

BEHAVIORAL PSYCHOLOGY

Risk preference shares the psychometric structure of major psychological traits

Renato Frey,^{1,2*} Andreas Pedroni,^{3,4} Rui Mata,¹ Jörg Rieskamp,⁴ Ralph Hertwig²

To what extent is there a general factor of risk preference, *R*, akin to *g*, the general factor of intelligence? Can risk preference be regarded as a stable psychological trait? These conceptual issues persist because few attempts have been made to integrate multiple risk-taking measures, particularly measures from different and largely unrelated measurement traditions (self-reported propensity measures assessing stated preferences, incentivized behavioral measures eliciting revealed preferences, and frequency measures assessing actual risky activities). Adopting a comprehensive psychometric approach (1507 healthy adults completing 39 risk-taking measures, with a subsample of 109 participants completing a retest session after 6 months), we provide a substantive empirical foundation to address these issues, finding that correlations between propensity and behavioral measures were weak. Yet, a general factor of risk preference, *R*, emerged from stated preferences and generalized to specific and actual real-world risky activities (for example, smoking). Moreover, *R* proved to be highly reliable across time, indicative of a stable psychological trait. Our findings offer a first step toward a general mapping of the construct risk preference, which encompasses both general and domain-specific components, and have implications for the assessment of risk preference in the laboratory and in the wild.

INTRODUCTION

The term “risk” refers to properties of the world, yet without a clear agreement on its definition, which has ranged from probability, chance, outcome variance, expected values, undesirable events, danger, losses, to uncertainties (1, 2). People’s responses to those properties, on the other hand, are typically described as their “risk preference.” For the behavioral sciences, particularly psychology and economics, risk preference is one of the key building blocks of human behavior and hence an important aspect to be captured by choice theories. However, there is still considerable disagreement on how to conceptualize risk preference on at least two accounts. First, there is no conclusive answer as to whether risk preference represents a unitary construct, a multidimensional construct that varies across life domains, or even a combination of both (3–7). Second, there is disagreement concerning whether people’s risk preferences can be thought of as relatively stable—like other psychological traits (8) and as in the classic economic view of enduring tastes (9, 10)—or, alternatively, as varying substantially over time and thus more resembling a state (11, 12).

Paralleling the conceptual disagreement, there is also a lack of consensus regarding how to best measure risk preference (13), with one of the reasons being that different disciplines have used different definitions of risk: Whereas economists usually define risk as variance of monetary outcomes, psychologists and clinicians often refer to risk as behavior with potentially harmful consequences (14). In turn, three major but largely unrelated measurement traditions have evolved across recent decades: Proponents of the stated-preference tradition harness people’s introspective abilities and rely on stated preferences obtained in response either to relatively abstract questions (“Are you generally a risk-taking person or do you try to avoid risks?”) or to more specific but hypothetical scenarios (for example, “How likely would you be to go white-water rafting at high water in the spring?”) (11, 15–18). These

self-reported “propensity measures” are widely used in practice, not least because they are relatively easy to implement. For example, financial companies often use propensity measures to assess their clients’ risk preference in accordance with the legal requirements for the sale of financial products (19). Proponents of the revealed-preference tradition (20), in contrast, hold that “talk is cheap” to the extent that stated preferences can be expressed gratuitously without real consequences and that only actual behavior elicited with tasks involving monetary incentives reveals true risk preference. Behavioral measures range from abstract tasks such as monetary gambles to more naturalistic, game-like tasks. They have often been designed to capture specific cognitive processes (14, 21), such as the integration of gains and losses or the role of learning and experience. Finally, the construct risk preference has also been used in clinical and epidemiological studies. There, the frequency of actual risky activities is assessed (for example, “How many cigarettes do you smoke per day?”) to examine their long-term effects on morbidity and mortality (22–24). These frequency measures are typically self-reported, as in the stated-preference tradition, but they focus on specific and observable behavior, as in the revealed-preference tradition.

Across various disciplines such as psychology and economics, the numerous extant risk-taking measures from the different traditions are often used interchangeably, as if they all capture the same underlying construct (13, 14, 21). However, to date, there have been few integrative attempts to study the conceptual issues raised above. For example, do stated preferences (obtained through propensity measures), revealed preferences (obtained through behavioral measures), and assessments of the engagement in actual risky activities (obtained through frequency measures) capture the same underlying general construct or set of constructs? Can these general or specific constructs of risk preference be thought of as stable traits?

The psychometric structure of risk preference

In answering questions about the psychometric structure of risk preference, it is helpful to draw an analogy with fields that have made substantial progress in addressing similar questions. For example, the field of intelligence has relied on extensive psychometric modeling of large batteries of tasks to assess the convergent validity of different intelligence

Copyright © 2017
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Center for Cognitive and Decision Sciences, Department of Psychology, University of Basel, Basel, Switzerland. ²Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany. ³Methods of Plasticity Research, Department of Psychology, University of Zurich, Zurich, Switzerland. ⁴Center for Economic Psychology, Department of Psychology, University of Basel, Basel, Switzerland.

*Corresponding author. Email: renato.frey@unibas.ch

measures (that is, the extent to which measures that are supposed to capture the same underlying construct correlate with each other). Research on the construct of intelligence suggests that although there are specific dissociable cognitive faculties, such as verbal and spatial abilities, there is also the persistent finding of a “positive manifold” across measures, which can be captured by a general factor of intelligence, g (25). This general factor accounts for about 50% of the variance—consistent with the observation that those who do well in one cognitive domain also tend to do well in others (interestingly, despite decades of research, the exact mechanisms that lead to this positive manifold remain disputed) (26). Similar patterns have been documented in the realm of psychopathology, in which a general factor, p , also accounts for a large proportion of the observed variance (27). Crucially, mapping the psychometric structure of intelligence and identifying its general factor have been instrumental in making significant progress in understanding its neural (28) and genetic bases (29), suggesting that this could be an important path for the understanding of the construct risk preference and respective theories as well.

To date, it is unknown whether a positive manifold also exists across the extant risk-taking measures and, consequently, to what extent there is a general factor of risk preference, R , that captures commonalities between measures or domains. The possible existence of a general factor of risk preference operating across different measures and life domains, such as health, wealth, or recreation, would inform our theoretical conception of the psychological construct: Is it akin to a domain-general, unitary trait, or instead, is it more appropriate to assume multiple psychological constructs that need to be invoked to account for inter-individual differences in risk preference across domains? In the extreme, observing no common variance in risk preference across measures and life domains would call into question the idea of risk preference as a unitary psychological trait and challenge the way classic economic theorizing has conceived of risk preference. In contrast, extracting a general factor across the diverse measures of risk preference would imply that such a general factor captures systematic common variance, over and above the improvements in reliability that are to be expected when aggregating measures that were designed to capture the same construct (because various risk-taking measures were purposely designed to capture domain-specific risk preference, for example, recreational risk taking and health risk taking).

For several reasons, past empirical evidence does not permit a conclusive answer regarding the general versus domain-specific nature of risk preference. First, the vast majority of past empirical work on risk preference has typically used single measures or not explicitly explored the convergent validity of multiple measures. For example, a meta-analysis of age differences in risk preference found that only 2 in 29 studies reported multiple measures of risk preference (with one study reporting two measures and another study reporting three measures of risk preference) (30).

Second, the few studies that have systematically estimated the empirical convergence between different measures have produced mixed results: Although some studies show significant associations between propensity measures as well as links between the latter and behavioral measures (11, 16, 31), others suggest poor convergence between measures from the same (behavioral) tradition (32–34). One should note that the significant effects reported in these studies are typically small. In addition, these studies were often quite restricted in the number of implemented measures, thus often not encompassing measures that represent different measurement traditions, such as stated and revealed preferences.

Third, and finally, unlike in the fields of intelligence or psychopathology, past studies on the construct of risk preference have typically not adopted state-of-the-art techniques of psychometric modeling, with only few exceptions that either implemented relatively few measures (35) or used psychometric modeling only within one specific questionnaire (36). However, these techniques are indispensable for a clear decomposition of the measures’ variance into shared and unique components that indicate how risk preference should be conceptualized: as a unitary construct, a collection of independent factors, or a combination of both.

In sum, the existing evidence cannot conclusively answer the key question of the extent to which risk preference should be thought of as a general construct that explains variance across different types of measures and life domains. To address this issue, we apply psychometric tools as have previously been implemented, for instance, in research on intelligence.

The temporal stability of risk preference

The issue of temporal stability is paramount to a conceptualization of constructs as either traits or states. The definition of a psychological trait hinges largely on consistency across time (37, 38). This definition does not exclude the possibility that a trait can show sizable variation as a function of specific life stages or momentary shocks, but it assumes some basic degree of rank-order stability (11). Accordingly, an interactional view regards states as a person’s adaptation to particular situations, whereas traits are assumed to remain stable across time and situations (37, 38). Consequently, a state is typically defined as a relatively rapid and reversible variation around a person’s mean-level behavior or preferences, which may be associated with either the exogenous environment (situational factors) or the endogenous environment (for example, cognitive or emotional processes occurring within a person) (38). Finally, “trait change” refers to mean-level changes across time, which, as outlined above, does not preclude the possibility of substantial rank-order stability across persons.

