Personality and Individual Differences 189 (2022) 111485

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Individual differences in decision-making: A test of a one-factor model of rationality

Vincent Berthet [a,*], David Autissier [b], Vincent de Gardelle [c]

[a] Université de Lorraine, 2LPN, F-54000 Nancy, France, and Centre d'Économie de la Sorbonne, CNRS UMR 8174, France
[b] IAE Paris Est Créteil, Institut de Recherche en Gestion, EA 2354, France
[c] CNRS and Paris School of Economics, France

ARTICLE INFO

Keywords:
Decision-making
Heuristics and biases
Rationality
Individual differences
Measurement

ABSTRACT

The study of individual differences in rational decision-making has led to two close streams of research. While the study of scores to the Adult-Decision Making Competence (A-DMC) tasks has provided evidence in favor of a general decision-making competence (DMC) factor, studies investigating individual differences in performance on heuristics and biases tasks have challenged a one-factor model of rationality. Assuming that heuristics and biases are part of DMC and considering that the A-DMC assesses just a few of them, the aim of the present study was to test whether a general DMC factor still emerges when adding four heuristics and biases tasks to the six A-DMC tasks, while ensuring satisfactory levels of score reliability. Exploratory factor analyses revealed that while performance on the A-DMC tasks can be reasonably aggregated into a general DMC measure, a two-factor model provided the best statistical and conceptual fit of the 10 tasks combined, the two factors reflecting Mindware gaps and Contaminated mindware.

1. Introduction

Grounded on the seminal work of Stanovich and West (1998, 2000), the last two decades have seen a growing number of studies investigating individual differences in the tendency to deviate from normative standards of decision-making and rational judgment (Stanovich et al., 2011). Two close streams of research have emerged in this endeavor. The first one has primarily focused on individual differences in performance on tasks from the so-called heuristics and biases tradition (Kahneman et al., 1982) to investigate rational thinking (Stanovich & West, 1998, 2000). The second one has investigated individual differences in performance on various decision-making tasks (including heuristics and biases ones) to address the issue of a general decision-making competence (DMC) defined as "the ability to make better decisions, as defined by decision-making principles posited by models of rational choice" (Bruine de Bruin et al., 2020, p. 1). While heuristics and biases might be viewed as part of DMC, these two streams of research have led to divergent results regarding a one-factor model of rational decision-making.

1.1. Individual differences in rational decision-making: One or multiple factors?

On the one hand, studies that explored the correlations between performance on heuristics and biases tasks have provided mixed results. Stanovich and West (1998) found a positive manifold of correlations between scores to disparate heuristics and biases tasks (e.g., belief bias in syllogistic reasoning, base-rate use, covariation detection, hypothesis testing, outcome bias), suggesting that a single rationality factor could explain the substantial amount of variance shared among these tasks (though Stanovich and colleagues did not report results from factor analysis). However, the several subsequent factor analytic studies reported evidence of multiple-factor solutions, referring to various theoretical taxonomies of heuristics and biases.

Weaver and Stewart (2012) found that a two-factor model provided the best fit of data correlations between scores to 10 judgment tasks (e.g., the Linda problem, base-rate, judging baseball team success based on five team features). The two factors were the ability to make accurate judgments based on multiple cues (correspondence) and the ability to make coherent judgments based on probabilities comparison and combination (coherence). Using two equivalent versions of a set of 13 heuristics and biases tasks (Study 1), Aczel et al. (2015) showed that a four-

* Corresponding author at: Université de Lorraine, UFR SHS-Nancy, 23 Boulevard Albert 1er, BP 60446, 54001 Nancy Cedex, France.
E-mail address: vincent.berthet@univ-lorraine.fr (V. Berthet).

https://doi.org/10.1016/j.paid.2021.111485
Received 16 May 2021; Received in revised form 10 December 2021; Accepted 21 December 2021

factor structure fitted best the data (with relatively low model-fit indices), but the robustness of this solution was questionable as the factor-loading pattern varied greatly between the two versions of the tasks. Teovanović et al. (2015) reported that a two-factor solution accounted for a small amount (17%) of the total variance of scores to seven heuristics and biases tasks. Belief bias and outcome bias loaded highly onto the first factor which was viewed as Normative irrationality (the tendency to make predictably irrational judgments). As it primarily accounted for the anchoring effect and hindsight bias, the second factor was viewed as Ecological rationality, that is, the ability to move one's initial answers towards subsequently given external information. In a recent and valuable attempt, Ceschi et al. (2019) sampled as much as 17 heuristics and biases tasks based on empirical and theoretical taxonomies. Exploratory and confirmatory factor analyses supported a three-factor solution including Mindware gaps (e.g., representativeness heuristic, availability heuristic), Valuation biases (e.g., optimism bias, time discounting), and Anchoring and Adjustment (e.g., anchoring heuristic, reference price).

