be interpreted in light of the theories reviewed in the introduction, which have put forth a range of predictions concerning the associations of certain correlates and risk preference. Other patterns remain speculative and can be seen as food-for-thought for future theoretical and empirical work.

Concerning a person's economic situation, different theoretical accounts predict positive and negative associations with risk preference. Our results tend to support the former (i.e., negative associations) but only in particular domains of life, namely, investment and driving. As far as we know, there is no theoretical account that defines or puts limits on domain-specific effects of wealth, but it may be interesting to consider such aspects in future theorizing.

Concerning the effects of sex and age, essentially all previous theoretical accounts predict a reduced risk preference for women and for older people. Indeed, the results for self-reported propensity measures supported these predictions and demonstrated that the associations of these two candidate correlates with risk preference are the most consistent and strongest of all investigated correlates. On a more fine-grained level, however, our analyses pointed out nuanced patterns that are harder to account for under the current theories: for instance, the strongest negative effects of sex (being female) on risk preference emerged for the driving and recreational domains. This may be in line with evolutionary theories that emphasize the signaling function of risk taking for males, which is arguably easier in these domains, but could also result from injunctive norms about gender differences in risk taking for these domains. One should note that such domain-specific patterns remain speculative and require further theorizing in the future.

[3] Note that here we do not refer to a person's "risk aversion" as measured by the concavity of the utility curve, which could be identical for persons with different wealth. Rather, we refer to the fact that persons with different initial wealth are located at different positions on the (same) utility curve. Consequently, wealthier people will be willing to pay a smaller risk premium and thus be more likely to "accept" a risk.

Finally, concerning the role of cognitive ability and education, the empirical patterns provided support for only some theoretical accounts. For example, the observation that higher fluid intelligence is associated with a reduced risk preference in decisions from experience (i.e., more extensive pre-decisional exploration) is in line with the cognitive control hypothesis, but incompatible with the ontogenetic co-development and phylogenetic co-evolution hypotheses. One should note that the pattern of results for fluid intelligence suggests the need to consider the role of age: our results suggest that fluid intelligence may pick up some age effects in general and domain-specific risk preference that are not observed when age is controlled for, a result that calls into question the role of fluid intelligence as a robust and independent correlate of risk preference.

### *Implications for the measurement of risk preference*

Our work also has important implications for the use of different measurement approaches in both scientific and applied assessments of people's risk preferences. Our analyses showed that behavioral measures largely failed to pick up associations between the candidate correlates and risk preference. Crucially, this result emerged despite the behavioral measures involving substantial monetary incentives. Indeed, somewhat ironically, one of the few dependent variables stemming from behavioral measures (sample size in decisions from experience) that picked up associations with a candidate correlate (fluid intelligence) did not involve any monetary incentives (for a related finding, see Frey et al., 2017).

Taken together, the present findings suggest that self-reported propensity measures are better suited to capture individual differences in risk preferences related to socio-demographic variables, such as sex and age. This is true when considering both single measures of risk preference and aggregate indices. In other words, for many applications in both scientific and practical contexts, self-reported propensity measures or aggregates of these may be the better choice relative to single behavioral measures or aggregates thereof.

In sum, these results highlight the importance of giving careful attention to how risk preference is operationalized (Frey et al., 2017) and may help inform future measurement efforts that aim to assess the genetic (Linnér et al., 2019), hormonal (Kurath & Mata, 2018), or neural (Grubb et al., 2016) basis of individual differences in risk preference.

## Limitations

We point out three main limitations of our work. First, we tested only a limited number of candidate correlates. To the extent that one is interested in other correlates of risk preference (e.g., birth order; Lejarraga et al., 2019; Rohrer et al., 2017), additional predictors could be included and tested using SCA. It is important to keep in mind, however, that the inclusion of each additional independent variable will double the number of specifications and, accordingly, increase the amount of computational time. Consequently, in practice there is a limit to the number of candidate correlates that can be considered and we have thus focused on the most prominent candidates that have previously been suggested by various psychological theories. Similarly, even though our analyses included a wide array of different measures of risk preference, future work could adopt similar approaches using yet others of the many existing operationalizations of this construct (Appelt et al., 2011; Frey et al., 2017).