To date, relatively little is known about the temporal stability of risk preference, particularly in terms of a potential general factor, making strong conclusions about whether risk preference may include a trait component impossible (39). However, recent work suggests considerable stability for stated preferences even across periods of years (11), and repeated assessments of stated risk preferences have rendered it possible to extract a reliable signal, which proved to have increased predictive validity for risk-taking behaviors as opposed to the single momentary assessments (40). In contrast, less is known about the stability of risk preference when assessed with behavioral measures, particularly involving longer delays (39). To uncover the potential trait- or state-like nature of the construct of risk preference, it is therefore essential to assess the temporal stability of stated and revealed preferences, and, in particular, of the psychometric factors extracted from them. Under the assumption that these psychometric factors capture somewhat stable constructs, they can be expected to reflect more error-free (and thus more reliable) indicators of a person’s risk preference.

Overview and aims of the present study

Our study was designed to close these gaps empirically by adopting a comprehensive psychometric framework based on a large battery of risk-taking measures that were sampled from the different measurement traditions. In so doing, we sought to examine (i) the convergent validity of different risk-taking measures and measurement traditions, (ii) the extent to which a general factor of risk preference exists across

measures and domains, and (iii) the temporal stability of risk preference as measured by single measures and the psychometric factors extracted from them (in a subsample of participants). That is, the study's goal was to clarify the degree to which risk preference should be conceptualized as a general construct, as separate distinct components, or as a combination of both, as well as to examine the temporal stability of these constructs.

Mapping the psychometric structure and stability of risk preference will be a crucial step toward uncovering its potential biological underpinnings (10, 41, 42) and its real-world consequences (43, 44). Relatedly, a comprehensive examination of the construct's psychometric structure may permit classifying it within the space of other psychological constructs and clarify to what extent interindividual differences in risk preference correspond with interindividual differences in potentially related variables. For example, socioeconomic factors such as income or education (16, 45–47), cognitive and numerical abilities (48, 49), and personality characteristics such as the Big Five personality traits (11, 50) have all been suggested to overlap or even drive risk preference. We will thus explore how predictive these variables are for the different measures and particularly for the psychometric factors extracted from them.

RESULTS

A daylong session in one of the two study centers, Basel (Switzerland) or Berlin (Germany), was completed by 1507 healthy adults (aged 20 to 36 years; table S1). This comprised a series of five questionnaires assessing stated preferences (22 propensity measures including all subscales), eight behavioral tasks assessing revealed preferences (seven of which were incentivized with monetary amounts as typically used in the literature; in total, there were nine behavioral measures including all dependent variables), and six scales assessing current and past risky activities, such as smoking, substance use, and gambling (eight frequency measures including all subscales; see Table 1). These 39 measures represent a broad sample of popular measures in research on risk preference (14, 21). Detailed information about the design and protocol of the study is provided in Materials and Methods and in the Supplementary Materials.

Convergent validity

Figure 1 depicts a network plot of the partial correlations between all measures (only correlations exceeding an absolute value of 0.1; the full correlation matrix is provided in table S2), after controlling for age, sex, and study center. We controlled for age and sex to examine the convergent validity of different measures beyond the influence of these two key variables underlying risk preference (that is, a tendency for a reduction in risk preference across the life span, starting in the early adulthood, as well as a stronger preference for risk in males) (11, 18, 30, 51–53). The network was generated using a force-directed algorithm such that correlated nodes (that is, risk-taking measures) attracted each other and uncorrelated nodes repulsed each other (54). Overall, there was a substantial gap between the stated- and revealed-preference measurement traditions, with the behavioral measures correlating only weakly with the propensity measures ($M = 0.06$, $HDI = 0.05$ to 0.06). Moreover, the correlations among the behavioral measures were substantially weaker ($M = 0.08$, $HDI = 0.06$ to 0.10) than those among the propensity measures ($M = 0.20$, $HDI = 0.18$ to 0.21 ; $\Delta M = 0.12$, $\Delta HDI = 0.09$ to 0.14 , $d = 1.22$) or the frequency measures ($M = 0.18$, $HDI = 0.14$ to 0.23 ; $\Delta M = 0.10$, $\Delta HDI = 0.05$ to 0.15 , $d = 1.15$). The frequency measures' correlations with the propensity measures were substantially stronger ($M = 0.13$, $HDI = 0.12$ to 0.15) than those with the behavioral measures ($M = 0.03$, $HDI = 0.03$ to

0.04). This pattern suggests that different propensity measures capture related components of the construct risk preference (which seem to be related to those captured by frequency measures), whereas each behavioral measure captures unique variance that is unrelated to that of the propensity measures, frequency measures, or even other behavioral measures.

These results were corroborated by rank-order analyses that relax the assumption of linear relations between measures (Fig. 1, panels on the right). Specifically, for each measure, we assigned a rank to every person (rank 1 for the most risk-taking person and rank 1507 for the least risk-taking person) and then determined each person's mean rank across measures, separately for the three measurement traditions. Provided that there is sufficient resolution within and perfect consistency across measures, the resulting mean ranks would be uniformly dispersed between 1 and 1507 (that is, forming a diagonal line in the panels on the right of Fig. 1). Because most measures do not provide a resolution for 1507 distinct values, we simulated perfectly consistent and random ranks across measures to obtain two extreme benchmarks for consistency. We then determined how much larger the SDs of the empirical ranks were, relative to those of the respective random ranks (a wide dispersion of ranks implies higher consistency across measures). Compared with their respective random rank orders, the empirical rank orders had a wider dispersion of 115.6 ranks (propensity measures; $HDI = 107.1$ to 124.0), 31.8 ranks (behavioral measures; $HDI = 23.2$ to 40.7), and 61.7 ranks (frequency measures; $HDI = 52.2$ to 71.4). That is, and as Fig. 1 shows (see the panels on the right), the empirical rank order resulting from the propensity measures fell in between perfectly consistent and random ranking, suggesting that the different propensity measures produced a somewhat but not perfectly consistent pattern. In contrast, the empirical rank order of the behavioral measures was almost identical to that expected from random ranking. The consistency of the frequency measures was somewhat higher but did not come close to that of the propensity measures.

Psychometric modeling

To investigate the extent to which the relation between measures could stem from a general risk-preference factor, R , while accommodating specific factors that stem from shared variance between specific measures and domains, we implemented a psychometric bifactor model (36, 55, 56). In contrast to a hierarchical model, in which the general (higher-order) factor merely summarizes shared variance across first-order factors, the general factor in a bifactor model directly accounts for shared variance at the level of measures, leaving the residual variance to be captured by specific, orthogonal factors. Hence, it is a more direct test for the existence of a general factor. Figure 2 shows the results of this analysis, namely, the general factor R and the specific factors $F1$ to $F7$, as well as the proportion of variance in each of the measures accounted for by these factors [depicted as colored stacked bars; factor loadings of the preceding exploratory factor analysis (EFA) are provided in table S4]. R accounted for a substantial portion of variance in several propensity measures and frequency measures but for little to no variance in the behavioral measures. Of the explained variance (31%), R accounted for 61%, and all the specific factors together accounted for 39%. Confirmatory factor analysis (CFA) indicated that the model fit was satisfactory [standardized root mean square residual (SRMR) = 0.05; root mean square error of approximation (RMSEA) = 0.04; $df = 632$; comparative fit index (CFI) = 0.94; Tucker-Lewis index (TLI) = 0.93]. However, overall, a substantial proportion of variance (in particular in the behavioral measures) could not be accounted for. To compare, in fields

Table 1. Risk-taking measures used in the Basel-Berlin Risk Study. DV, dependent variable. All measures were coded such that higher values indicate more risk taking, except for “DFEss” (a larger sample size may reflect stronger uncertainty reduction and thus less risk taking).

Measure	Subscale/DV	Abbreviation
Propensity measures		
Socioeconomic panel (90)	General risk taking	SOEP
	Financial	SOEPfin
	Health	SOEPhea
	Recreational	SOEPrec
	Occupational	SOEPocc
	Social	SOEPsoc
	Driving	SOEPdri
Domain-specific risk-attitude scale (4)	Investment	Dinv
	Gambling	Dgam
	Health	Dhea
	Recreational	Drec
	Ethical	Deth
	Social	Dsoc
Gambling Attitude and Beliefs Survey (91)	Total score	GABS
Personal Risk Inventory (92)	Total score	PRI
Sensation Seeking Scale (93)	Thrill and adventure seeking	SStas
	Experience seeking	SSexp
	Disinhibition	SSdis
	Boredom susceptibility	SSbor
Barratt's Impulsivity Scale (94)	Attentional	BISa
	Motor	BISm
	Nonplanning behavior	BISn
Behavioral measures		
Balloon Analogue Risk Task (95)	Number of pumps	BART
Decisions from experience (72)	Sample size	DFEss
	% Risky choices	DFEre
Decisions from description (72)	% Risky choices	DFD
Adaptive lotteries (96)	% Risky choices	LOT
Multiple price list (97)	Switching point (inverted)	MPL
Columbia Card Task (98)	Number of cards	CCT
Marbles task (99)	% Risky choices	MT
Vienna Risk-Taking Test Traffic (100)	Reaction latency	VRTTT
Frequency measures		
Alcohol use disorders identification test (22)	Total score	AUDIT

continued on next page

Measure	Subscale/DV	Abbreviation
Fagerström test for nicotine dependence (23)	Total score	FTND
Pathological gambling (24)	Total score	PG
Drug Abuse Screening Test (101)	Total score	DAST
Encounters with risky situations (102)	Aggressive behavior	CAREa
	Sexual behavior	CAREs
	Behavior at work	CAREw
Risky behaviors in the past month (4)	Total score	Dm

with stronger test-theoretic backgrounds, similar bifactor models accounted for substantially higher proportions of explained variance, ranging from 50% in intelligence (55) up to 62% in psychopathology (56). When we implemented a bifactor model with an inclusion criterion that was twice as strict (see Materials and Methods), resulting in a winnowed-down set of 13 of the 39 measures (fig. S3), the model (RMSEA = 0.06; SRMR = 0.05; *df* = 52; CFI = 0.97; TLI = 0.95) accounted for 53% of the variance, with *R* still accounting for 50% of the explained variance. Crucially, this latter solution excluded all behavioral measures.