On the other hand, the stream of research on DMC has provided more positive evidence in favor of a one-factor model. Parker and Fischhoff (2005) selected seven decision-making components (applying decision rules, consistency in risk perception, recognizing social norms, resistance to framing, overconfidence, resistance to sunk costs, and axiom of path independence) and designed tasks allowing for their reliable measurement in young adults, forming the Youth Decision-Making Competence (Y-DMC). These tasks were thought to be representative of four fundamental aspects of decision-making: assessing beliefs (e.g., consistency in risk perception); assessing values (e.g., recognizing social norms); integrating beliefs and values (e.g., applying decision rules); and having a meta-cognitive understanding of one's abilities, resources, and constraints (e.g., overconfidence) (Edwards, 1954; Raiffa, 1968). Bruine de Bruin et al. (2007) adapted these tasks for adults (A-DMC) and improved their reliability (one of the seven original tasks, path independence, was later removed because of low reliability and validity). The six A-DMC measures show adequate reliability with regard both to internal consistency and temporal stability. A single factor was able to explain 25% of the variance in the Y-DMC tasks and 30% of the variance in the A-DMC tasks with all tasks having loadings of at least 0.30, supporting the claim that performance in these tasks can be reasonably aggregated into a general DMC measure (but see Blacksmith, Behrend, Dalal, & Hayes,2019, for opposite results).

This state of the art is somewhat reminiscent of the single vs. multiple factors debate, which occurred in several research areas in psychology. In particular, early studies on the structure of intelligence have opposed a single-factor theory (*g*) powered by Spearman and the Primary Mental Abilities theory of Thurstone (Gustafsson, 1984). More recently, research on individual differences in risk preference has debated the same issue (Frey et al., 2017). In fact, risk preference can be measured either with behavioral decision tasks (revealed preferences) or self-report surveys (stated preferences). Results revealed that these measures were weakly correlated, and that while the former suffer from poor construct validity, a general factor of risk preference (*R*) emerged from stated preferences (Frey et al., 2017; Millroth et al., 2020).

1.2. Methodological considerations

Given the evidence of a general DMC factor, and assuming that heuristics and biases are part of DMC, one might expect the first stream of research to point out a general rationality factor as well. In fact, the discrepancy between the findings might be related to two methodological considerations. On the one hand, it is worth noting that the set of tasks used has been inconsistent across the studies, which seems problematic with regard to the issue of data dimensionality. This fluctuating sampling of tasks may be one of the primary causes of the discrepancies in the results (Stankov, 2017). Besides, one could expect that a higher number of tasks would result in a higher number of factors extracted. Indeed, the two studies that used the most tasks (Aczel et al., 2015; Ceschi et al., 2019) reported the highest numbers of factors (four and three, respectively). While the selection of the decision-making tasks included in the A-DMC was theory-driven, this selection resulted in a limited number of tasks (especially with regard to heuristics and biases), which might favor the emergence of a one-factor solution.

On the other hand, the discrepancy between the findings might also be related to score reliability. While the A-DMC measures show adequate reliability, several heuristics and biases studies used tasks including a single or a few equivalent items (e.g., Aczel et al., 2015; Toplak et al., 2011; Weaver & Stewart, 2012; Toplak, 2011; Ceschi et al., 2019), leading to scores with poor or moderate reliability, and thus questioning the robustness of the results of factor analyses. In particular, while the valuable study of Ceschi et al. (2019) included a significant number of tasks, the authors reported no information regarding the reliability of scores to each task. While previous studies have reported evidence that individual differences in decision-making and heuristics and biases tasks can be reliably measured (Bruine de Bruin et al., 2007; Teovanović et al., 2015), the issue of measurement is still relevant as some measures show poor or moderate reliability such as the A-DMC measures of framing and sunk cost fallacy (Berthet, 2021; Blacksmith et al., 2019; Scopelliti et al., 2015), and reliable measures of several common heuristics and biases such as availability and confirmation bias are still lacking (see Berthet, 2022, for a review).

