Supplemental Material: Through the Window of My Mind: Mapping Information Integration and the Cognitive Representations Underlying Self-Reported Risk Preference

Markus D. Steiner
University of Basel

Florian I. Seitz
University of Basel

Renato Frey
University of Basel and Princeton University

Contents

1 Inclusion Criteria

Only participants who had completed at least 500 tasks on MTurk with an approval rate of at least 95% where eligible for our studies. Participants had to pass two out of two instructional manipulation checks (IMCs) and provide ratings of at least 25 out of 100 on questions asking how focused they were and how much effort they put into the study. Moreover, when inspecting the aspects listed in the pilot study (see preregistration), we found that some participants either just entered a few letters or entered the words *no*, *nope*, or *none*. We thus decided to also exclude all aspects with fewer than four characters, or that consisted only of the words *nope* or *none*. Finally, participants had to complete the study either on a desktop computer or a laptop. In study 1, we collected data of new participants until we had data of 250 participants that fulfilled all these inclusion criteria. In study 2, the same criteria applied. Of the 164 participants who completed study 2, we had to exclude 14 because they did not meet the specified criteria. More specifically, two participants failed an IMC, one participant reported a focus of less than 25, and seven participants reported to not have completed the study on either a desktop computer or a laptop, and four participants were mistakenly included in study 2, even though they were excluded in study 1. Thus, our final sample size for study 2 was 150.

2 Sentiment Analysis

To obtain the aspects' sentiments, we used the *tidytext* R package (Silge & Robinson, 2016), which provides different methods of scoring the sentiment of words. To this end, we first removed stop words (e.g., *and*), and then explored different scoring methods to find out which method works best given the resulting data structure. More specifically, we first tried the *afinn* lexicon which provides sentiment ratings of words between -5 and 5. However, as it has only a relatively small word lexicon, we could not obtain a sentiment for 13.9% of participants and thus there were many missing values. Because of this, and in line with our preregistration, we used the *bing* lexicon, which only provides binary sentiment ratings of words (positive or negative), but has a much larger lexicon. For every aspect, we then counted the positive and negative words and used the difference as sentiment score.

3 Deviations from Preregistered Analysis Plan

In this section we describe where and why we deviated from our preregistered analysis plan. There were four relatively minor deviations occurred in the analysis of study 2.

First, we preregistered that the similarity ratings would be conducted by three independent raters who would rate the similarity of all possible aspect pairs (4,286 pairs). To complete this work the raters would have been able to split the ratings over a maximum of four consecutive days. We recruited workers on Amazon MTurk for this task, but it turned out that many of them dropped out. We thus decided to split the workload and split the 4,286 aspect pairs in 21 groups of about 200 aspects, which were each rated by three raters. We thus obtained data of 63 workers (instead of only three, as originally planned). The workers had to pass at least 6 out of 8 IMCs to ensure a high data quality.

Second, we adapted the way we calculated the proportion of overlapping aspects. We preregistered that the number of aspects listed in study 1 would be the upper bound of possible overlapping aspects (i.e., the denominator in the fraction to obtain the proportion of overlap) and that two aspects would be considered to overlap if the mean similarity rating was at least four. As this approach sometimes led to values exceeding one, we adapted the way of obtaining the proportion of overlap in two respects. First, we set the threshold to be more conservative, more specifically, to five. Second, we used the smaller number of the number of aspects listed in study 1 and study 2, respectively, as denominator.

Third, we preregistered that we would use a linear model (with gaussian distribution and identity link function) to test the relation between aspect stability and the stability of self-reported risk preferences. This model identified no credible association (b = -1.21, , 95% CI: [-2.79, 0.36]). However, as both the outcome and the predictor variable were absolute difference scores, the errors were not normally distributed and the model thus likely biased. Therefore, we decided to run (and report in the main text) a gamma regression (i.e., with gamma distribution and log-link function). Note that this analysis found no credible association between aspect stability and the stability of self-reported risk preferences either. Relatedly, we explored additional ways of aggregating the similarity ratings across raters. These are reported in the additional analyses section below.

Fourth and finally, we preregistered that we would test the association between evidence stability and the stability of self-reported risk preferences with a linear model using the absolute difference scores between study 1 and study 2 of the two variables. With this model, evidence stability was a credible predictor of the stability of the reported risk preferences ($b = 0.04$, 95% CI [0.03, 0.05], $r_s = .41$). However, because only a directional test (i.e., without taking the absolute value of the differences) yielded normally distributed errors of the model, we also ran a directional Bayesian linear model in which, for better interpretability, we also z-standardized both difference scores. Using this model, we again found a credible association between evidence stability and the stability of the reported risk preference ($\beta = 0.63$, 95% CI [0.50, 0.75], $r_s = .45$). We decided to report the second model as some of the necessary assumptions in the first model were not met.

4 Details on Reported Analyses

4.1 Description of the Cognitive Models

Table 1

Properties of Implemented Models

Property	QT	SUM	EXT	FIRST	LAST	VUM
Compensatory	X	X				X
Weight of evidence	X					
Strength of evidence		X	X	X	X	X
Serial position	X			X	X	X
N free parameters	4	0	0	0	0	1

Note: QT = Query Theory. SUM = Sum of evidence. EXT = Most extreme evidence. FIRST = First aspect. LAST = Last aspect. VUM = Value updating model. Compensatory = Whether a model is compensatory (X) or noncompensatory.

We systematically selected a set of cognitive models that make different assumptions about how the three properties of evidence determine a judgment. The initial set of models was sampled from the literature on judgment and decision making and implemented both *compensatory* and *noncompensatory* processes. In compensatory processes, a single piece of information does not necessarily determine a judgment by itself because it may be outweighed cumulatively by other pieces of information (Gigerenzer & Goldstein, 1996, 1999; Hertwig, Barron, Weber, & Erev, 2006; Payne, Bettman, & Johnson, 1988). Conversely, in noncompensatory processes as few as "one good reason" (potentially identified in a lexicographic way) is considered (e.g., Gigerenzer & Goldstein, 1996, 1999). Models that resulted in highly correlated predictions and/or were not recovered well in an extensive simulation analysis (see preregistration) were excluded from the initial set of selected models. An overview of the final model space is provided in Table 1. In what follows we describe the retained models in detail.

4.1.1 Query theory (QT). QT (E. J. Johnson, Häubl, & Keinan, 2007; Weber et al., 2007) describes a compensatory process assuming that the *number of aspects* retrieved from memory (i.e., the weight of evidence) is closely related to people's judgments, irrespective of these aspects' strengths of evidence: That is, the more aspects supporting a positive judgment are retrieved (i.e., pro-aspects), the higher the resulting judgment. Conversely, the more aspects supporting a negative judgment are retrieved (i.e., contra-aspects), the lower the resulting judgment. Moreover, it is assumed that the order in which the aspects are retrieved matters, such that aspects retrieved earlier are more important; that is, receive a larger weight. To measure the clustering of pro- and contra-aspects and thus the influence of the aspects' serial positions, studies testing QT typically make use of the standardized median rank difference (SMRD), which is defined as $2(MR_p - MR_c)/n$, where MR_c is the median rank of contra-aspects, MR_p is the median rank of pro-aspects, and n is the total number of aspects retrieved for a given judgment (E. J. Johnson et al., 2007).

We formalize QT as a linear model that predicts the self-reported risk preference (RP) by

$$RP = \beta_I + \beta_p N_p + \beta_c N_c + \beta_{SMRD} SMRD \tag{1}$$

where β_I is the intercept, and β_c, β_p, and β_{SMRD} are the coefficients for the number of contra-aspects, number of pro-aspects, and the SMRD. Due to the free parameters, QT can in principle emulate diverse other models that only take into account the weight of evidence (e.g., a model that assumes that the judgment depends on the tally of the direction of the retrieved aspects; cf. Jarecki & Wilke, 2018).