In sum, a general factor of risk preference, *R*, could be extracted. This factor explained substantial variance across propensity measures and frequency measures of risky activities but did not generalize to behavioral measures. Moreover, there was only one specific factor that captured common variance across behavioral measures, specifically, choices among different types of risky lotteries (F7). Beyond the variance accounted for by *R*, the remaining six factors captured specific variance associated with health risk taking (F1), financial risk taking (F2), recreational risk taking (F3), impulsivity (F4), traffic risk taking (F5), and risk taking at work (F6).

Temporal stability

To assess whether any of the risk-taking measures or extracted factors exhibit some basic degree of stability, we assessed test–retest correlations in a subsample of 109 participants, who completed the battery of measures twice, at a 6-month interval (see Materials and Methods). Factor values were computed on the basis of the model identified in the main sample; that is, no new model was estimated in the smaller retest subsample. As Fig. 3 shows, the test–retest correlations of the propensity measures (*M* = 0.68, HDI = 0.64 to 0.72) and the frequency measures (*M* = 0.65, HDI = 0.52 to 0.77) tended to be higher than those of the behavioral measures (*M* = 0.46, HDI = 0.29 to 0.63; $\Delta M = 0.22$, $\Delta HDI = 0.04$ to 0.39, *d* = 1.3; $\Delta M = 0.19$, $\Delta HDI = -0.02$ to 0.40, *d* = 0.92). Notably, the general risk-preference factor *R*, which summarized the largest commonality across all measures, had the second highest test–retest reliability of all factors and measures at 0.85.

Examining test–retest reliability in isolation may be misleading because some measures could have obtained such a reliability at the expense of not capturing any interindividual differences (that is, for a “trivial” measure to which all participants respond identically, it would not be surprising to observe high test–retest reliability). Therefore, we also report each measure’s coefficient of variation as an indicator of standardized variance captured across participants. As Fig. 3 shows, the propensity measures (*M* = 0.32, HDI = 0.23 to 0.41) and behavioral measures (*M* = 0.30, HDI = 0.16 to 0.45) captured similar amounts of interindividual differences, albeit less than the frequency measures did

(*M* = 0.84, HDI = 0.45 to 1.21). That is, propensity measures did not achieve high test–retest reliability as a result of being trivial stimuli to which most participants provide the same (or similar) responses.

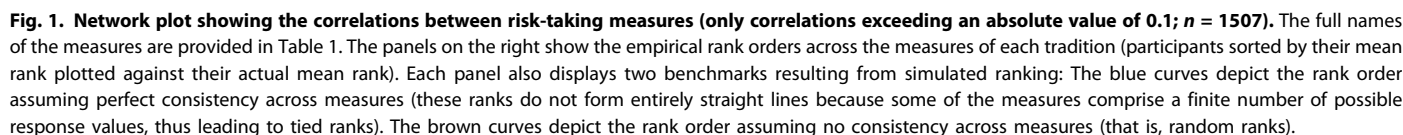
The reliability analyses illustrated that some of the measures are far from yielding error-free responses. Measurement error in the different measures may stem from changes in observable factors over time, as well as from random changes that cannot be attributed to external factors (40). The decomposition of error into these two sources has rarely been made, and with little success in attributing error to systematic changes in external variables (for example, changes in sociodemographic variables did not account for measurement error in time preference) (57). However, irrespective of its source, measurement error may provide an explanation for the partially low correlations between measures. Specifically, the reliabilities of any two measures impose an upper limit for the correlation that can be expected between them. Because reliabilities are almost never perfect, most empirical correlations are thus “attenuated” (58). To control for measurement error, we disattenuated each correlation by dividing the empirical correlation by the geometric mean of the two measures’ reliabilities and replotted Fig. 1. As to be expected, correlations between measures substantially increased (fig. S2 and table S3). However, the overall pattern with a clear gap between behavioral measures and propensity as well as frequency measures persisted. Because of the low reliabilities of some behavioral measures, disattenuating the correlations led to some very strong inflations. Thus, we did not extract latent variables from the latter, in line with previous recommendations to only disattenuate correlations between measures with reliabilities of at least 0.7 (59).

Relation of risk preference with sociodemographic, cognitive, and personality characteristics

It is beyond the scope of this article to systematically examine the various potential drivers underlying interindividual differences in risk preference. Yet, at least three categories of variables have previously been considered as such driving forces and are thus predestined to be tested for their associations with the extracted psychometric factors: socioeconomic circumstances, such as a person’s previous economic experiences, current economic situation (that is, socioeconomic status), or educational level (16, 45–47); cognitive and numerical abilities, to the extent that the evaluation of risky options requires some degree of arithmetical skills (48, 49); and personality characteristics, such as the Big Five personality traits (11, 50).

Consequently, we ran a series of Bayesian regression analyses (see Materials and Methods) to examine whether any of these variables predict interindividual differences in *R* and the other, specific factors of risk preference. Personality measures were available for 297 of the participants,

Downloaded from <http://advances.sciencemag.org/> on October 11, 2018



psychometric modeling analyses, can be summarized as follows: First, we found a substantial gap between stated and revealed preferences. This gap was observed at the level of simple correlations and rank orders, as well as in terms of the general psychometric factor R that captured shared variance across propensity measures and frequency measures but not across behavioral measures. Second, propensity measures showed substantially higher test–retest reliability across 6 months than behavioral measures, and the general factor R proved to be even more reliable across time—providing support for the idea of risk preference as a psychological construct with a certain degree of temporal stability. What does this suggest about the nature of risk preference?

At the outset, we reviewed the lack of consensus about how to conceptualize risk preference—a construct that is often regarded as a key building block of human behavior in both psychology and economics. Our results inform and enrich this discussion in several respects.

First, the present work helps to overcome the overly simplistic view of risk preference as either general or domain-specific. Current theories of intelligence and psychopathology suggest that there are both general and domain- or facet-specific components to each construct (26, 27). Akin to this theory development, our results suggest that such a view may also be most suitable to understand the nature of risk preference. Specifically, a general factor of risk preference accounted for about half

Our study evaluated the convergent validity and test–retest reliability of a range of risk-taking measures across different measurement traditions (39 risk-taking measures collected from 1507 participants). On this empirical basis, let us revisit the questions pertaining to the nature of the construct risk preference. Our findings, obtained from state-of-the-art

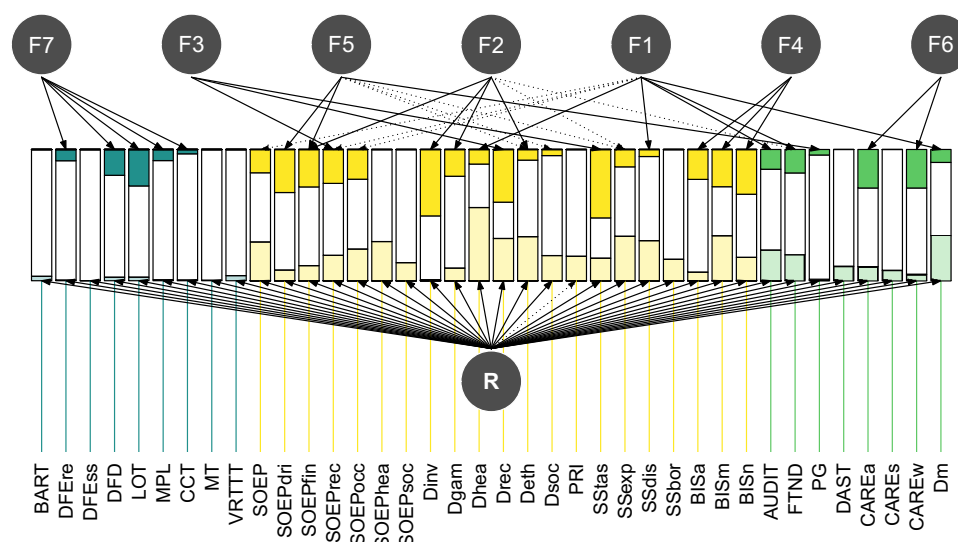


Fig. 2. Bifactor model ($n = 1507$) with all risk-taking measures, grouped by measurement tradition (Table 1). R reflects a general factor of risk preference, and F1 to F7 reflect a series of specific factors. The specific factors were formed by selecting all measures that loaded ≥ 0.25 on at least one factor in a preceding EFA with bifactor rotation. The stacked bars indicate the proportion of variance in each of the measures explained by the factors. Negative loadings are represented by dotted lines.

of the explained variance, and a series of specific factors accounted for approximately the other half. That is, after accounting for general risk preference, some specific domains (for example, investment or recreation) persist, because these may differ psychologically in terms of the risks and benefits that respondents perceive (4). This finding converges with observations that have recently been made within a single questionnaire of risk preference (36). In our findings, several of the specific factors capture variance from diverse measures (that is, scales), providing insight into some of the underlying mechanisms: For example, F3 captured recreational risk taking, and the factor structure suggests that this tendency may be triggered by a desire for “thrill and adventure seeking.” In contrast, F1 captured alcohol consumption and smoking, and the respective factor structure indicates that these behaviors occur because of problems with disinhibitory processes. In line with these insights, risk taking is considered to be a subdimension of disinhibition in psychopathology (60). Future studies may thus benefit from examining these (domain-specific) clusters of risk-taking measures in more detail to identify the sources of interindividual differences.