1.3. Aims

Assuming that heuristics and biases are part of DMC, the goal of the present study was to test whether a general DMC factor still emerges when adding other heuristics and biases tasks to the six original A-DMC tasks, while ensuring satisfactory levels of reliability. The A-DMC includes three heuristics and biases tasks: framing, overconfidence, and sunk costs fallacy. With regard to the taxonomy of Stanovich et al. (2008), framing (failure in probabilistic thinking) and sunk costs fallacy (failure in rational decision-making) fall into the category of Mindware gaps (lack of relevant knowledge) while overconfidence, viewed as egocentric thinking, is an instance of Contaminated mindware (faulty mindware). We aimed to expand this set of tasks by adding four heuristics and biases: two belonging to the category of Cognitive miserliness (insufficient reflective thinking), availability (overreliance on immediate examples that come to mind) and anchoring (overreliance on a particular reference point), and two belonging to the category of Contaminated mindware, hindsight/outcome bias (faulty evaluation of a decision), and confirmation bias (faulty hypothesis testing).[1]

2. Method

Open Science Practices: All data files and all materials used in the heuristics and biases tasks are available at: https://osf.io/eh52k/.

2.1. Participants

Participants were recruited by the PRISME Human Behavior Core Facility of the Institut du Cerveau in France to complete three sessions (each of approximately 1 h) one week apart. A total of 239 participants completed the first session; 228 completed sessions 1 and 2; and 223 completed the three sessions. The mean age was 32.5 ($SD = 14.04$). 168 participants were females (74%) and 60 were males (26%). Our sample was quite educated as 67.5% of the participants reported having at least a bachelor's degree. This research was reviewed and approved by Institutional Review Board–Paris School of Economics (approval

[1] Note that these four heuristics and biases also happen to be recurrent in the literature on professional decision-making in four occupational areas (management, finance, law, and medicine; Berthet, in press).

number: 2020–022). Participants gave their informed consent before taking part in the study.

2.2. Measures

After the description of each task and the scoring procedure, we provide the main descriptive statistics of the observed scores in Session 1 or Session 2 (see Table 1 for the results on Session 3).

2.2.1. A-DMC measures

We used the same measures as defined in Bruine de Bruin et al. (2007) (items were simply translated in French). Regarding the measurement of framing and overconfidence, we used the improved version introduced by Berthet (2021).

2.2.1.1. Framing. Framing is the tendency of people to be affected by how information is structured (Kahneman & Tversky, 1984). We used the improved version of the risky choice framing task of the A-DMC as described by Berthet (2021). Participants were presented with decision problems and were required to choose between a sure-thing option (A) and a risky-choice option (B). They responded on a 6-point scale ranging from 1 ("I would definitely choose option A") to 6 ("I would definitely choose option B"). Each decision scenario was framed in two versions, a gain version and a loss version (the objective information being the same in both versions), and the wording of the scenario was slightly changed between the two versions (e.g., "Imagine that an autonomous car out of control is rushing into a city crowd. If nothing is done, the accident will cause 120 deaths. Public authorities must choose between two interventions" vs. "Imagine that a train out of control is about to derail near a village. If nothing is done, the accident will cause 120 deaths. Public authorities must choose between two interventions"). The gain and loss items were blocked and separated in the test in order to minimize the likelihood that participants remember their original responses. The framing effect score was calculated as the difference between the mean ratings of the gain items and the mean ratings of the loss items, the potential range of scores being −5 to 5 ($M = 1.00$, $SD = 1.10$). The same eight decision problems (8 pairs of frames) as in Berthet (2021) were used.

2.2.1.2. Overconfidence. The A-DMC measurement of overconfidence follows a classic procedure in which participants perform a cognitive task while indicating the confidence in their responses (e.g., Lichtenstein & Fischhoff, 1977). In this measurement, overconfidence is a measure of calibration (i.e., the difference between the mean confidence ratings and the mean accuracy). Bruine de Bruin et al. (2007) used dichotomous general knowledge items for the performance task. However, Berthet (2021) obtained unreliable overconfidence scores using a general knowledge task as scores in this task were themselves unreliable. Reliable overconfidence scores were obtained when using a cognitive task (scores in such tasks being typically reliable). As in Berthet (2021), we used the matrix reasoning task of the International Cognitive Ability Resource (Condon & Revelle, 2014), which included 11 items. Participants indicated their confidence on a scale ranging from 0 to 100%, so that the potential range of overconfidence scores was −100 to 100 ($M = 8.54$, $SD = 23.00$).

2.2.1.3. Recognizing social norms. This measure evaluates the ability of participants to estimate the frequency of undesirable behaviors among their peers. Participants were first asked to judge whether "it is sometimes OK" or not to engage in 16 undesirable behaviors (e.g., to keep things you find in the street). This first set of responses allowed us to calculate the actual percentage of participants who endorsed each behavior. In a second phase, participants estimated the percentage of people of their age that would endorse each behavior. Bruine de Bruin et al. (2007) calculated the score as the rank-order correlation between the actual percentage and the estimated percentage of peers' behavior across the 16 items. We used an alternative scoring method by which the score for each item is calculated as the absolute difference between the actual percentage and the estimated percentage of peers' behavior (this scoring method is more appropriate to compute Cronbach's alpha). The potential range of scores was 0 to 100 ($M = 21.84$, $SD = 6.02$).