4.1.2 Sum of evidence (SUM). SUM (a weighted additive model; cf. Payne et al., 1988) also assumes a compensatory process and takes into account the strengths of evidence of all listed aspects, yet it does not incorporate any order effects. Specifically, SUM predicts the self-reported risk preference to be the cumulative strength of evidence of all retrieved aspects. As we operationalized the strength of evidence from -50 to 50, the cumulative evidence reduces to the sum of the aspects' strength of evidence. That is, SUM can be formalized as

$$RP = \sum_{i=1}^{n} e(a_i) \tag{2}$$

where $e(a_i)$ is the strength of evidence of aspect a_i and n is the total number of aspects for a given judgment.

4.1.3 Most extreme evidence (EXT). EXT assumes a noncompensatory lexicographic process, inspired by the *take-the-best* heuristic (Gigerenzer & Goldstein, 1996, 1999). Specifically, EXT relies on the aspect with the most extreme strength of evidence (i.e., with the highest absolute difference from 0, likely the most salient aspect) to predict the self-reported risk preference. EXT can be formalized as

$$RP = \begin{cases} \max e(a_i), & \text{if } a_i \geq 0 \\ \min e(a_i), & \text{if } a_i < 0 \end{cases} \tag{3}$$

where $e(a_i)$ is the strength of evidence of the aspect a_i. In the case of ties, that is a pro- and a contra-aspect with equally extreme strengths of evidence, EXT predicts a neutral risk preference.

4.1.4 First aspect's strength of evidence (FIRST). FIRST (for similar implementations are also known as take-the-first; see Jarecki & Wilke, 2018; J. G. Johnson & Raab, 2003) is another model assuming a noncompensatory process, and implements a very strong primacy effect—that is, the assumption that information retrieved at the beginning of a sequence of aspects predominantly impacts a person's judgment (cf. Hogarth & Einhorn, 1992; Murdock, 1962). Specifically, FIRST uses the strength of evidence of the first aspect listed to predict the self-reported risk preference. FIRST can be formalized as

$$RP = e(a_1) \tag{4}$$

where $e(a_1)$ is the strength of evidence of the first aspect a_1.

4.1.5 Last aspect's strength of evidence (LAST). Conversely, LAST implements a very strong recency effect—the assumption that information retrieved at the end of a sequence of aspects predominantly impacts a person's judgment (cf. Hogarth & Einhorn, 1992; Murdock, 1962). LAST uses the strength of evidence of the last aspect listed to predict the self-reported risk preference. LAST can be formalized as

$$RP = e(a_n) \tag{5}$$

where $e(a_n)$ is the strength of evidence of the last (nth) aspect (a_n).

4.1.6 Value updating model (VUM). Finally, the VUM (i.e., an instance of a fractional-adjustment model, also known as anchoring-and-adjustment model; Hertwig et al., 2006; Hogarth & Einhorn, 1992) assumes a compensatory process and was originally developed as a sequential learning model (see also Sutton & Barto, 1998; Yechiam & Busemeyer, 2005). Applied to the current context, VUM assumes that people rely on the strength of evidence of each retrieved aspect, to then update their current judgment of their own risk preference "by computing a weighted average of the previously estimated value and the value of the most recently experienced outcome" (Hertwig et al., 2006, p. 84). The VUM can be formalized as

$$RP_t = (1 - \omega_t) RP_{t-1} + \omega_t e(a_t) \tag{6}$$

where RP_t is the predicted risk preference after t retrieved aspects, ω_t is the weight given to the aspect t, RP_{t-1} is the estimated risk preference before integrating aspect t, and $e(a_t)$ denotes the strength of evidence of aspect a_t. The weight ω_t conferred to $e(a_t)$ is given by

$\omega_t = (1/t)^\phi$. Thus, VUM has one free parameter (ϕ) with which it can gradually account for primacy ($\phi > 1$) or recency effects ($\phi < 1$). With $\phi = 1$ it reduces to the arithmetic mean.

4.1.7 Estimated Parameters for Query Theory. The positive regression coefficient for the SMRD in query theory (QT; averaged over the five coefficients obtained from the cross-validation procedure, mean b = 1.76, 95% CI: [1.36, 2.17]) indicated that there was a clustering of pro- and contra-aspects, respectively. Moreover, 90.4% of the SMRDs were either one or negative one, and 82.4% of the participants listed only pro- or only contra-aspects. Thus, this clustering was present very strongly. Furthermore, in line with previous findings (E. J. Johnson et al., 2007), there was a positive correlation between participants' SMRD and their self-reported risk preference (r_s = .80), indicating that participants with a higher risk preferences tended to first report pro-aspects. The other parameters of QT indicated that the number of pro-aspects listed by a participant was credibly associated with higher risk-taking propensity ratings (mean b = 0.49, 95% CI: [0.31, 0.68]), and that the number of contra-aspects listed by a participant was credibly associated with lower risk-taking propensity ratings (mean b = -0.25, 95% CI: [-0.41, -0.08]), indicating that also the weight of evidence is a valid predictor of risk preference (but note that the weight of evidence and the strength of evidence correlate strongly, as the former is a binarized version of the latter).

4.1.8 Robustness Check of Study 1. To test the robustness of the findings of study 1, we reran the analyses of study 1 with the data of study 2. In the aspect listing, again, a majority of 121 participants (78.6%) listed between one and four aspects (M = 3.58; range: 1 - 13). Of these listed aspects, again 63% were contra-aspects. Moreover, most participants (N = 119; 77.3%) listed either only contra-aspects or only pro-aspects, and only a minority listed aspects of both types. Through this, the strength of evidence ratings within participants were rather stable with an intraclass correlation of .71.

The cognitive models were again very successful with r_s ranging from .72 to .81 (note that these values resulted from out-of-sample predictions using the independent hold-out sets). Specifically, the models achieved the following performances: $r_{s,VUM}$ = .82, $r_{s,EXT}$ = .73, $r_{s,QT}$ = .75, $r_{s,FIRST}$ = .72, $r_{s,LAST}$ = .74, and $r_{s,SUM}$ = .74. Moreover, these models again clearly outperformed the three reference models (see also Figure 1 in the article). However, the r_s values of the models were somewhat lower compared to those obtained in study 1, yet still substantial.

4.2 Mixed Models Used to Quantify the Differences Between Pro- and Contra-Aspects

To quantify the effects of pro- versus contra-aspects, we ran generalized linear mixed effects models with the content variables (e.g., the rating whether an aspect involved an active choice or a passive experience) as outcome variables, and aspect valence (pro or contra) as dummy coded predictor. Additionally, we included by-subjects random slopes and intercepts, that is, we specified the maximal model (see, Barr, Levy, Scheepers, & Tily, 2013). Note that these analyses were not preregistered. We found that the estimated probability for an aspect describing a personal experience was credibly higher for the pro-aspects (p = .854) than the contra-aspects (p = .474; b = 1.87, 95% CI: [0.82, 3.17]). Moreover, the estimated probability that an aspect would involve an active choice rather than a passive experience was higher for pro- (p = .913) than for contra-aspects (p = .713; $b = 1.44$, 95% CI: [0.81, 2.22]). For the other two comparisons we found no credible effects of pro- versus contra-aspects (social comparison: $b = 0.61$, 95% CI: [-0.61, 1.91]; controllable vs. uncontrollable: $b = 0.24$, 95% CI: [-0.52, 1.08]). To quantify the difference in the reported frequencies between pro- an contra-aspects, we ran an ordinal mixed effects model using the *brms* R package (Bürkner, 2017), with the frequency categories as outcome variable, and aspect valence as dummy coded predictor. We again included by-subjects random slopes and intercepts. There was no credible effect of pro- or contra-aspects on the frequency categories ($b = 0.28$, 95% CI: [-0.24, 0.81]).

To also quantify the differences in the sentiment for pro- and contra-aspects, we then ran a linear mixed effects model with the sentiment as outcome variable and the aspect valence as dummy coded predictor. Moreover, we included by-subjects random slopes and random intercepts. The analysis showed a credible effect of the aspect valence indicating that pro-aspects, on average, had a higher sentiment than contra-aspects ($b = 0.60$, 95% CI: [0.40, 0.78]).