Second, our results speak to the important question of whether risk preference—in terms of a latent variable that is assessed by multiple and diverse measures—can be considered a stable psychological construct. Our test–retest reliability analyses indicate that R was impressively stable across time, paralleling the findings from intelligence and personality research that have obtained similar or even higher values across periods of years or decades (8, 61). Whereas the aggregation of multiple measures can be expected to reduce unsystematic error, an increase in test–retest reliability does not automatically follow and may only be observed if the captured construct indeed remains stable across time: For example, the aggregation of multiple risky lotteries in one factor (F7) did not lead to a substantial improvement in test–retest reliability. Conversely, the entity underlying the positive manifold across the various risk-taking measures, as captured by R , appears to constitute a relatively stable trait. Future work that carries this research forward (by including a comprehensive measurement of risk preference in longitudinal designs across longer periods) will be needed to thus establish the degree of temporal stability and relative predictive validity of the different types of measures discussed

here. Moreover, the (domain-)specific psychometric factors, which were orthogonal to R , turned out to have substantially lower test–retest reliability as compared to R . In line with an integrated trait-state perspective (38), these specific factors may thus rather reflect particular states, which complement a general trait (also see the next subsection).

Third and finally, we explored the relationship of different variables that have previously been suggested to be associated with or even to drive risk preference. In the investigated range of potential predictor variables, we found that risk preference, as manifested in the general factor R , most closely relates to the Big Five personality traits. Two of these traits (openness to experience and extraversion) were positively associated with R , and two traits (conscientiousness and agreeableness) were negatively associated with R . However, these associations were weak, suggesting that risk preference may reflect a partly related but independent construct. This interpretation is in line with previous research that has concluded that risk taking is sufficiently independent from personality factors and thus may constitute a separate construct (62). Furthermore, we did not find associations between socioeconomic and cognitive dimensions and the general factor R . It is important to note that we evaluated potential links between risk preference and covariates while controlling for the effects of age and sex that have shown systematic links to risk preference (11, 30). Consequently, our results represent associations of individual differences in personality that go beyond any age or sex effects. In future research, the convergent and divergent validity of both the general and specific factors of risk preference with yet other psychological traits will need to be studied systematically, for example, with multitrait-multimethod approaches (63).

The gap between propensity and behavioral measures

Previous work on the convergent validity of various measures of risk preference has been inconclusive because of its reliance on small sets of measures and the lack of psychometric approaches that allow distinguishing between general and domain- or measurement-specific variance (32, 33). Our work provides more conclusive evidence in this regard by suggesting the following main findings that deserve close consideration: the substantial convergence between the propensity

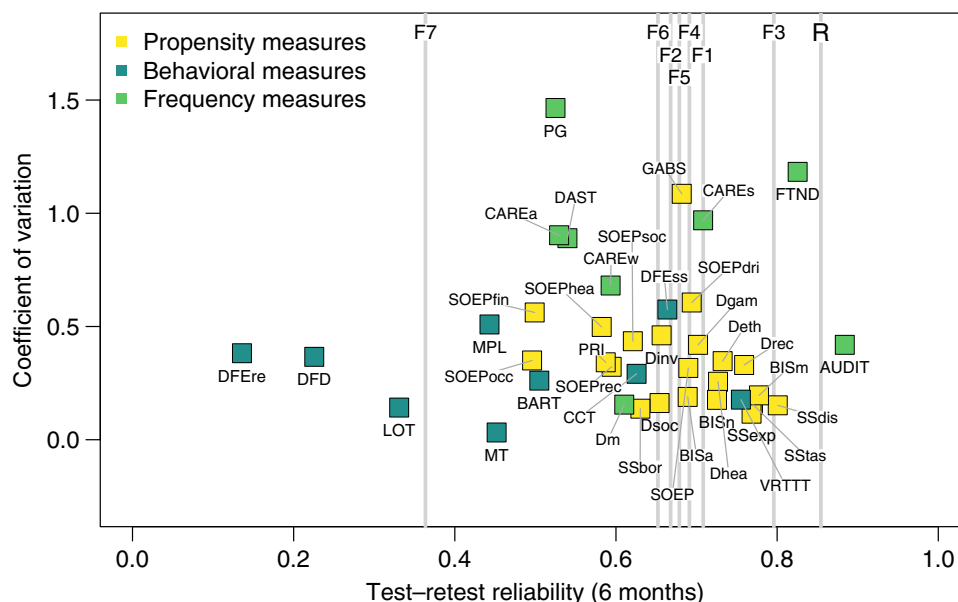


Fig. 3. Test-retest reliability and coefficient of variation across participants (that is, a standardized measure of dispersion that allows the amount of variance captured by different measures to be compared; $n = 109$). Note that we do not report the coefficients of variation for the extracted factors because the factor values were determined on the basis of standardized measures (making a comparison of the variance futile).

measures and their relatively high test–retest reliability, the lack of convergence between behavioral measures and their relatively low test–retest reliability, and the gap in convergent validity between propensity and behavioral measures. That is, unlike in research on intelligence but akin to the results from personality or psychopathology research, our results suggest a primacy of self-reports over behavioral measures, which extends both to convergent validity and to test–retest reliability. These results are of theoretical but also practical utility because the observed gap between stated and revealed preferences suggests that measures from the propensity and behavioral measurement traditions cannot be used interchangeably to capture risk preference.

When considering the relatively high convergent validity of the self-reported propensity measures, one needs to discuss the extent to which this convergence represents a reliable signal or, alternatively, undesired bias (for example, a distorted self-perception). Past and current psychological theories suggest that self-reports contain a “kernel of truth” that goes beyond bias (64) and that “intentions to perform behaviors of different kinds can be predicted with high accuracy from attitudes toward the behavior” (65). These conclusions are supported both by evidence of convergent validity between self-reports and informant reports (66) and by the predictive validity of self-reports for real-world behavior (43, 67). For example, in the neighboring field of research on impulsivity, a meta-analysis across different impulsivity measures found substantially higher correlations between different self-reports and informant reports (~ 0.5) than between behavioral tasks and informant reports (~ 0.1) (66). Similarly, self-reported impulsivity and self-control have been shown to have high predictive validity, and more so than behavioral measures, for a number of real-world outcomes, such as teenage pregnancy, drug use, and financial security (43, 67). Thus, it is difficult to dismiss the convergence of self-reported propensity measures simply as bias. Moreover, their convergence and their high temporal stability may be rooted in the way that they elicit

preferences based on respondents retrieving episodes of real-world behaviors from memory. Past psychological research suggests that people use chronically accessible and stable sources (for example, prototypical situations from everyday life) to make personality judgments (for example, regarding life satisfaction) (68). Consequently, if different propensity measures tap into similar prototypical situations and episodes from memory, then they are likely to produce consistent results that are anchored in people's actual experiences of risk and risk taking.

Concerning the low convergent validity of behavioral measures (and lower test–retest reliability as compared to propensity and frequency measures), it is important to ask what factors may contribute to this pattern of results. According to the constructed-preferences approach (69), people construct preferences online in response to a specific context defined by queries, cues, and internal states (70). Because behavioral measures differ widely in their choice architectures (in terms of their presentation format, specific instructions, or framing), they may be designed, perhaps unwittingly, to elicit different queries and cues and thus trigger diverging constructed preferences (34). Moreover, it has previously been observed that specific task manipulations can override traits (71). Behavioral tasks may thus not be suitable instruments for measuring general risk preference, but they may be indispensable if one is interested in capturing responses to specific choice architectures, such as whether risks are presented in a descriptive or experiential format (72). Equally, they may be useful for examining how different choice architectures interact with intraindividual (73) or interindividual differences (74). Hence, the various behavioral measures may capture states rather than a general and stable trait, in line with an interactional view of how both person and situation characteristics determine behavior (37, 38). That is, a potential reason for the high inconsistencies between the various behavioral measures may be that they capture various cognitive processes (beyond risk preference) and may give rise to the use of different strategies. A promising outcome of the current analysis,

however, is that there are a number of behavioral tasks involving described risks that capture common variance (F7). Given its relatively low temporal stability, this factor may capture a particular state, which is independent from the general trait (R). Future work will need to assess the extent to which the particular characteristics of these measures contribute to this convergence. Except for numeracy, which was weakly associated with F7, overall, our regression analyses did not suggest an important role for cognitive ability or numeracy in accounting for the patterns captured by the specific factors (tables S5 and S6). In sum, future work on the potential driving forces underlying interindividual differences in (domain-specific) risk preference may thus profit from an integrated trait-state perspective (38).