Second, in our model specifications we only included linear additive effects and no interaction terms between the various independent variables. We based this decision on two principles proposed by Hastie and Dawes (2010): according to the *mathematical principle*, ordinal (i.e., monotone) interactions are well approximated by linear additive effects. That is, in the context of our SCAs a potential interaction between two IVs can be approximated in specifications that include both variables as main effects. Although this approximation is naturally not perfect, according to the mathematical principle it should not systematically distort the overall interpretation as long as such interactions are not disordinal (i.e., crossed interactions). In fact, according to the *principle of nature*, most interactions that exist are ordinal rather than disordinal, and "it is easy to hypothesize crossed interactions, but extraordinarily difficult to find them in everyday situations, especially in the areas of psychology..." (p. 58). To illustrate, age by sex interactions may indeed result in an ordinal pattern, with the differences between men and women potentially decreasing across the lifespan due to a main effect of age. Such an interaction could be approximated in our SCAs. Yet, it is unlikely that age can completely reverse a main effect of sex (Josef et al., 2016), which could not be accounted for in our SCAs.

Third, personality characteristics including people's risk preferences cannot be manipulated experimentally and typically have to be studied, almost by definition, in correlational ways. That is, our work does not directly reveal any causal links between specific mechanisms and individual differences in risk preference. Nevertheless, our findings suggest it may be important to investigate more thoroughly the source of some robust effects, such as those concerning sex and age, for which there are a plethora of theorized causal mechanisms.

## Exhaustive modeling as a general tool in personality research

Our results suggest that exhaustive modeling approaches such as specification curve analysis are a powerful and useful tool to investigate the role of both predictors and operationalizations of constructs of interest. For example, we detected potential confounds concerning the role of age and fluid intelligence on risk preference, which illustrates the benefits of extensive robustness checks. Further, our analyses showed that the links between the six candidate correlates and risk preference depend mostly on how risk preference is measured, rather than whether and which control variables (e.g., employment status) are included in the models.

In the future, SCAs and related exhaustive modeling approaches may prove useful in the methodological toolbox of personality researchers to investigate the correlates of personality characteristics more broadly and systematically. For example, future conceptual work might profit from using SCA to identify the correlates of constructs closely related to risk preference, such as impulsivity and self-control (Nigg, 2016). The extent to which similar (or divergent) patterns concerning the correlates of these constructs are found will inform the need to distinguish or unify constructs. Such future efforts could thus help assess the usefulness of various constructs and ground our theories on a more solid empirical basis.

One conceptual issue surrounding the use of SCA is whether it should be seen as a confirmatory or an exploratory tool. As our application has illustrated, the two possibilities are not mutually exclusive. We employed SCA in a confirmatory way to run the most exhaustive meta-test of the candidate correlates proposed by various theoretical accounts (Tab. 1). At the same time, however, our implementation of SCA also revealed patterns that none of the reviewed theories predict, which may thus be instrumental in future theory building. For example, the strength of the associations between sex and risk preference varied as a function of domain. Even though various theories predict an association between being female and reduced risk preference, these theories are currently not detailed enough to predict any domain-specific differences.

All in all, we believe that for both confirmatory and exploratory applications, approaches such as SCA can foster transparency, reproducibility, and thus, ultimately, the credibility of personality research and the behavioral sciences in general (Baribault et al., 2018; Simonsohn et al., 2015; Steegen et al., 2016; Wacker, 2017).

## Conclusions

Our analyses shed light on the robustness of six candidate correlates of risk preference that were put forth by previous theories of risk preference. The results suggest that sex and age have robust and consistent associations with risk preference, whereas other candidate correlates are associated with risk preference in more nuanced ways. Furthermore, our analyses demonstrated the important role of operationalization in the context of assessing people's risk preference: clearly, not all measures of risk preference are created equal—self-reported propensity measures identified various associations with the candidate correlates, whereas behavioral measures by and large did not. Our work suggests that a careful choice of measures is needed when examining the role of individual characteristics in people's appetite for risk.