4.3 Priors Used in the Regression Models

Here we report the priors used in our Bayesian regression analyses. In all regression analyses we relied on the default priors of *rstanarm* and *brms* (Bürkner, 2017; Goodrich, Gabry, Ali, & Brilleman, 2018). That is, for all linear regressions we used, the prior for the intercept was $\mathcal{N}(0, 10)$, for the coefficients the prior was $\mathcal{N}(0, 2.5)$. In the case of the models with a gamma distribution and log link function there is an additional shape parameter that had the prior $exp(1)$. For the additional parameters of the (ordinal) mixed models, we refer to the R code for the detailed priors (also here we used the default priors).

4.4 Robustness Check of Modeling Results Using Participants Who Listed More Than One Aspect

In the case where only one aspect is listed, many of the models (SUM, FIRST, EXT, and LAST) make identical predictions. Thus, to test the robustness of our modeling analysis, we reran the cross-validation procedure presented in study 1 of the main text, but including only the data of participants who had listed at least two aspects (N = 223 in study 1 and N = 137 in study 2). The results of this robustness check were very similar to the original analysis with all participants, that is, the VUM still clearly outperformed the second best model (see Figure S1).

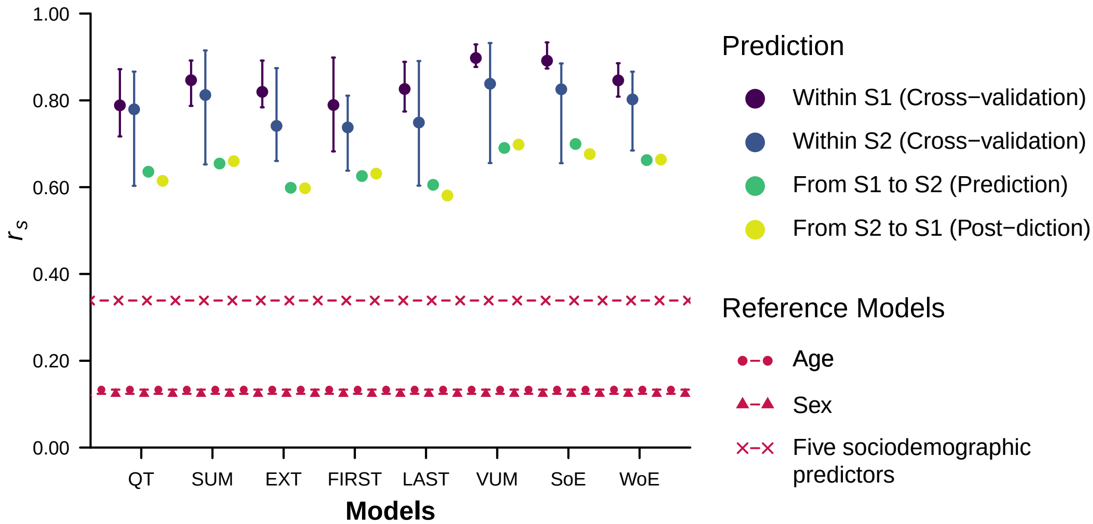

Figure S1. Spearman correlations between self-reported risk preference and the different model predictions. QT = Query theory; SUM = Sum of evidence; EXT = Most extreme evidence; FIRST = First aspect's evidence; LAST = Last aspect's evidence; VUM = Value updating model. SoE = Ordinal regression model with the average strength of evidence per participant as predictor. WoE = Ordinal regression model with the average weight of evidence per participant as predictor. Whiskers depict the range of r_s in the five folds of the cross-validation within studies. For the pre-/post-diction across studies, the models only used the aspects participants listed in one study to pre-/post-dict their risk preferences in the other study. "Five sociodemographic predictors" = Reference model using age, sex, years of education, income, and employment status as predictors. All reference models were implemented separately for study 1 and study 2 and their respective r_ss averaged for this plot.

4.5 Parameter Recovery of VUM and QT

To test whether the parameters of the two models including free parameters—VUM and QT—were recoverable, we simulated participants' risk preferences based on different parameter values. We then fitted the models and compared whether the recovered parameters were in line with the ones used to fit the data.

4.5.1 Parameter recovery of QT. The QT parameters could be recovered almost perfectly (see Figure S2). The correlation between the simulated and the recovered data was $r = 1$. Note however, that for this parameter recovery analysis, we relied on an OLS implementation, rather than a Bayesian implementation of the models, as the latter would have taken exceedingly long to run. However, as we relied on the weakly informative default priors in our Bayesian implementation in the modeling analyses, this yields essentially the same result.

4.5.2 Parameter recovery of the VUM. The VUM parameters could not be properly recovered (see Figure S3). That is, although we varied ϕ between 0.1 and 3, the recovered parameters were always around 1.03. The correlation between the simulated data and the recovered predictions was $r = .98$. That is, even though the parameters used to simulate the data varied strongly—implementing strong versions of recency and primacy

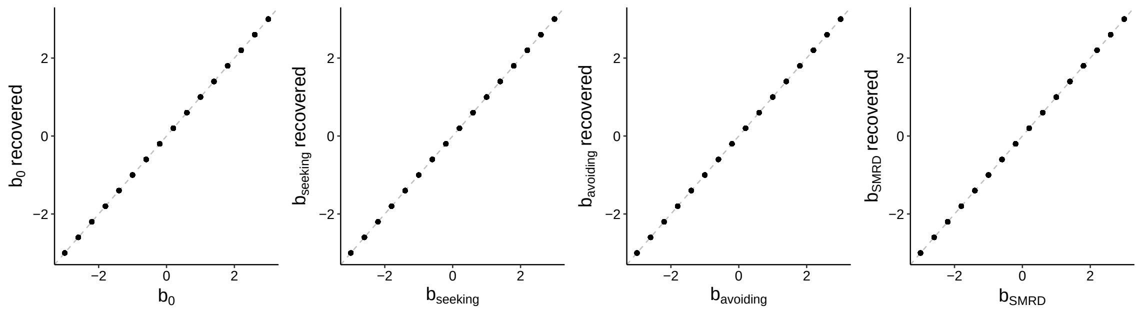

Figure S2. Model recovery analysis for query theory. b_0 = Intercept. $b_{seeking}$ = Coefficient for the number of pro-aspects. $b_{avoiding}$ = Coefficient for the number of contra-aspects. b_{SMRD} = Coefficient for the SMRD. The model is described in detail above in section 4.1.

effects—this had hardly any effects on the model predictions. This is likely due to the fact that most participants listed (a) between one and four aspects, and (b) either only pro-aspects or contra-aspects.

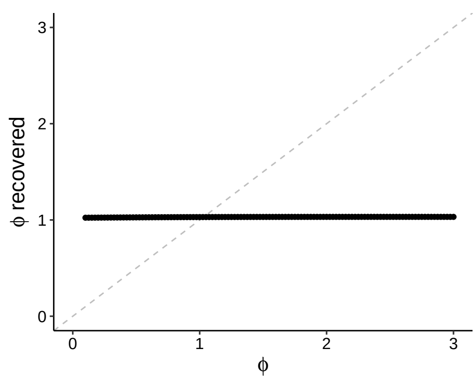

Figure S3. Model recovery analysis for the value updating model. ϕ = The parameter implementing a recency effect (< 1) or primacy effect (> 1). The model is described in detail above in section 4.1.

4.5.3 Testing the impact of different values of ϕ. Our modeling analysis indicated that the VUM performs best. Moreover, the obtained value of ϕ across the different cross-validation-folds suggested a slight recency effect. To test the robustness of this interpretation, we tested post-hoc, whether different values of ϕ—including ones suggesting a strong recency effect ($\phi << 1$), as well as ones suggesting a strong primacy effect ($\phi >> 1$)—led to striking differences in the model performance. To this end, we tested the model performance for values of $\phi \in \{0.001, 0.002, \ldots, 2.500\}$ using the complete data of study 1. Figure S4 shows the r_s values as a function of ϕ. It is clear that only rather extreme values of ϕ really influence the r_s values and even then the r_s is large enough for the VUM to outperform most other models. This is likely due to the limited number of aspects listed by participants such that only extreme values strongly affect the predictions. We conclude that the VUM mainly shows the importance of using a central tendency measure to aggregate *all* the strengths of evidence reported rather than relying only on certain aspects such as, for example, EXT, FIRST, or LAST.