An additional potential explanation for the lack of consistency between different behavioral measures is of a motivational nature: It is possible that the payoff magnitudes used in the typical studies are simply too small to engage participants sufficiently and elicit actual risk preference. If this were true, then the results of our study, in which we implemented the typical payment schemes as used in the literature, would indicate a serious problem with current state-of-the-art implementation of behavioral measures. However, the potential effects of incentives in behavioral tasks can be diverse, and their usefulness remains disputed (75, 76). Moreover, consistency across measures can be achieved even in the complete absence of any incentives, as shown in the propensity measures discussed above.

Limitations

One potential limitation of this study is that we did not specifically include any extreme groups of risk takers. Future research should thus investigate the extent to which our findings generalize to more diverse populations. Moreover, the risk-taking measures were administered in a fixed order, potentially leading to sequence effects and thus inflated correlations between neighboring measures. For example, the relatively high correlations between propensity measures may have partly resulted from this effect, because the measures were administered within the same block of the experimental session. However, two arguments speak against this interpretation: First, such a sequence effect should also have emerged for the behavioral measures, which were also presented together, yet we did not find high intercorrelations between these measures. Moreover, sequence effects, if they occurred, could not explain the high test-retest reliability of propensity measures that we observed.

Implications for the measurement of risk preference

The present findings have wide-ranging scientific and practical implications: For example, previous studies examining the genetic and neural underpinnings of risk preference have used single propensity measures (10) or single behavioral measures (77, 78). However, our results suggest that future empirical work on risk preferences may profit from using several measures to reduce measurement error and, in particular, to extract variance that is shared across diverse measures of risk preference, beyond the systematic variance that is specific to the individual (domain-specific) measures. Moreover, our findings would advise against some commonly used measures on the basis of their low test-retest reliability. Overall, any improvements in establishing a reliable phenotypic construct of risk preference could foster future studies of its genetic and neural underpinnings (10, 41, 42) and real-world consequences (43, 44) that may involve small effect sizes. For example, a reliable assessment of risk preference could prove useful for designing targeted interventions in domains such as recreation and health, in which accurate personality descriptions may be key to targeting the right persons (79).

Of course, it will not always be feasible to collect large batteries of risk-taking measures in practice, such as when pooling multiple studies in the context of consortia designed to robustly examine the genetic underpinnings of certain phenotypes (80). In these cases, specific shorter subscales or even individual measures may be used as proxies for R , for example, the health risk-taking subscale of the domain-specific risk-attitude scale (that is, “Dhea”) or the general risk-taking item of the German socioeconomic panel (that is, “SOEP”), which correlated with R at 0.79 and 0.57, respectively. In practice, the use of self-reported risk preference measures is already common in some domains, such as finance (19). According to our analyses, these propensity measures exhibit desirable psychometric properties, such as convergent validity and test-retest reliability. However, one has to bear in mind that these proxy assessments do not capture the full breadth of risk preference and also cannot take advantage of the reductions in measurement error and increases in reliability, as promised by a more comprehensive assessment of the trait-like construct R .

CONCLUSION

Our work suggests that risk preference has a psychometric structure akin to other major psychological traits, such as intelligence. It involves both a general, stable component that can account for about half of the explained variance and a series of facets, each capturing more specific aspects of risk preference. These results contribute to the debate about the domain-specific nature of risk preference and indicate that this construct encompasses both general and domain-specific components. We also identified a marked gap between measures of stated and revealed preferences, suggesting that more consideration needs to be given to the measurement of risk preference. These results have implications for both basic and applied research because a solid measurement of risk preference will be needed to uncover both its biological basis and its consequences for many momentous decisions in the real world.

MATERIALS AND METHODS

Experimental design

Inclusion criteria, sample size, and sociodemographic data

We recruited 1512 healthy participants between 20 and 35 years of age who did not have any neurological (for example, epilepsy) or psychiatric (for example, depression and schizophrenia) disorders. Of these, we excluded 5 participants who did not complete four or more of the behavioral tasks or questionnaires, resulting in a total sample size of 1507 participants (746 in Basel and 761 in Berlin). This sample size established a solid participants-to-item ratio of 39:1 and thus promised to yield reliable results from the latent variable modeling (17). Note that we do not report classic significant tests for the correlations but interpret them directly as effect sizes, because the relatively large sample size would render even small correlations significant. The retest subsample ($n = 109$) was recruited from the Berlin sample. We also recruited a retest subsample from the Basel sample, but because this subsample was smaller ($n = 64$) and assessed across a shorter interval (3 months), we only report the Berlin retest subsample here (data for the Basel retest subsample are available upon request). Table S1 shows sociodemographic information.

Approvals of institutional review boards

The respective local ethics committees (Ethikkommission beider Basel, Ethikkommission des Max-Planck-Institut für Bildungsforschung, Berlin) approved the study, which was conducted in accordance with the Declaration of Helsinki. Participants received a detailed explanation of the study, and written informed consent was obtained.

Compensation

Participants received a fixed payment plus a bonus contingent on their decisions in the incentivized behavioral tasks. We took into account the different wage levels at the Swiss and German study centers to make the monetary incentives comparable: The fixed amount was based on the typical hourly wage for research assistants at the local universities, namely, 15 CHF in Basel and 10 EUR in Berlin (it took most participants 8 hours to complete the study, resulting in a fixed payment of 120 CHF or 80 EUR). In addition, participants collected bonus points in the incentivized tasks, before each of which they were informed about the conversion rate of points into CHF or EUR (we used the same scaling factor of 0.66 between study centers as for the fixed payment). Participants started with an initial bonus of 15 CHF or 10 EUR (mimicking the incentives typically used in the literature) and were informed that in the two most extreme cases, they could either maximally double or entirely lose this amount, depending on their choices. At the end of the study, one of the incentivized tasks was randomly selected, and the respective outcome was either added to or subtracted from the initial bonus.

Study schedule

The study consisted of a daylong session in one of the two laboratories, Basel or Berlin. It started at 9:00 a.m. when participants were given written information about the study (that is, inclusion and exclusion criteria, monetary compensation, duration, and data protection). Participants then provided written informed consent, completed a questionnaire tapping sociodemographic data, and rated their general and current affective state. At 9:15 a.m., participants started completing the behavioral tasks, with 10-min breaks at about 10:30 a.m. and 11:20 a.m. Before the second break, participants provided a saliva sample and again rated their current affective state. After another behavioral task, participants had a lunch break at around 12 noon. The study recommenced at 1:00 p.m. with a test of participants' working memory capacity and another rating of their current affective state. After a 10-min break, participants completed the last two behavioral tasks and another memory test. After a final 10-min break, participants completed a series of self-report questionnaires. At around 4:00 p.m., participants provided final ratings of their current affective state and were compensated for their participation. Throughout the study, a server-side framework ensured that the procedure at both study centers was identical (for example, the same task order, the same duration of breaks, etc.) and automatically loaded the different tasks and questionnaires. Time-sensitive tasks were run offline. Participants completed the study in private cubicles (a maximum of six participants were present at any time), and at least one laboratory manager was permanently available to assist participants if they needed any support. Details regarding the methods of the implemented tasks can be found in the Supplementary Materials.

Statistical analyses

Data preprocessing and robustness checks

The following measures were heavily skewed: GABS, FTND, PG, DAST, CAREa, CAREs, and CAREw. We therefore transformed them into ordinal bins (we created bins such that at least 50 participants were present in each and the number of bins was maximized, to get the best resolution—in the extreme case, this resulted in a binary classification of participants, such as smokers versus nonsmokers in the case of the FTND). The final distributions of all risk-taking measures are shown in fig. S1.

Past literature suggests that there are important age and sex differences associated with risk preference (11, 18, 30, 51–53). To ensure that

the correlations between risk-taking measures were not inflated by age, sex, or study center, we first ran linear models with these three variables as predictors of each of the risk-taking measures (for measures without a continuous scale, we ran ordinal regressions). For the main analyses, we used the resulting residuals. Of the 58,773 data points (1507 participants \times 39 measures), 82 were missing (0.14%). To avoid convergence issues in the main analyses, particularly in the latent variable modeling, we imputed these missing data points by means of Gibbs sampling using the R package mice (81). The imputation of the missing data points affected the correlations between the different measures only negligibly, with the absolute values of the correlations changing on average by 0.0007 and the most affected correlation changing by an absolute value of 0.016. All reported correlations are based on a heterogeneous correlation matrix obtained with the R package polycor (82), that is, Pearson's correlations between continuous measures, polyserial correlations between continuous and ordinal measures, and polychoric correlations between ordinal measures.

Bayesian inference statistics

The reported inference statistics for convergent validity (that is, correlational analyses and rank analyses) and temporal stability (test–retest correlations and coefficients of variation) were estimated using a Bayesian approach. Specifically, we used Markov chain Monte Carlo sampling to obtain posterior distributions, and we report their modes (which tend to converge with the medians of the empirical distributions) as measures of central tendency (denoted by “M”) and their 95% highest-density intervals as confidence intervals (denoted by “HDI”). The tests were conducted using broad default priors as implemented in the R package BESTmcmc (83). In addition to its conceptual clarity, a Bayesian approach has several advantages over classic frequentist approaches (for example, providing full distributions over means, SDs, group differences, and effect sizes). Moreover, the method renders transitive group comparisons possible and handles outliers (83).