## Author Contributions

R.F., R.H., and R.M. designed research; R.F. performed research and conducted the data analysis; R.F. and R.M. wrote the article; D.R., J.S., and R.H. provided critical revisions to the article. All authors approved the final version.

## Declaration of Conflicting Interests

The authors declare no conflict of interest.

## References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Appelt, K. C., Milch, K. F., Handgraaf, M. J., & Weber, E. U. (2011). The Decision Making Individual Differences Inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*, *6*(3), 252–262. http://journal.sjdm.org/11/11218/jdm11218.html

Aven, T. (2012). The risk concept-historical and recent development trends. *Reliability Engineering & System Safety*, *99*, 33–44. https://doi.org/10.1016/j.ress.2011.11.006

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, *66*(3), 411–421.

Baker, M. D., & Maner, J. K. (2009). Male risk-taking as a context-sensitive signaling device. *Journal of Experimental Social Psychology*, *45*(5), 1136–1139. https://doi.org/10.1016/j.jesp.2009.06.006

Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., van Ravenzwaaij, D., White, C. N., Boeck, P. D., & Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences*, *115*(11), 2607–2612. https://doi.org/10.1073/pnas.1708285114

Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, *103*(6), 2499–2529. https://doi.org/10.1257/aer.103.6.2499

Batty, G. D., Deary, I. J., & Gottfredson, L. S. (2007). Premorbid (early life) iq and later mortality risk: Systematic review. *Annals of Epidemiology*, *17*(4), 278–288. https://doi.org/10.1016/j.annepidem.2006.07.010

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, *88*(4), 354–364. https://doi.org/10.1037/0033-295X.88.4.354

Bernoulli, D. (1738). Exposition of a new theory on the measurement of risk. *Econometrica*, *22*(1), 23–36. https://doi.org/10.2307/1909829

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. https://doi.org/10.1037/t13084-000

Bleidorn, W., Klimstra, T. A., Denissen, J. J. A., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2013). Personality maturation around the world: A cross-cultural examination of social-investment theory. *Psychological Science*, *24*(12), 2530–2540. https://doi.org/10.1177/0956797613498396

Boyle, P. A., Yu, L., Buchman, A. S., Laibson, D. I., & Bennett, D. A. (2011). Cognitive function is associated with risk aversion in community-based older persons. *BMC geriatrics*, *11*(1), 53. https://doi.org/10.1186/1471-2318-11-53

Brailovskaia, J., Schillack, H., Assion, H.-J., Horn, H., & Margraf, J. (2018). Risk-taking propensity and (un)healthy behavior in Germany. *Drug and Alcohol Dependence*. https://doi.org/10.1016/j.drugalcdep.2018.08.027

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, *125*(3), 367–383. https://doi.org/10.1037/0033-2909.125.3.367

Calcagno, V., & de Mazancourt, C. (2010). Glmulti: An r package for easy automated model selection with (generalized) linear models. *Journal of Statistical Software*, *34*(12), 1–29. https://doi.org/10.18637/jss.v034.i12

Carstensen, L. L., Pasupathi, M., Mayr, U., & Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology*, *79*(4), 644. https://doi.org/10.1037//0022-3514.79.4.644

Clark, W. A. V., & Lisowski, W. (2017). Prospect theory and the decision to move or stay. *Proceedings of the National Academy of Sciences*, *114*(36), E7432–E7440. https://doi.org/10.1073/pnas.1708505114

Conrod, P. J., O'Leary-Barrett, M., Newton, N., Topper, L., Castellanos-Ryan, N., Mackie, C., & Girard, A. (2013). Effectiveness of a selective, personality-targeted prevention program for adolescent alcohol use and misuse: A cluster randomized controlled trial. *JAMA Psychiatry*, *70*(3), 334–342. https://doi.org/10.1001/jamapsychiatry.2013.651