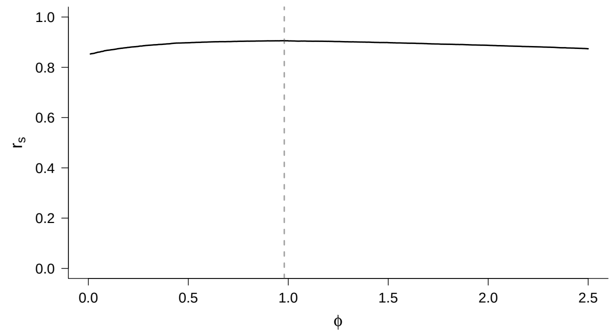

Figure S4. r_s values between the self-reported risk preferences and the predictions of the value updating model as a function of the ϕ parameter. The dashed vertical line indicates the ϕ value that maximizes the r_s across the complete data of study 1.

4.6 MDS Plots of the Identical Predictions of the Ordinal Regression Models

In line with the tournament approach to evaluate different models and their similarity when making predictions, we plotted the multi-dimensional scaling (MDS) solutions mapping the five considered ordinal regression models based on their proportions of identical predictions (see Figure S5).

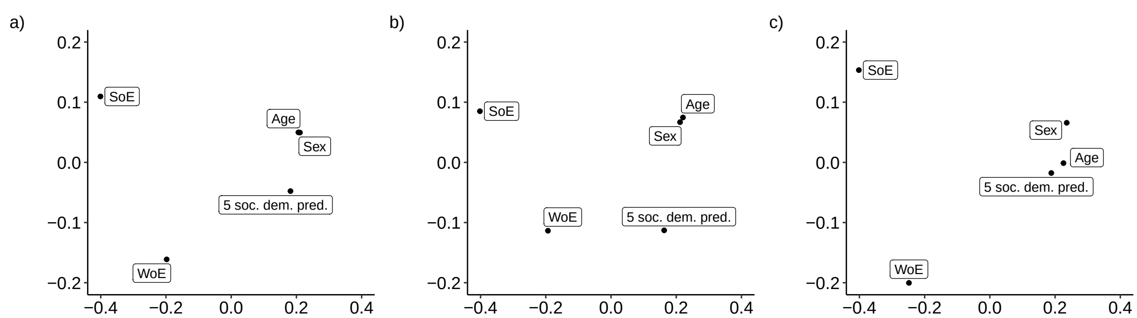

Figure S5. MDS solutions mapping the proportion of identical predictions of the five considered ordinal regression models. a) = Within study 1. b) = Cross-study prediction from study 1 to study 2. c) = Cross-study prediction from study 2 to study 1. SoE = Model with the average strength of evidence per participant as predictor. WoE = Model with the average weight of evidence per participant as predictor. Age = Model using age as predictor. Sex = Model using sex as predictor. 5 soc. dem. pred. = Model using age, sex, years of education, income, and employment status as predictors.

4.7 Tables of the Proportions of Identical (Correct) Model Predictions

In line with the tournament approach to evaluate different models and their similarity when making predictions, we also compared the proportions of identical model predictions, as well as the proportions of identical correct model predictions. Table 2 shows these values

Table 2

Proportion of Identical Model Predictions (Below Diagonal) and of Identical Correct Model Predictions (Above Diagonal) From Models Fit in Study 1.

Model	SoE	WoE	Age	Sex	5 soc-dem
Study 1					
SoE	-	0.29	0.18	0.17	0.18
WoE	0.65	-	0.20	0.19	0.20
Age	0.38	0.54	-	0.20	0.18
Sex	0.38	0.54	0.97	-	0.18
5 soc. dem. pred.	0.39	0.56	0.79	0.79	-
Study 2					
SoE	-	0.23	0.18	0.17	0.19
WoE	0.61	-	0.19	0.19	0.21
Age	0.36	0.49	-	0.20	0.21
Sex	0.37	0.48	0.95	-	0.20
5 soc. dem. pred.	0.37	0.50	0.87	0.83	-

Note: Proportion of identical (below diagonal), and identical correct predictions (above diagonal) of the five considered ordinal regression models. SoE = Model with the average strength of evidence per participant as predictor. WoE = Model with the average weight of evidence per participant as predictor. Age = Model using age as predictor. Sex = Model using sex as predictor. 5 soc-dem = Model using age, sex, years of education, income, and employment status as predictors.

for the models fit on the data from study 1. Table 3 shows these values for the predictions from study 1 to study 2, and the post-dictions from study 2 to study 1.

Table 3

Proportion of Identical Model Predictions (Below Diagonal) and of Identical Correct Model Predictions (Above Diagonal), in the Cross Study Analyses.

Model	SoE	WoE	Age	Sex	5 soc-dem
Fitting in study 1, prediction to study 2					
SoE	-	0.21	0.15	0.15	0.15
WoE	0.68	-	0.19	0.19	0.20
Age	0.37	0.54	-	0.21	0.19
Sex	0.38	0.55	0.97	-	0.19
5 soc. dem. pred.	0.40	0.60	0.77	0.77	-
Fitting in study 2, prediction to study 1					
SoE	-	0.25	0.17	0.17	0.19
WoE	0.61	-	0.19	0.18	0.2
Age	0.35	0.49	-	0.20	0.21
Sex	0.35	0.45	0.93	-	0.20
5 soc. dem. pred.	0.38	0.51	0.88	0.82	-

Note: Proportion of identical (below diagonal), and identical correct predictions (above diagonal) of the five considered ordinal regression models. The models were fitted with the data from study 1 (study 2), to then generate out-of-sample predictions of the self-reported risk preferences from study 2 (study 1) with the aspects of study 1 (study 2) . SoE = Model with the average strength of evidence per participant as predictor. WoE = Model with the average weight of evidence per participant as predictor. Age = Model using age as predictor. Sex = Model using sex as predictor. 5 soc-dem = Model using age, sex, years of education, income, and employment status as predictors.

5 Additional Analyses

5.1 Distribution of Self-Reported Risk Preferences

There were signs for a bimodal distribution of self-reported risk preferences, in line with observations of previous investigations (Dohmen et al., 2011; Frey, Pedroni, Mata, Rieskamp, & Hertwig, 2017), yet slightly stronger so (see Figure S6). A possible explanation for this amplification could lie in the deliberate reasoning of participants in our study: The explicit listing of (positive and negative) aspects may have led to slightly more extreme ratings. Furthermore, other patterns from the past literature could also be replicated in our dataset, such as that risk preferences were negatively associated with age (r_s = -.23 in study 1, and r_s = -.19 in study 2; Josef et al., 2016; Mamerow, Frey, & Mata, 2016.