Latent variable modeling

To test the hypothesis of a general risk-preference factor, we implemented a psychometric bifactor model, which has a long tradition in research on intelligence (84) and has recently been rediscovered as an effective method to model construct-relevant multidimensionality (27, 85, 86). Multidimensionality is modeled such that a general factor expresses itself directly across all manifest variables, and the remaining variance is partitioned into a series of orthogonal factors. Specifically, we first ran an EFA across all measures using maximum-likelihood estimation and a bifactor rotation as implemented in the R package GPArotation (87). One measure (GABS) had to be excluded before the latent variable modeling, as it hindered model convergence because of its problematic distribution. The number of factors was informed by a preceding parallel analysis on the heterogeneous correlation matrix (see above). However, the suggested number of eight factors yielded trivial factors (that is, factors on which only one measure loaded substantially) as well as cross-loadings, and we therefore reduced the number of factors to seven. Next, to determine the fit of the resulting factor structure, we implemented a CFA, estimating the factor loadings of all measures on a general risk-preference factor, as well as the factor loadings of measures that loaded at least 0.25 on any of the factors in the preceding EFA (the other loadings were fixed to 0). The model was estimated using the R package lavaan (88) with the weighted least-squares mean and variance estimator using diagonally weighted least squares and computing robust SEs (this estimator takes both continuous and ordinal measures into account). All factors were forced to be orthogonal, as defined by the standard bifactor model.

For the reduced model, we first ran an EFA without a general factor and allowed for correlated factors, to maximize the chance that shared variance between any of the measures could be extracted (that is, no bifactor but a Promax rotation). On the basis of this EFA, we then applied a more stringent threshold of 0.5 (as opposed to 0.25 in the full model) and implemented a CFA using the same method as described for the full model to assess the model fit. The reduced model is shown in fig. S3. Of the explained variance (53%), *R* still accounted for 50%, and all the specific factors together accounted for the other 50%.

Bayesian regression analyses

With a set of sociodemographic, cognitive, and personality variables, we performed Bayesian regression analyses on each of the risk-taking measures (table S5) and the psychometric factors extracted from them (table S6). The regression analyses were based on the raw values of the measures (that is, not the residuals with control for age, sex, and study center, as in the previous analyses). We used the R package *rstanarm* (89) for running Bayesian regressions and implemented the mildly informative default priors $N(0,3)$ for the intercept as well as the predictors. Hence, these regressions implement some regularization and minimize the chance of obtaining spurious results.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/3/10/e1701381/DC1>

Methods and DVs of behavioral tasks

fig. S1. Distributions of scores on risk-taking measures ($n = 1507$).

fig. S2. Network plot showing the disattenuated correlations between risk-taking measures (only correlations exceeding an absolute value of 0.1; $n = 1507$).

fig. S3. Reduced bifactor model ($n = 1507$).

table S1. Sociodemographic and related variables.

table S2. Correlations between all risk-taking measures and extracted factors.

table S3. Disattenuated correlations between all risk-taking measures and extracted factors.

table S4. Factor loadings resulting from the EFA across all risk-taking measures.

table S5. Bayesian regression analyses: Individual measures as dependent variables.

table S6. Bayesian regression analyses: Psychometric factors as dependent variables.

table S7. Decision problems used in the MPL task.

table S8. Initial lotteries used in the LOT task.

REFERENCES AND NOTES

1. T. Aven, The risk concept—Historical and recent development trends. *Reliab. Eng. Syst. Saf.* **99**, 33–44 (2012).
2. F. Ewald, in *The Foucault Effect: Studies in Governmentality*, G. Burchell, C. Gordon, P. Miller, Eds. (Harvester Wheatsheaf, 1991), pp. 197–210.
3. P. Slovic, Assessment of risk taking behavior. *Psychol. Bull.* **61**, 220–233 (1964).
4. E. U. Weber, A.-R. Blais, N. E. Betz, A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *J. Behav. Decis. Mak.* **15**, 263–290 (2002).
5. Y. Hanoch, J. G. Johnson, A. Wilke, Domain specificity in experimental measures and participant recruitment: An application to risk-taking behavior. *Psychol. Sci.* **17**, 300–304 (2006).
6. E. Yechiam, E. Ert, Risk attitude in decision making: In search of trait-like constructs. *Top. Cogn. Sci.* **3**, 166–186 (2011).
7. L. Einav, A. Finkelstein, I. Pascu, M. R. Cullen, How general are risk preferences? Choices under uncertainty in different domains. *Am. Econ. Rev.* **102**, 2606–2638 (2012).
8. I. Anusic, U. Schimmack, Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *J. Pers. Soc. Psychol.* **110**, 766–781 (2016).
9. G. J. Stigler, G. S. Becker, De gustibus non est disputandum. *Am. Econ. Rev.* **67**, 76–90 (1977).
10. D. J. Benjamin, D. Cesarini, M. J. H. M. van der Loos, C. T. Dawes, P. D. Koellinger, P. K. E. Magnusson, C. F. Chabris, D. Conley, D. Laibson, M. Johannesson, P. M. Visscher, The genetic architecture of economic and political preferences. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8026–8031 (2012).
11. A. K. Josef, D. Richter, G. R. Samanez-Larkin, G. G. Wagner, R. Hertwig, R. Mata, Stability and change in risk-taking propensity across the adult life span. *J. Pers. Soc. Psychol.* **111**, 430–450 (2016).
12. M. M. Galizzi, S. R. Machado, R. Miniati, “Temporal stability, cross-validity, and external validity of risk preferences measures: Experimental evidence from a UK representative sample,” Working paper, London School for Economics and Political Science, London, UK, 2016.
13. G. Charness, U. Gneezy, A. Imas, Experimental methods: Eliciting risk preferences. *J. Econ. Behav. Organ.* **87**, 43–51 (2013).
14. T. Schonberg, C. R. Fox, R. A. Poldrack, Mind the gap: Bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends Cogn. Sci.* **15**, 11–19 (2011).
15. A.-R. Blais, E. U. Weber, A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgm. Decis. Mak.* **1**, 33–47 (2006).
16. T. Dohmen, A. Falk, D. Huffman, U. Sunde, J. Schupp, G. G. Wagner, Individual risk attitudes: Measurement, determinants, and behavioral consequences. *J. Eur. Econ. Assoc.* **9**, 522–550 (2011).
17. A. B. Costello, J. W. Osborne, Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pract. Assess. Res. Eval.* **10**, 1–9 (2005).
18. R. Mata, A. K. Josef, R. Hertwig, Propensity for risk taking across the life span and around the globe. *Psychol. Sci.* **27**, 231–243 (2016).
19. N. Marinelli, C. Mazzoli, An insight into suitability practice: Is a standard questionnaire the answer?, in *Bank Strategy, Governance and Ratings*, P. Molyneux, Ed. (Palgrave Macmillan, London, 2011), pp. 217–245.
20. P. A. Samuelson, A note on the pure theory of consumer's behaviour. *Economica* **5**, 61–71 (1938).
21. K. C. Appelt, K. F. Milch, M. J. J. Handgraaf, E. U. Weber, The Decision Making Individual Differences Inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgm. Decis. Mak.* **6**, 252–262 (2011).
22. J. P. Allen, D. F. Reinert, R. J. Volk, The alcohol use disorders identification test: An aid to recognition of alcohol problems in primary care patients. *Prev. Med.* **33**, 428–433 (2001).
23. T. F. Heatherton, L. T. Kozlowski, R. C. Frecker, K.-O. Fagerström, The Fagerström test for nicotine dependence: A revision of the Fagerström tolerance questionnaire. *Br. J. Addict.* **86**, 1119–1127 (1991).
24. J. Brodbeck, S. Duerrenberger, H. Znoj, Prevalence rates of at risk, problematic and pathological gambling in Switzerland. *Eur. J. Psychiatry* **23**, 67–75 (2009).
25. C. Spearman, “General intelligence,” objectively determined and measured. *Am. J. Psychol.* **15**, 201–292 (1904).
26. I. J. Deary, Intelligence. *Annu. Rev. Psychol.* **63**, 453–482 (2012).
27. A. Caspi, R. M. Houts, D. W. Belsky, S. J. Goldman-Mellor, H. Harrington, S. Israel, M. H. Meier, S. Ramrakha, I. Shalev, R. Poulton, T. E. Moffitt, The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).
28. I. J. Deary, L. Penke, W. Johnson, The neuroscience of human intelligence differences. *Nat. Rev. Neurosci.* **11**, 201–211 (2010).
29. R. Plomin, I. J. Deary, Genetics and intelligence differences: Five special findings. *Mol. Psychiatry* **20**, 98–108 (2015).
30. R. Mata, A. K. Josef, G. R. Samanez-Larkin, R. Hertwig, Age differences in risky choice: A meta-analysis. *Ann. N. Y. Acad. Sci.* **1235**, 18–29 (2011).
31. L. Menkhoff, S. Sakha, Estimating risky behavior with multiple-item risk measures. *J. Econ. Psychol.* **59**, 59–86 (2017).
32. J. Berg, J. Dickhaut, K. McCabe, Risk preference instability across institutions: A dilemma. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4209–4214 (2005).
33. D. Friedman, R. M. Isaac, D. James, S. Sunder, in *Risky Curves: On the Empirical Failure of Expected Utility*, D. Friedman, R. M. Isaac, D. James, S. Sunder, Eds. (Routledge, 2014), pp. 20–53.
34. P. Crosetto, A. Filippin, A theoretical and experimental appraisal of four risk elicitation methods. *Exp. Econ.* **19**, 613–641 (2016).
35. K. P. Harden, N. Kretsch, F. D. Mann, K. Herzhoff, J. L. Tackett, L. Steinberg, E. M. Tucker-Drob, Beyond dual systems: A genetically-informed, latent factor model of behavioral and self-report measures related to adolescent risk-taking. *Dev. Cogn. Neurosci.* **25**, 221–234 (2017).
36. S. Highhouse, C. D. Nye, D. C. Zhang, T. B. Rada, Structure of the Dospert: Is there evidence for a general risk factor? *J. Behav. Decis. Mak.* **30**, 400–406 (2017).
37. A. H. Buss, Personality as traits. *Am. Psychol.* **44**, 1378–1388 (1989).
38. E. L. Hamaker, J. R. Nesselroade, P. C. M. Molenaar, The integrated trait-state model. *J. Res. Pers.* **41**, 295–315 (2007).
39. Y. Chuang, L. Schechter, Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *J. Dev. Econ.* **117**, 151–170 (2015).
40. J. Beauchamp, D. Cesarini, M. Johannesson, “The psychometric properties of measures of economic risk preferences,” Working Paper, New York University, New York, 2011.