Corter, J. E., & Chen, Y.-J. (2005). Do investment risk tolerance attitudes predict portfolio risk? *Journal of Business and Psychology*, *20*(3), 369–381. https://doi.org/10.1007/s10869-005-9010-5

Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex differences in impulsivity: A meta-analysis. *Psychological Bulletin*, *137*(1), 97–130. https://doi.org/10.1037/a0021591

Cross, C. P., Cyrenne, D.-L. M., & Brown, G. R. (2013). Sex differences in sensation-seeking: A meta-analysis. *Scientific Reports*, *3*, 2486. https://doi.org/10.1038/srep02486

Defoe, I. N., Dubas, J. S., Figner, B., & an Aken, M. A. G. (2015). A meta-analysis on age differences in risky decision making: Adolescents versus children and adults. *Psychological Bulletin*, *141*(1), 48–84. https://doi.org/10.1037/a0038088

Depping, M. K., & Freund, A. M. (2011). Normal aging and decision making: The role of motivation. *Human Development*, *54*, 349–367. https://doi.org/10.1159/000334396

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. https://doi.org/10.1111/ecoj.12322

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, *100*(3), 1238–1260. https://doi.org/10.1257/aer.100.3.1238

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2018). On the relationship between cognitive ability and risk preference. *Journal of Economic Perspectives*, *32*(2), 115–134. https://doi.org/10.1257/jep.32.2.115

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550. https://doi.org/10.1111/j.1542-4774.2011.01015.x

Düzel, E., Bunzeck, N., Guitart-Masip, M., & Düzel, S. (2010). NOvelty-related motivation of anticipation and exploration by dopamine (NOMAD): Implications for healthy aging. *Neuroscience & Biobehavioral Reviews*, *34*, 660–669. https://doi.org/10.1016/j.neubiorev.2009.08.006

Dwyer, P. D., Gilkeson, J. H., & List, J. A. (2002). Gender differences in revealed risk taking: Evidence from mutual fund investors. *Economics Letters*, *76*(2), 151–158. https://doi.org/10.1016/S0165-1765(02)00045-9

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380–417. https://doi.org/10.1037/h0053870

Eeckhoudt, L., Gollier, C., & Schlesinger, H. (2005). Risk aversion. In *Economic and financial decisions under risk*. Princeton University Press.

Figner, B., & Weber, E. U. (2011). Who takes risks when and why? Determinants of risk taking. *Current Directions in Psychological Science*, *20*(4), 211–216. https://doi.org/10.1177/0963721411415790

Frankenhuis, W. E., Dotsch, R., Karremans, J. C., & Wigboldus, D. H. J. (2010). Male physical risk taking in a virtual environment. *Journal of Evolutionary Psychology*, *8*(1), 75–86. https://doi.org/10.1556/JEP.8.2010.1.6

Frey, R., Mata, R., & Hertwig, R. (2015). The role of cognitive abilities in decisions from experience: Age differences emerge as a function of choice set size. *Cognition*, *142*, 60–80. https://doi.org/10.1016/j.cognition.2015.05.004

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances*, *3*, e1701381. https://doi.org/10.1126/sciadv.1701381

Frey, R., Rieskamp, J., & Hertwig, R. (2015). Sell in May and go away? Learning and risk taking in nonmonotonic decision problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 193–208. https://doi.org/10.1037/a0038118

Granger, C. W., & Uhlig, H. F. (1990). Reasonable extreme-bounds analysis. *Journal of Econometrics*, *44*(1-2), 159–170. https://doi.org/10.1016/0304-4076(90)90077-7

Grubb, M. A., Tymula, A., Gilaie-Dotan, S., Glimcher, P. W., & Levy, I. (2016). Neuroanatomy accounts for age-related changes in risk preferences. *Nature Communications*, *7*, 13822. https://doi.org/10.1038/ncomms13822