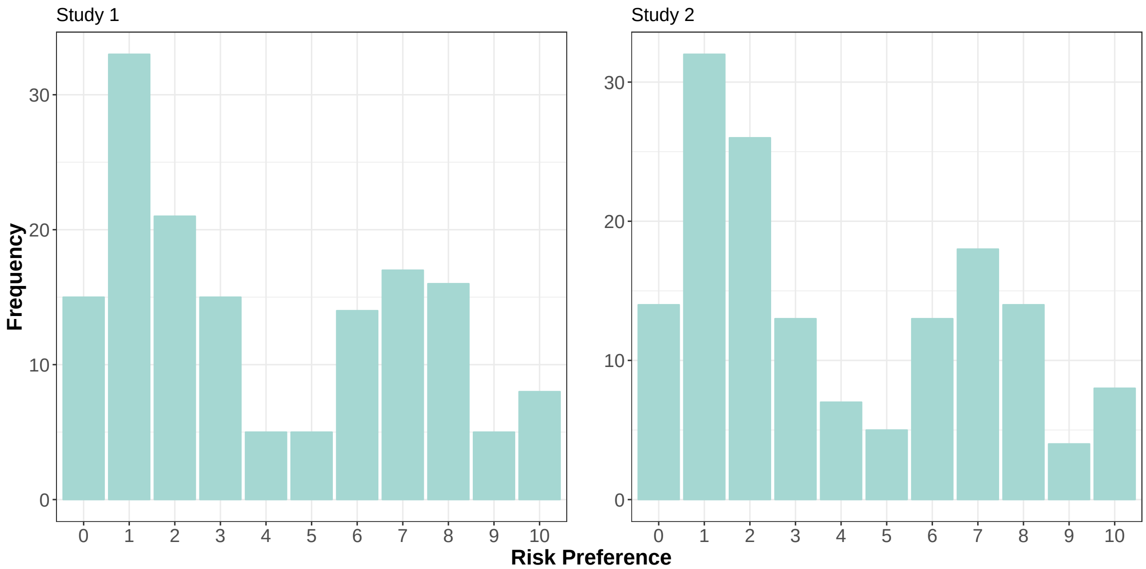

Figure S6. Distributions of self-reported risk preference in study 1 (left) and study 2 (right).

5.2 Distribution of the Strength of Evidence Ratings

The strengths of evidence were clearly bimodally distributed, that is, there were almost no aspects with a “neutral” strength of evidence, such that they did not speak in favor or against taking a risk (see Figure S7).

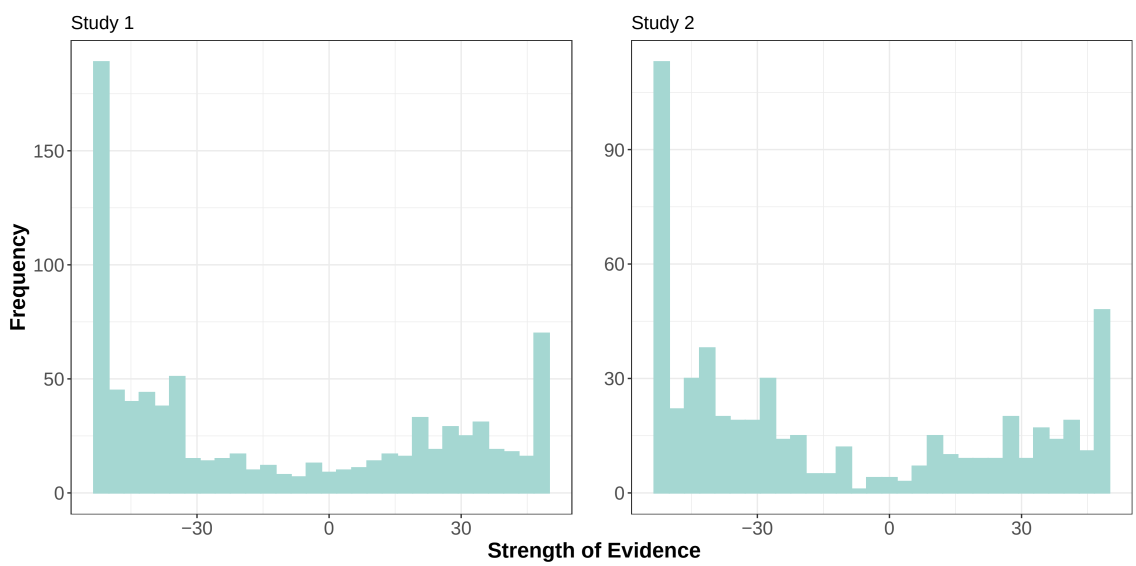

Figure S7. Distributions of the aspects’ strengths of evidence in study 1 (left) and study 2 (right).

5.3 Distribution of Similarities

The mean similarity per aspect tended to be negatively skewed but otherwise relatively normally distributed (see Figure S8). The similarity ratings were obtained for all possible combinations of aspects.

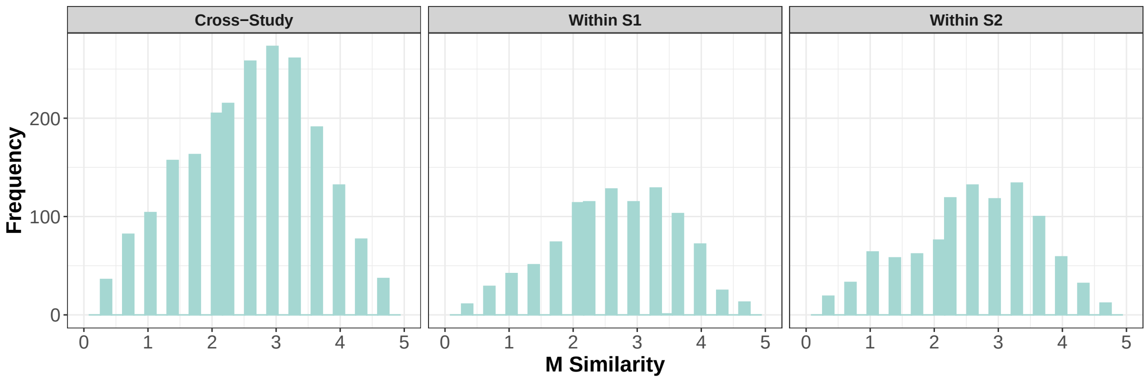

Figure S8. Distributions of the mean similarity per aspect across studies (left), within study 1 (middle), and within study 2 (right).

5.4 Additional Ways to Aggregate Similarity Ratings

We ran additional robustness checks on the relation between aspect stability and the stability of self-reported risk preferences. Specifically, we computed the similarities in two additional ways, (a) by taking the median instead of the mean, and (b) by identifying the rater pair with the highest Kendall's W of all three combinations of each rater triplet, and by then computing the average similarity only from the ratings of these two raters.

When we used the median to aggregate the similarity ratings, the average similarities within and between studies were very similar, though slightly higher compared to those reported in the main text (between studies: $M = 2.80$; within study 1: $M = 2.81$; within study 2: $M = 2.74$). From these similarity ratings, we then also computed the proportion of overlap. On average, we again found only a rather small proportion of overlaps (proportion of overlap across studies: .09). Using these overlap measures, we found no credible association between the absolute difference in the risk-taking propensity between study 1 and study 2, and the proportion of overlap between study 1 and study 2 in a participant's listed aspects using a GLM with a gamma distribution and a log link function (b = -1.15, 95% CI [-2.88, 1.06]).

When we used the mean of the ratings of the two raters with the highest Kendall's W to aggregate the similarity ratings, the average similarities within and between studies were again very similar, though this time slightly lower compared to those reported in the article (between studies: $M = 2.67$; within study 1: $M = 2.72$; within study 2: $M = 2.64$). From these similarity ratings, we then again computed the proportion of overlap. On average, we again found only a rather small proportion of overlaps (proportion of overlap across studies: .07). Using these overlap measures, we again did not find a credible association between the absolute difference in the risk-taking propensity between study 1 and study 2, and the proportion of overlap between study 1 and study 2 in a participant's listed aspects using a

GLM with a gamma distribution and a log-link function (b = -2.52, 95% CI [-4.99, 0.40]). In sum, whereas we found clear associations between evidence stability and the stability of self-reported risk preferences, these robust checks indicate that further investigations are needed to clarify the impact of aspect stability.

5.5 Additional Analyses of the Aspects' Contents

5.5.1 Selection of domains and properties. To learn more about the aspects' contents, we rated each aspect concerning the domain(s) it captures, as well as for which of a set of properties it mentioned or alluded to. We relied on domains suggested in one of the most popular domain-specific risk-taking questionnaires (Blais & Weber, 2006; Weber, Blais, & Betz, 2002), those suggested in the SOEP (e.g., Dohmen et al., 2011), as well as those of an evolutionary domain-specific risk-taking questionnaire (Wilke et al., 2014). Based on this, and by collapsing domains suggested by more than one of the questionnaires, we ended up with 19 different domains (see Table 4). The selection of properties was mostly theory driven based on previous literature, yet also included a data-driven part, based on what had commonly been described in the aspects (see Table 5 for some key references for the theory-driven properties).