41. C. M. Kuhnen, J. Y. Chiao, Genetic determinants of financial risk taking. *PLOS ONE* **4**, e4362 (2009).
42. S. Zhong, S. H. Chew, E. Set, J. Zhang, H. Xue, P. C. Sham, R. P. Ebstein, S. Israel, The heritability of attitude toward economic risk. *Twin Res. Hum. Genet.* **12**, 103–107 (2009).
43. T. E. Moffitt, L. Arseneault, D. Belsky, N. Dickson, R. J. Hancox, H. Harrington, R. Houts, R. Poulton, B. W. Roberts, S. Ross, M. R. Sears, W. M. Thomson, A. Caspi, A gradient of childhood self-control predicts health, wealth, and public safety. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2693–2698 (2011).
44. D. W. Belsky, T. E. Moffitt, D. L. Corcoran, B. Domingue, H. Harrington, S. Hogan, R. Houts, S. Ramrakha, K. Sugden, B. S. Williams, R. Poulton, A. Caspi, The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychol. Sci.* **27**, 957–972 (2016).
45. U. Malmendier, S. Nagel, Depression babies: Do macroeconomic experiences affect risk taking? *Q. J. Econ.* **126**, 373–416 (2011).
46. K. L. Shaw, An empirical analysis of risk aversion and income growth. *J. Labor Econ.* **14**, 626–653 (1996).
47. L. X. Zhong, J. X. Xiao, Determinants of family bond and stock holdings. *J. Financ. Couns. Plann.* **6**, 107–114 (1995).
48. T. Dohmen, A. Falk, D. Huffman, U. Sunde, Are risk aversion and impatience related to cognitive ability? *Am. Econ. Rev.* **100**, 1238–1260 (2010).
49. N. Y. Park, Domain-specific risk preference and cognitive ability. *Econ. Lett.* **141**, 1–4 (2016).
50. S. Mishra, M. L. Lalumière, Individual differences in risk-propensity: Associations between personality and behavioral measures of risk. *Pers. Individ. Differ.* **50**, 869–873 (2011).
51. I. N. Defoe, J. S. Dubas, B. Figner, M. A. G. van Aken, A meta-analysis on age differences in risky decision making: Adolescents versus children and adults. *Psychol. Bull.* **141**, 48–84 (2015).
52. J. P. Byrnes, D. C. Miller, W. D. Schafer, Gender differences in risk taking: A meta-analysis. *Psychol. Bull.* **125**, 367–383 (1999).
53. C. R. Harris, M. Jenkins, D. Glaser, Gender differences in risk assessment: Why do women take fewer risks than men. *Judgm. Decis. Mak.* **1**, 48–63 (2006).
54. T. M. J. Fruchterman, E. M. Reingold, Graph drawing by force-directed placement. *Softw. Pract. Exper.* **21**, 1129–1164 (1991).
55. M. R. Reynolds, T. Z. Keith, Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence* **35**, 267–281 (2007).
56. B. B. Lahey, B. Applegate, J. K. Hakes, D. H. Zald, A. R. Hariri, P. J. Rathouz, Is there a general factor of prevalent psychopathology during adulthood? *J. Abnorm. Psychol.* **121**, 971–977 (2012).
57. S. Meier, C. D. Sprenger, Temporal stability of time preferences. *Rev. Econ. Stat.* **97**, 273–286 (2015).
58. C. Spearman, The proof and measurement of association between two things. *Am. J. Psychol.* **15**, 72–101 (1904).
59. J. W. Osborne, Effect sizes and the disattenuation of correlation and regression coefficients: Lessons from educational psychology. *Pract. Assess. Res. Eval.* **8** (2003).
60. R. F. Krueger, K. E. Markon, The role of the DSM-5 personality trait model in moving toward a quantitative and empirically based approach to classifying personality and psychopathology. *Annu. Rev. Clin. Psychol.* **10**, 477–501 (2014).
61. I. J. Deary, L. J. Whalley, H. Lemmon, J. R. Crawford, J. M. Starr, The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence* **28**, 49–55 (2000).
62. S. V. Paunonen, D. N. Jackson, The Jackson Personality Inventory and the five-factor model of personality. *J. Res. Pers.* **30**, 42–59 (1996).
63. D. T. Campbell, D. W. Fiske, Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* **56**, 81–105 (1959).
64. S. T. McAbee, B. S. Connelly, A multi-rater framework for studying personality: The trait-reputation-identity model. *Psychol. Rev.* **123**, 569–591 (2016).
65. I. Ajzen, The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* **50**, 179–211 (1991).
66. A. L. Duckworth, M. L. Kern, A meta-analysis of the convergent validity of self-control measures. *J. Res. Pers.* **45**, 259–268 (2011).
67. J. L. White, T. E. Moffitt, A. Caspi, D. J. Bartusch, D. J. Needles, M. Stouthamer-Loeber, Measuring impulsivity and examining its relationship to delinquency. *J. Abnorm. Psychol.* **103**, 192–205 (1994).
68. U. Schimmack, E. Diener, S. Oishi, Life-satisfaction is a momentary judgment and a stable personality characteristic: The use of chronically accessible and stable sources. *J. Pers.* **70**, 345–384 (2002).
69. S. Lichtenstein, P. Slovic, Eds., *The Construction of Preference* (Cambridge Univ. Press, 2006).
70. E. J. Johnson, G. Häubl, A. Keinan, Aspects of endowment: A query theory of value construction. *J. Exp. Psychol. Learn. Mem. Cogn.* **33**, 461–474 (2007).
71. W. Mischel, E. B. Ebbesen, A. R. Zeiss, Selective attention to the self: Situational and dispositional determinants. *J. Pers.* **27**, 129–142 (1973).
72. R. Hertwig, G. Barron, E. U. Weber, I. Erev, Decisions from experience and the effect of rare events in risky choice. *Psychol. Sci.* **15**, 534–539 (2004).
73. R. Frey, R. Hertwig, J. Rieskamp, Fear shapes information acquisition in decisions from experience. *Cognition* **132**, 90–99 (2014).
74. R. Frey, R. Mata, R. Hertwig, The role of cognitive abilities in decisions from experience: Age differences emerge as a function of choice set size. *Cognition* **142**, 60–80 (2015).
75. D. Read, Monetary incentives, what are they good for? *J. Econ. Methodol.* **12**, 265–276 (2005).
76. G. W. Harrison, E. Johnson, M. M. McInnes, E. E. Rutström, Risk aversion and incentive effects: Comment. *Am. Econ. Rev.* **95**, 897–901 (2005).
77. S. M. Helfinstein, T. Schonberg, E. Congdon, K. H. Karlsgodt, J. A. Mumford, F. W. Sabb, T. D. Cannon, E. D. London, R. M. Bilder, R. A. Poldrack, Predicting risky choices from brain activity patterns. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2470–2475 (2014).
78. S. Gilaie-Dotan, A. Tymula, N. Cooper, J. W. Kable, P. W. Glimcher, I. Levy, Neuroanatomy predicts individual risk attitudes. *J. Neurosci.* **34**, 12394–12401 (2014).
79. D. Cross, H. Monks, M. Hall, T. Shaw, Y. Pintabona, E. Erceg, G. Hamilton, C. Roberts, S. Waters, L. Lester, Three-year results of the Friendly Schools whole-of-school intervention on children's bullying behaviour. *Brit. Educ. Res. J.* **37**, 105–129 (2011).
80. C. A. Rietveld, S. E. Medland, J. Derringer, J. Yang, T. Esko, N. W. Martin, H.-J. Westra, K. Shakhbuzov, A. Abdellaoui, A. Agrawal, E. Albrecht, B. Z. Alizadeh, N. Amin, J. Barnard, S. E. Baumeister, K. S. Benke, L. F. Bielak, J. A. Boatman, P. A. Boyle, G. Davies, C. de Leeuw, N. Eklund, D. S. Evans, R. Ferhmann, K. Fischer, C. Gieger, H. K. Gjessing, S. Hägg, J. R. Harris, C. Hayward, C. Holzapfel, C. A. Ibrahim-Verbaas, E. Ingelsson, B. Jacobsson, P. K. Joshi, A. Jugessur, M. Kaakinen, S. Kanoni, J. Karjalainen, I. Kolcic, K. Kristiansson, Z. Kutalik, J. Lahti, S. H. Lee, P. Lin, P. A. Lind, Y. Liu, K. Lohman, M. Loitfelder, G. McMahon, P. M. Vidal, O. Meirelles, L. Milani, R. Myhre, M.-L. Nuotio, C. J. Oldmeadow, K. E. Petrovic, W. J. Peyrot, O. Polašek, L. Quaye, E. Reinmaa, J. P. Rice, T. S. Rizzi, H. Schmidt, R. Schmidt, A. V. Smith, J. A. Smith, T. Tanaka, A. Terracciano, M. J. H. M. van der Loos, V. Vitart, H. Völzke, J. Wellmann, L. Yu, W. Zhao, J. Allik, J. R. Attia, S. Bandinelli, F. Bastardot, J. Beauchamp, D. A. Bennett, K. Berger, L. J. Bierut, D. I. Boomsma, U. Bültmann, H. Campbell, C. F. Chabris, L. Cherkas, M. K. Chung, F. Cucca, M. de Andrade, P. L. De Jager, J.-E. De Neve, I. J. Deary, G. V. Dedoussis, P. Deloukas, M. Dimitriou, G. Eiriksdóttir, M. F. Elderson, J. G. Eriksson, D. M. Evans, J. D. Faul, L. Ferrucci, M. E. Garcia, H. Grönberg, V. Gudnason, P. Hall, J. M. Harris, T. B. Harris, N. D. Hastie, A. C. Heath, D. G. Hernandez, W. Hoffmann, A. Hofman, R. Holle, E. G. Holliday, J.-J. Hottenga, W. G. Iacono, T. Illig, M.-R. Jarvelin, M. Kahonen, J. Kaprio, R. M. Kirkpatrick, M. Kowgier, A. Latvala, L. J. Launer, D. A. Lawlor, T. Lehtimäki, J. Li, P. Lichtenstein, P. Lichtner, D. C. Liewald, P. A. Madden, P. K. E. Magnusson, T. E. Mäkinen, M. Masala, M. McGue, A. Metspalu, A. Mielck, M. B. Miller, G. W. Montgomery, S. Mukherjee, D. R. Nyholt, B. A. Oostra, L. J. A. van Rooij, D. R. Van Penninx, M. Perola, P. A. Peyser, M. Preisig, K. Räikkönen, O. T. Raitakari, A. Realo, S. M. Ring, S. Ripatti, F. Rivadeneira, I. Rudan, A. Rustichini, V. Salomaa, A.-P. Sarin, D. Schlessinger, R. J. Scott, H. Snieder, B. St Pourcain, J. M. Starr, J. H. Sul, I. Surakka, R. Svento, A. Teumer, The LifeLines Cohort Study, H. Tiemeier, F. J. A. van Rooij, D. R. Van Wagoner, E. Vartiainen, J. Viikari, P. Vollenweider, J. M. Vonk, G. Waeber, D. R. Weir, H.-E. Wichmann, E. Widen, G. Willemsen, J. F. Wilson, A. F. Wright, D. Conley, G. Davey-Smith, L. Franke, P. J. F. Groenen, A. Hofman, M. Johannesson, S. L. R. Kardia, R. F. Krueger, D. Laibson, N. G. Martin, M. N. Meyer, D. Posthuma, A. R. Thurik, N. J. Timpson, A. G. Uitterlinden, C. M. van Duijn, P. M. Visscher, D. J. Benjamin, D. Cesarini, P. D. Koellinger, GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
81. S. van Buuren, K. Groothuis-Oudshoorn, Mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).
82. J. Fox, *Polycor: Polychoric and Polyserial Correlations* (2010); R package version 0.7-8.
83. J. K. Kruschke, Bayesian estimation supersedes the *t* test. *J. Exp. Psychol. Gen.* **142**, 573–603 (2013).
84. K. J. Holzinger, F. Swineford, The Bi-factor method. *Psychometrika* **2**, 41–54 (1937).
85. S. P. Reise, Invited paper: The rediscovery of bifactor measurement models. *Multivariate Behav. Res.* **47**, 667–696 (2012).
86. R. I. Jennrich, P. M. Bentler, Exploratory bi-factor analysis. *Psychometrika* **76**, 537–549 (2011).
87. C. A. Benaards, R. I. Jennrich, Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educ. Psychol. Meas.* **65**, 676–696 (2005).
88. Y. Rosseel, Lavaan: An R package for structural equation modeling. *J. Stat. Softw.* **48**, 1–36 (2012).
89. Stan Development Team, *Rstanarm: Bayesian Applied Regression Modeling via Stan* (Stan Development Team, 2016).
90. A. Falk, A. Becker, T. J. Dohmen, D. Huffman, U. Sunde, "The Preference Survey Module: A validated instrument for measuring risk, time, and social preferences," Discussion Paper No. 9674, Institute for the Study of Labor, Bonn, 2016.