Hanoch, Y., Johnson, J. G., & Wilke, A. (2006). Domain specificity in experimental measures and participant recruitment. An application to risk-taking behavior. *Psychological Science*, *17*(4), 300–304. https://doi.org/10.1111/j.1467-9280.2006.01702.x

Harris, C. R., Jenkins, M., & Glaser, D. (2006). Gender differences in risk assessment: Why do women take fewer risks than men. *Judgment and Decision Making*, *1*(1), 48–63. http://journal.sjdm.org/06016/jdm06016.htm

Hastie, R., & Dawes, R. M. (2010). A general framework for judgment. In *Rational choice in an uncertain world: The psychology of judgment and decision making* (pp. 47–72). SAGE.

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

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*. https://doi.org/10.1002/bdm.1953

Hills, T. T., Noguchi, T., & Gibbert, M. (2013). Information overload or search-amplified risk? Set size and order effects on decisions from experience. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-013-0422-3

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*(5), 1157–1164. https://doi.org/10.3758/s13423-013-0572-3

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*(4), 382–401.

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 life span. *Journal of Personality and Social Psychology*, *111*(3), 430–450. https://doi.org/10.1037/pspp0000090

Kurath, J., & Mata, R. (2018). Individual differences in risk taking and endogeneous levels of testosterone, estradiol, and cortisol: A systematic literature search and three independent meta-analyses. *Neuroscience & Biobehavioral Reviews*, *90*, 428–446. https://doi.org/10.1016/j.neubiorev.2018.05.003

Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review*, *73*(1), 31–43. Retrieved March 19, 2018, from http://www.jstor.org/stable/1803924

Lejarraga, T., Frey, R., Schnitzlein, D. D., & Hertwig, R. (2019). No effect of birth order on adult risk taking. *Proceedings of the National Academy of Sciences*, 201814153. https://doi.org/10.1073/pnas.1814153116

Lilleholt, L. (2019). Cognitive ability and risk aversion: A systematic review and meta analysis. *Judgment and Decision Making*, *14*(3), 234–279. http://journal.sjdm.org/18/181011/jdm181011.html

Linnér, R. K., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., Lebreton, M., Tino, S. P., Abdellaoui, A., Hammerschlag, A. R., Nivard, M. G., Okbay, A., Rietveld, C. A., Timshel, P. N., Trzaskowski, M., de Vlaming, R., Zünd, C. L., Bao, Y., Buzdugan, L., … Beauchamp, J. P. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, *51*(2), 245. https://doi.org/10.1038/s41588-018-0309-3

Mamerow, L., Frey, R., & Mata, R. (2016). Risk taking across the life span: A comparison of self-report and behavioral measures of risk taking. *Psychology and Aging*, *31*(7), 711–723. https://doi.org/10.1037/pag0000124

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*(1), 18–29. https://doi.org/10.1111/j.1749-6632.2011.06200.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. https://doi.org/10.1257/jep.32.2.155

Mata, R., Josef, A. K., & Hertwig, R. (2016). Propensity for risk taking across the life span and around the globe. *Psychological Science*, *27*(2), 231–243. https://doi.org/10.1177/0956797615617811

Mishra, S. (2014). Decision-making under risk: Integrating perspectives from biology, economics, and psychology. *Personality and Social Psychology Review*, 1088868314530517. https://doi.org/10.1177/1088868314530517

Mishra, S., Gregson, M., & Lalumière, M. L. (2011). Framing effects and risk-sensitive decision making. *British Journal of Psychology*, 83–97. https://doi.org/10.1111/j.2044-8295.2011.02047.x

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*(4), 674–701. https://doi.org/10.1037/0033-295X.100.4.674

Nigg, J. T. (2016). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.12675

Olschewski, S., Rieskamp, J., & Scheibehenne, B. (2018). Taxing cognitive capacities reduces choice consistency rather than preference: A model-based test. *Journal of Experimental Psychology: General*, *147*(4), 462–484. https://doi.org/10.1037/xge0000403