On the one hand, the domains provide information about external attributes of the behaviors, such as what kind of behavior was reported. On the other hand, the properties provide information about the lower level, internal attributes of the behaviors, such as how large the gains or losses involved where, or whether the behavior was the result of a state of need—that is, for example, a risk was taken because only by taking a risk could a goal be reached—or from a state of ability—that is, for example, a risk was taken because the agent could stomach a potential loss and could thus take the risk for the chance of a large gain (see e.g., Mishra, Barclay, & Sparks, 2017). The properties reflect six different categories that involve both stable traits and situational characteristics (states). The categories and respective entries are listed in Table 5. Note, however, that the selection of domains and properties, although also driven by previous research, is largely arbitrary. However, we tried to be inclusive in that we included properties suggested by various theoretical accounts.

Table 4
Domains Aspects were Checked for

Domain	Key References
Ethical	Blais and Weber (2006); Weber et al. (2002)
Financial	Blais and Weber (2006); Dohmen et al. (2011); Rolison and Shenton (2019); Weber et al. (2002)
Health/ Safety	Blais and Weber (2006); Rolison and Shenton (2019); Weber et al. (2002)
Recreational/ Leisure	Blais and Weber (2006); Dohmen et al. (2011); Rolison and Shenton (2019); Weber et al. (2002)
Social	Blais and Weber (2006); Rolison and Shenton (2019); Weber et al. (2002)
Faith	Dohmen et al. (2011)
Traffic	Dohmen et al. (2011)

(continued)

Domain	Key References
Occupation	Dohmen et al. (2011)
General statement	Dohmen et al. (2011)
Between-group competition	Wilke et al. (2014)
Within-group competition	Wilke et al. (2014)
Status/power	Wilke et al. (2014)
Environmental exploration	Wilke et al. (2014)
Food selection	Wilke et al. (2014)
Food acquisition	Wilke et al. (2014)
Parent-offspring conflict	Wilke et al. (2014)
Kinship	Wilke et al. (2014)
Mate attraction	Wilke et al. (2014)
Mate retention	Wilke et al. (2014)

Note: Domains used when rating the aspects. The key references are not exhaustive and merely provide an arbitrary selection.

Table 5

Properties Aspects were Checked for

Property/Domain	Key References
Outcomes	
Positive outcomes	Atkinson (1957); Bernoulli (1738); Carver and White (1994); Coombs (1969); Fromme, Katz, and Rivet (1997); Gray (1987); Kahneman and Tversky (1979); Loxton and Dawe (2001); Mishra et al. (2017); Rode, Cosmides, Hell, and Tooby (1999); Sitkin and Pablo (1992); von Neuman and Morgenstern (1947); Weber and Milliman (1997)
Probability of positive outcomes	Atkinson (1957); Bernoulli (1738); Kahneman and Tversky (1979); Lopes (1984); March and Shapira (1987); Mishra et al. (2017); von Neuman and Morgenstern (1947)
Negative outcomes	Atkinson (1957); Bernoulli (1738); Carver and White (1994); Coombs (1969); Fromme et al. (1997); Gray (1987); Hillier and Morrongiello (1998); Kahneman and Tversky (1979); Loxton and Dawe (2001); Mishra et al. (2017); Rode et al. (1999); Sitkin and Pablo (1992); von Neuman and Morgenstern (1947)
Probability of negative outcomes	Atkinson (1957); Bernoulli (1738); Kahneman and Tversky (1979); Lopes (1984); March and Shapira (1987); Mishra et al. (2017); von Neuman and Morgenstern (1947)

(continued)

Property/Domain	Key References
Long-term perspective	Bickel and Marsch (2001); Green, Myerson, and Ostaszewski (1999); Kirby (1997); Mitchell and Wilson (2010); Reynolds (2006)
Short-term perspective	Bickel and Marsch (2001); Green et al. (1999); Kirby (1997); Mitchell and Wilson (2010); Reynolds (2006)
Goal/state	
Need	Adler et al. (1994); Bernoulli (1738); Capaldi, Stoolmiller, Clark, and Owen (2002); Kahneman and Tversky (1979); Lopes (1984); Marmot, Shipley, and Rose (1984); Mishra et al. (2017); Mishra and Lalumière (2010); Rode et al. (1999); Schoemaker (1993); Tversky and Kahneman (1992); Twenge, Catanese, and Baumeister (2002)
Ability	Anderson and Galinsky (2006); Bernoulli (1738); Kahneman and Tversky (1979); Lopes (1984); Mishra et al. (2017); Schoemaker (1993); Tversky and Kahneman (1992)
Keep status quo	Kahneman and Tversky (1979); Lopes (1984); Mishra et al. (2017); Tversky and Kahneman (1992)
Change status quo	Kahneman and Tversky (1979); Lopes (1984); Mishra et al. (2017); Tversky and Kahneman (1992)
Signaling	Jellison and Riskind (1970); Morrongiello and Lasenby-Lessard (2007)
Cultural role and personality	
Experience/habit	Malmendier and Nagel (2011, 2015); March and Shapira (1987); Morrongiello and Lasenby-Lessard (2007); Sitkin and Pablo (1992)
Social norm/pressure	Morrongiello and Lasenby-Lessard (2007); Sitkin and Pablo (1992)
Personality	Frey et al. (2017); Nicholson, Soane, Fenton-O'Creevy, and Willman (2005)
Religion/fate	*Based on aspects*
Affect	
Positive affect	Frey, Hertwig, and Rieskamp (2014); Loewenstein, Weber, Hsee, and Welch (2001); Morrongiello and Lasenby-Lessard (2007)
Thrill	Kloep, Güney, Çok, and Simsek (2009); Zuckerman (2002)
Regret	Bell (1982, 1985); Loomes and Sugden (1982); Mellers, Schwartz, Ho, and Ritov (1997)

(continued)

Property/Domain	Key References
Fear	Cohn, Engelmann, Fehr, and Maréchal (2015); Fischhoff, Slovic, Lichtenstein, Read, and Combs (1978); Frey et al. (2014); Guiso, Sapienza, and Zingales (2018); Lerner, Gonzalez, Small, and Fischhoff (2003); Loewenstein et al. (2001); Morrongiello and Lasenby-Lessard (2007); Slovic (1987)
Other negative affect	Fessler, Pillsworth, and Flamson (2004); Lerner et al. (2003); Loewenstein et al. (2001)
Life-history related variables	
Age	Frey, Richter, Schupp, Hertwig, and Mata (2020); Mamerow et al. (2016); Mata, Josef, and Hertwig (2016); Wang, Kruger, and Wilke (2009)
Sex	Byrnes, Miller, and Schafer (1999); Frey et al. (2020); Wang et al. (2009)
Parental status	Wang et al. (2009)
Birth order	Wang et al. (2009)
Number of siblings	Wang et al. (2009)
Reproductive goal	Wang et al. (2009)
Subjective life expectancy	Wang et al. (2009)
Family	*Based on aspects*
SES	Amir, Jordan, and Rand (2018); Griskevicius, Tybur, Delton, and Robertson (2011)
Other	
Relativize (depends on situation)	*Based on aspects*
Nonword	*Category to classify semantically invalid aspects*

Note: Properties used when rating the aspects. The key references are not exhaustive and merely provide an arbitrary selection.

5.5.2 Rating procedure.

The ratings of the aspects' strength of evidence where performed on a visual analog scale with the same labels as used in study 1. For the different domains and properties, it was then coded whether they were applicable for a given aspect (1) or not (0). We then used a majority rule in the codings, that is, if at least two raters coded an aspect to contain information about, for example, the magnitude of the positive outcomes involved, this was coded as 1, otherwise as 0. We rated a subsample of 300 randomly selected of the total of 857 aspects listed in study 1 (about one-third).