91. R. B. Breen, M. Zuckerman, 'Chasing' in gambling behavior: Personality and cognitive determinants. *Pers. Individ. Differ.* **27**, 1097–1111 (1999).
 92. G. R. J. Hockey, A. J. Maule, P. J. Clough, L. Bdzola, Effects of negative mood states on risk in everyday decision making. *Cogn. Emot.* **14**, 823–855 (2000).
 93. M. Zuckerman, E. A. Kolin, L. Price, I. Zoob, Development of a Sensation-Seeking Scale. *J. Consult. Psychol.* **28**, 477–482 (1964).
 94. E. S. Barratt, Factor analysis of some psychometric measures of impulsiveness and anxiety. *Psychol. Rep.* **16**, 547–554 (1965).
 95. C. W. Lejuez, J. P. Read, C. W. Kahler, J. B. Richards, S. E. Ramsey, G. L. Stuart, D. R. Strong, R. A. Brown, Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *J. Exp. Psychol. Appl.* **8**, 75–84 (2002).
 96. J. Rieskamp, The probabilistic nature of preferential choice. *J. Exp. Psychol. Learn. Mem. Cogn.* **34**, 1446–1465 (2008).
 97. C. A. Holt, S. K. Laury, Risk aversion and incentive effects. *Am. Econ. Rev.* **92**, 1644–1655 (2002).
 98. B. Figner, R. J. Mackinlay, F. Wilkening, E. U. Weber, Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *J. Exp. Psychol. Learn. Mem. Cogn.* **35**, 709–730 (2009).
 99. G. Dutilh, J. Rieskamp, Comparing perceptual and preferential decision making. *Psychon. Bull. Rev.* **23**, 723–737 (2016).
 100. A. Hergovich, M. E. Arendasy, M. Sommer, B. Bogner, The Vienna Risk-Taking Test - Traffic. A new measure of road traffic risk-taking. *J. Individ. Differ.* **28**, 198–204 (2007).
 101. H. A. Skinner, The drug abuse screening test. *Addict. Behav.* **7**, 363–371 (1982).
 102. K. Fromme, E. C. Katz, K. Rivet, Outcome expectancies and risk-taking behavior. *Cognit. Ther. Res.* **21**, 421–442 (1997).
- Acknowledgments:** We are grateful for many helpful comments from the members of the Center for Adaptive Rationality at the Max Planck Institute for Human Development, Berlin, and the Center for Cognitive and Decision Sciences and the Center for Economic Psychology at the University of Basel. We thank S. Goss and L. Wiles for editing the manuscript.
- Funding:** This work was supported by the Swiss National Science Foundation with a grant to J.R. and R.H. (CRSII1_136227). **Author contributions:** R.F., A.P., R.M., J.R., and R.H. designed the research and wrote the paper. R.F. and A.P. implemented and conducted the study. R.F. analyzed the data. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. The full data set including a detailed codebook can be downloaded from the public repository <https://osf.io/rce7g>.
- Submitted 28 April 2017
 Accepted 13 September 2017
 Published 4 October 2017
 10.1126/sciadv.1701381
- Citation:** R. Frey, A. Pedroni, R. Mata, J. Rieskamp, R. Hertwig, Risk preference shares the psychometric structure of major psychological traits. *Sci. Adv.* **3**, e1701381 (2017).

Risk preference shares the psychometric structure of major psychological traits

Renato Frey, Andreas Pedroni, Rui Mata, Jörg Rieskamp and Ralph Hertwig

Sci Adv 3 (10), e1701381.
DOI: 10.1126/sciadv.1701381

ARTICLE TOOLS

<http://advances.sciencemag.org/content/3/10/e1701381>

SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2017/10/02/3.10.e1701381.DC1>

REFERENCES

This article cites 92 articles, 6 of which you can access for free
<http://advances.sciencemag.org/content/3/10/e1701381#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Advances* is a registered trademark of AAAS.