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*(2), 173–182. https://doi.org/10.1038/s41562-018-0506-1

Pleskac, T. J. (2008). Decision making and learning while taking sequential risks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 167–185. https://doi.org/10.1037/0278-7393.34.1.167

Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, *143*(5), 2000. https://doi.org/10.1037/xge0000013

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/271063

Roberts, B. W., Wood, D., & Smith, J. L. (2005). Evaluating Five Factor Theory and social investment perspectives on personality trait development. *Journal of Research in Personality*, *39*(1), 166–184. https://doi.org/10.1016/j.jrp.2004.08.002

Rohrer, J. M., Egloff, B., & Schmukle, S. C. (2017). Probing birth-order effects on narrow traits using specification-curve analysis. *Psychological Science*, *28*(12), 1821–1832. https://doi.org/10.1177/0956797617723726

Rolison, J. J., Hanoch, Y., & Wood, S. (2012). Risky decision making in younger and older adults: The role of learning. *Psychology and Aging*, *27*(1), 129–140. https://doi.org/10.1037/a0024689

Rolison, J. J., Hanoch, Y., Wood, S., & Liu, P.-J. (2014). Risk-taking differences across the adult life span: A question of age and domain. *The Journals of Gerontology: Series B*, *69*(6), 870–880. https://doi.org/10.1093/geronb/gbt081

Sala-i-Martin, X. X. (1997). I just ran four million regressions. *The American Economic Review*, *87*(2), 178–183. https://doi.org/10.3386/w6252

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61–71. https://doi.org/10.2307/2548836

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

Schubert, R., Brown, M., Gysler, M., & Brachinger, H. W. (1999). Financial decision-making: Are women really more risk-averse? *American Economic Review*, *89*(2), 381–385. https://doi.org/10.1257/aer.89.2.381

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification curve: Descriptive and inferential statistics on all reasonable specifications. *SSRN*

*Electronic Journal*, 1–18. https://doi.org/10.2139/ssrn.2694998
Slovic, P. (1966). Risk-taking in children: Age and sex differences. *Child Development*, *37*(1), 169–176.
Slovic, P. (1964). Assessment of risk taking behavior. *Psychological Bulletin*, *61*(3), 220–233. https://doi.org/10.1037/h0043608
Stan Development Team. (2016). Rstanarm: Bayesian applied regression modeling via Stan. http://mc-stan.org/
Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637
Steinberg, L. (2007). Risk taking in adolescence: New perspectives from brain and behavioral science. *Current Directions in Psychological Science*, *16*(2), 55–59. https://doi.org/10.1111/j.1467-8721.2007.00475.x
Steinberg, L. (2013). Does recent research on adolescent brain development inform the mature minor doctrine? *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, *38*(3), 256–267. https://doi.org/10.1093/jmp/jht017
Steiner, M., D., & Frey, R. (2020). Representative design in psychological assessment: A case study using the Balloon Analogue Risk Task (BART). *PsyArXiv*. https://doi.org/10.31234/osf.io/dg4ks
Steiner, M., D., Seitz, F., & Frey, R. (2019). Through the window of my mind: Mapping the cognitive processes underlying self-reported risk preference. *PsyArXiv*. https://doi.org/10.31234/osf.io/sa834
Stephens, D. W. (1981). The logic of risk-sensitive foraging preferences. *Animal Behaviour*, *29*(2), 628–629. https://doi.org/10.1016/S0003-3472(81)80128-5
Trivers, R. (1972). *Parental investment and sexual selection* (Vol. 136). Biological Laboratories, Harvard University Cambridge.
van den Bos, W., & Hertwig, R. (2017). Adolescents display distinctive tolerance to ambiguity and to uncertainty during risky decision making. *Scientific Reports*, *7*, 40962. https://doi.org/10.1038/srep40962
van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03
Wacker, J. (2017). Increasing the reproducibility of science through close cooperation and forking path analysis. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01332
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
Yechiam, E., & Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making*, *19*(1), 1–16. https://doi.org/10.1002/bdm.509