To validate these ratings, we additionally rated the 60 items of the domain-specific risk-taking scale (Blais & Weber, 2006) and the evolutionary risk scale (Wilke et al., 2014), to test whether the domains could be recovered. Indeed in 13 of the 15 distinct domains suggested by these two scales, more than half of the respective items were correctly recovered, and in eight of the 15 domains all items were correctly recovered (overall 50 out of 60 items—83%—were correctly "recovered").

We descriptively analyzed the domain and property ratings, gauging which properties and domains occurred how frequently. Moreover, we tested whether the external raters could recover the strengths of evidence reported by our participants. To this end, we correlated the average strength of evidence across the three external raters per aspect with the one indicated by the respective participant. Finally, we tested the proportion of agreement in the aspect valence (i.e., the proportion of correctly "recovered" pro- and contra-aspects).

5.5.3 Mentions of the different properties. Figure S9 provides an overview of the proportion of aspects in which the different properties were mentioned.

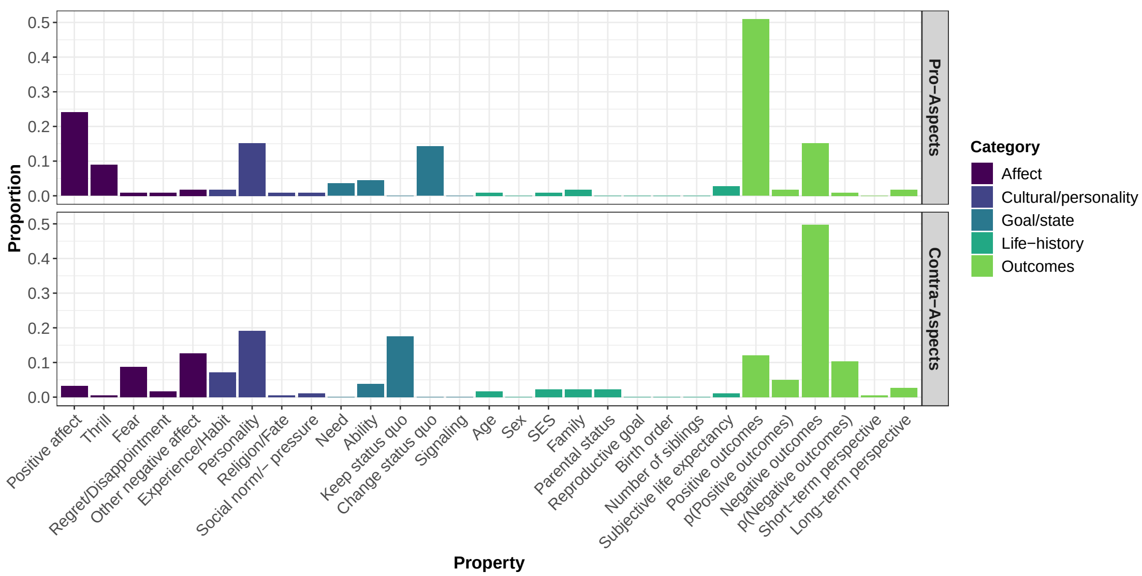

Figure S9. Proportions of aspects in which the different properties are mentioned, separate for pro-aspects (top) and contra-aspects (bottom).

5.6 Comparison of Participants Who Completed the Retest With Those Who Did Not

We ran several post-hoc analyses to test whether the participants who completed both studies were similar to those who only completed study 1. This was to exclude sample composition as potential influential factor. To this end, we splitted participants from study 1 in two subsamples, (a) those who had completed both study 1 and study 2, and (b) those who had only completed study 1. We ran Bayesian regression analyses, where we entered different demographic and study specific parameters as dependent variables and a subsample indicator as dummy coded predictor variable. We tested the group differences in age, proportion female, years of education, self-reported risk preference (the response to the SOEP general risk item), the average strength of evidence, the number of aspects listed, and the average sentiment across all aspects listed by a participant. The regression model with proportion female was run with a binomial family and logistic link function, the others with a gaussian family and identity link function. On average, participants who completed both studies were slightly older and listed more aspects, but otherwise were comparable to those who had only completed study 1 (see Table 6).

Table 6

Group differences between participants who had completed both studies and those who had only completed study 1

Dependent variable	Group difference (regression coefficient b)
Age	**2.90 [0.16, 5.40]**
Years of education	0.53 [-0.23, 1.26]
Sex	0.21 [-0.30, 0.72]
Risk preference	-.30 [-1.07, 0.51]
Mean strength of evidence	-1.08 [-5.41, 3.05]
N aspects listed	**0.73 [0.26, 1.15]**
Mean sentiment	-0.06 [-0.29, 0.15]

Note: Group differences are the median and 95% CI of the posterior. Credible differences are indicated in bold.

6 References

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health: The challenge of the gradient. *American Psychologist*, *49*(1), 15–24. doi: 10.1037/0003-066X.49.1.15

Amir, D., Jordan, M. R., & Rand, D. G. (2018). An uncertainty management perspective on long-run impacts of adversity: The influence of childhood socioeconomic status on risk, time, and social preferences. *Journal of Experimental Social Psychology*, *79*, 217–226. doi: 10.1016/j.jesp.2018.07.014

Anderson, C., & Galinsky, A. D. (2006). Power, optimism, and risk-taking. *European Journal of Social Psychology*, *36*(4), 511–536. doi: 10.1002/ejsp.324

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*, 359–372. doi: 10.1037/h0043445

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi: 10.1016/j.jml.2012.11.001

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, *30*(5), 961–981. doi: 10.1287/opre.30.5.961

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations Research*, *33*(1), 1–27. doi: 10.1287/opre.33.1.1

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

Bickel, W. K., & Marsch, L. A. (2001). Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction*, *96*(1), 73–86. doi: 10.1046/j.1360-0443.2001.961736.x

Blais, A.-R., & Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, *1*, 33–47. doi: 10.1037/t13084-000

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10.18637/jss.v080.i01

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

Capaldi, D. M., Stoolmiller, M., Clark, S., & Owen, L. D. (2002). Heterosexual risk behaviors in at-risk young men from early adolescence to young adulthood: Prevalence, prediction, and association with STD contraction. *Developmental Psychology*, *38*(3), 394–406. doi: 10.1037/0012-1649.38.3.394

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of personality and social psychology*, *67*(2), 319.

Cohn, A., Engelmann, J., Fehr, E., & Maréchal, M. A. (2015). Evidence for countercyclical risk aversion: An experiment with financial professionals. *American Economic Review*, *105*(2), 860–85.

Coombs, C. H. (1969). Portfolio theory: A theory of risky decision making. In G. T. Guilbaud (Ed.), *La decision.* Paris: Centre National de la Recherche Scientifique.

Dohmen, T. J., 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*, 522–550. doi: 10.1111/j.1542-4774.2011.01015.x

Fessler, D. M., Pillsworth, E. G., & Flamson, T. J. (2004). Angry men and disgusted women: An evolutionary approach to the influence of emotions on risk taking. *Organizational Behavior and Human Decision Processes*, *95*, 107-123. doi: 10.1016/j.obhdp.2004.06.006

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences*, *9*(2), 127–152. doi: 10.1007/BF00143739

Frey, R., Hertwig, R., & Rieskamp, J. (2014). Fear shapes information acquisition in decisions from experience. *Cognition*, *132*, 90–99. doi: 10.1016/j.cognition.2014.03.009

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. doi: 10.1126/sciadv.1701381

Frey, R., Richter, D., Schupp, J., Hertwig, R., & Mata, R. (2020). Identifying robust correlates of risk preference: A systematic approach using specification curve analysis. *Journal of Personality and Social Psychology.* doi: 10.1037/pspp0000287

Fromme, K., Katz, E. C., & Rivet, K. (1997). Outcome expectancies and risk-taking behavior. *Cognitive Therapy and Research*, *21*(4), 421–442. doi: 10.1023/A:1021932326716

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. doi: 10.1037/0033-295X.103.4.650

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 75–95). New York: Oxford University Press.

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2018). *rstanarm: Bayesian applied regression modeling via Stan.* Retrieved from http://mc-stan.org/ (R package version 2.17.4)

Gray, J. A. (1987). *The psychology of fear and stress* (Vol. 5). Cambridge, UK: Cambridge University Press.

Green, L., Myerson, J., & Ostaszewski, P. (1999). Amount of reward has opposite effects on the discounting of delayed and probabilistic outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 418–427. doi: 10.1037/0278-7393.25.2.418

Griskevicius, V., Tybur, J. M., Delton, A. W., & Robertson, T. E. (2011). The influence of mortality and socioeconomic status on risk and delayed rewards: A life history theory approach. *Journal of Personality and Social Psychology*, *100*(6), 1015–1026. doi: 10.1037/a0022403

Guiso, L., Sapienza, P., & Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics*, *128*(3), 403–421.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2006). The role of information sampling in risky choice. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 72–91). New York: Cambridge University Press.

Hillier, L. M., & Morrongiello, B. A. (1998). Age and Gender Differences in School-Age Children's Appraisals of Injury Risk. *Journal of Pediatric Psychology*, *23*(4), 229–238. doi: 10.1093/jpepsy/23.4.229

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1-55. doi: 10.1016/0010-0285(92)90002-J

Jarecki, J. B., & Wilke, A. (2018). Into the black box: Tracing information about risks related to 10 evolutionary problems. *Evolutionary Behavioral Sciences*, *12*, 230–244. doi: 10.1037/ebs0000123

Jellison, J. M., & Riskind, J. (1970). A social comparison of abilities interpretation of risk-taking behavior. *Journal of Personality and Social Psychology*, *15*(4), 375-390. doi: 10.1037/h0029601

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: A query theory of value construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 461–474. doi: 10.1037/0278-7393.33.3.461

Johnson, J. G., & Raab, M. (2003). Take the first: Option-generation and resulting choices. *Organizational Behavior and Human Decision Processes*, *91*, 215–229. doi: 10.1016/S0749-5978(03)00027-X

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*, 430–450. doi: 10.1037/pspp0000090

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. doi: 10.2307/1914185

Kirby, K. N. (1997). Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General*, *126*, 54–70. doi: 10.1037/0096-3445.126.1.54

Kloep, M., Güney, N., Çok, F., & Simsek, Ö. F. (2009). Motives for risk-taking in adolescence: A cross-cultural study. *Journal of Adolescence*, *32*(1), 135–151. doi: 10.1016/j.adolescence.2007.10.010

Lerner, J. S., Gonzalez, R. M., Small, D. A., & Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychological Science*, *14*(2), 144–150. doi: 10.1111/1467-9280.01433

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267-286. doi: 10.1037/0033-2909.127.2.267

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, *92*(368), 805–824. doi: 10.2307/2232669

Lopes, L. L. (1984). Risk and distributional inequality. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 465-485. doi: 10.1037/0096-1523.10.4.465

Loxton, N. J., & Dawe, S. (2001). Alcohol abuse and dysfunctional eating in adolescent girls: The influence of individual differences in sensitivity to reward and punishment. *International Journal of Eating Disorders*, *29*(4), 455–462. doi: 10.1002/eat.1042

Malmendier, U., & Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics*, *126*(1), 373–416. doi: 10.1093/qje/qjq004

Malmendier, U., & Nagel, S. (2015). Learning from inflation experiences. *The Quarterly Journal of Economics*, *131*(1), 53–87. doi: 10.1093/qje/qjv037

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*, 711–723. doi: 10.1037/pag0000124

March, J. G., & Shapira, Z. (1987). Managerial perspectives on risk and risk taking. *Management science*, *33*(11), 1404–1418. doi: 10.1287/mnsc.33.11.1404

Marmot, M. G., Shipley, M. J., & Rose, G. (1984). Inequalities in death—specific explanations of a general pattern? *The Lancet*, *323*(8384), 1003–1006. doi: 10.1016/S0140-6736(84)92337-7

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. doi: 10.1177/0956797615617811
Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science*, *8*(6), 423–429. doi: 10.1111/j.1467-9280.1997.tb00455.x
Mishra, S., Barclay, P., & Sparks, A. (2017). The relative state model: Integrating need-based and ability-based pathways to risk-taking. *Personality and Social Psychology Review*, *21*(2), 176-198. doi: 10.1177/1088868316644094
Mishra, S., & Lalumière, M. L. (2010). You can't always get what you want: The motivational effect of need on risk-sensitive decision-making. *Journal of Experimental Social Psychology*, *46*(4), 605–611. doi: 10.1016/j.jesp.2009.12.009
Mitchell, S. H., & Wilson, V. B. (2010). The subjective value of delayed and probabilistic outcomes: Outcome size matters for gains but not for losses. *Behavioural Processes*, *83*(1), 36–40. doi: 10.1016/j.beproc.2009.09.003
Morrongiello, B. A., & Lasenby-Lessard, J. (2007). Psychological determinants of risk taking by children: An integrative model and implications for interventions. *Injury Prevention*, *13*, 20-25. doi: 10.1136/ip.2005.011296
Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488. doi: 10.1037/h0045106
Nicholson, N., Soane, E., Fenton-O'Creevy, M., & Willman, P. (2005). Personality and domain-specific risk taking. *Journal of Risk Research*, *8*(2), 157–176. doi: 10.1080/1366987032000123856
Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552. doi: 10.1037/0278-7393.14.3.534
Reynolds, B. (2006). A review of delay-discounting research with humans: Relations to drug use and gambling:. *Behavioural Pharmacology*, *17*(8), 651–667. doi: 10.1097/FBP.0b013e3280115f99
Rode, C., Cosmides, L., Hell, W., & Tooby, J. (1999). When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory. *Cognition*, *72*(3), 269–304. doi: 10.1016/S0010-0277(99)00041-4
Rolison, J. J., & Shenton, J. (2019). How much risk can you stomach? Individual differences in the tolerance of perceived risk across gender and risk domain. *Journal of Behavioral Decision Making*, bdm.2144. doi: 10.1002/bdm.2144
Schoemaker, P. J. H. (1993). Determinants of risk-taking: Behavioral and economic views. *Journal of Risk and Uncertainty*, *6*(1), 49–73. doi: 10.1007/BF01065350
Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in r. *The Journal of Open Source Software*, *1*(3), 1–3. doi: 10.21105/joss.00037
Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the determinants of risk behavior. *Academy of Management Review*, *17*, 9–38. doi: 10.5465/amr.1992.4279564
Slovic, P. (1987). Perception of risk. *Science*, *236*, 280–285. doi: 10.1126/science.3563507
Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, MA: MIT press.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. doi: 10.1007/BF00122574
Twenge, J. M., Catanese, K. R., & Baumeister, R. F. (2002). Social exclusion causes self-defeating behavior. *Journal of Personality and Social Psychology*, *83*, 606–615. doi: 10.1037//0022-3514.83.3.606
von Neuman, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (Second ed.). Princeton, NJ: Princeton University Press.
Wang, X. T., Kruger, D. J., & Wilke, A. (2009). Life history variables and risk-taking propensity. *Evolution and Human Behavior*, *30*(2), 77–84. doi: 10.1016/j.evolhumbehav.2008.09.006
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. doi: 10.1002/bdm.414

Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice: A query-theory account. *Psychological Science*, *18*, 516–523. doi: 10.1111/j.1467-9280.2007.01932.x

Weber, E. U., & Milliman, R. A. (1997). Perceived risk attitudes: Relating risk perception to risky choice. *Management Science*, *43*(2), 123–144. doi: 10.1287/mnsc.43.2.123

Wilke, A., Sherman, A., Curdt, B., Mondal, S., Fitzgerald, C., & Kruger, D. J. (2014). An evolutionary domain-specific risk scale. *Evolutionary Behavioral Sciences*, *8*(3), 123–141. doi: 10.1037/ebs0000011

Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review*, *12*, 387–402. doi: 10.3758/BF03193783

Zuckerman, M. (2002). *Sensation Seeking and Risky Behavior.* Binghamton, NY: Maple-Vail Press.