2.2.1.4. Applying decision rules. This measure evaluates the ability of participants to apply decision rules in the context of a practical choice (consumers choosing between five equally priced DVD players following four criteria). In each item, participants were presented with a particular decision rule (e.g., Brian selects the DVD player with the highest number of ratings greater than "Medium") and were asked to select the correct DVD player(s) satisfying that rule. The task included 10 items and the performance was measured by the number of items to which an incorrect answer was provided (higher scores indicating a more pronounced bias) ($M = 4.85$, $SD = 2.44$).

2.2.1.5. Consistency in risk perception. This measure evaluates the ability of participants to estimate risk following probability rules. Participants were first asked to estimate the probability that particular events

Table 1
Descriptive statistics, discriminative properties, and reliability of scores to the 10 tasks ($N = 223$).

Component		Descriptive statistics								Reliability		
		Sessions 1 and 2			Session 3			Range		Sessions 1 and 2	Session 3	
	Number of items	Mean	SD	SKW	Mean	SD	SKW	Potential	Observed	Internal consistency		Test–retest
Framing	8	1.00	1.10	0.83	0.62	1.08	1.48	−5 to 5	−2.12 to 5	0.76	0.85	0.45***
Overconfidence	11	8.54	23.00	0.15	8.67	19.82	0.65	−100 to 100	−56.36 to 86.36	0.73	0.59	0.54***
Availability	4	3.33	3.79	0.87	2.41	3.44	1.86	−20 to 20	−6.5 to 20	0.67	0.81	0.48***
Anchoring	8	0.47	0.22	0.22	0.43	0.25	0.50	undefined	−0.12 to 1	0.67	0.75	0.63***
Hindsight bias	16	1.82	1.50	0.32	1.08	1.49	0.29	−10 to 10	−2.94 to 6.25	0.82	0.84	0.63***
Outcome bias	16	2.99	2.88	0.25	3.36	3.18	0.19	−10 to 10	−3.12 to 10	0.89	0.91	0.78***
Confirmation bias	4	4.40	1.37	−0.26	4.45	1.59	−0.31	0 to 8	0 to 7.5	0.83	0.88	0.75***
Recognizing social norms	16	21.84	6.02	1.04				0 to 100	9.23 to 44.22	0.62		
Applying decision rules	10	4.85	2.44	0.19				0 to 10	0 to 10	0.69		
Consistency in risk perception	20	5.70	3.00	0.86				0 to 20	0 to 16	0.73		
Sunk costs	10	2.19	0.64	−0.08				0 to 5	0.5 to 4	0.38		

Note. Higher scores indicate a more pronounced bias. Internal consistency was measured by Cronbach's alpha except for overconfidence and outcome bias (for these two measures, the split-half method was used). SKW: Skewness.

(e.g., having a car accident) will occur for the next year on a continuous scale ranging from 0 to 100%. In a second phase, they estimated the probability that the same events will occur for the next 5 years. Consistency in risk perception is measured with regard to three rules: the probability that an event occurs the next year should not be larger than the probability that it occurs in the next 5 years; the probability of a subset event (e.g., the probability of dying in a terrorist attack) should not exceed that of its superset event (e.g., the probability of dying); the probability of complementary events (e.g., getting into a car accident while driving and being accident free) should be equal to 1. A total of 20 pairs of responses were analyzed (10 related to time-consistency, 6 related to inclusion, 4 related to complementarity) and the performance was measured by the number of pairs with inconsistent responses (higher scores indicating a more pronounced bias) (M = 5.70, SD = 3.00).

2.2.1.6. Sunk cost fallacy. Sunk cost fallacy is the tendency of people to follow through on an endeavor if they have already invested resources (effort, time, money) into it (Arkes & Blumer, 1985). From a rational perspective, only future costs and benefits should be considered in the decision to follow through on the decision, regardless of past investment which cannot get back. Participants were presented with 10 hypothetical scenarios and responded on a 6-point scale ranging from 1 (the sunk-cost option) to 6 (the normatively correct option). The sunk cost bias score was calculated as 6 minus the mean rating across the 10 items (higher scores indicating a more pronounced bias), so that the potential range of scores was 0 to 5 (M = 2.19, SD = 0.64).

2.2.2. Additional heuristics and biases tasks

2.2.2.1. Availability bias. The availability heuristic is the tendency of people to judge events' likelihood or frequency based on ease of recall (Tversky & Kahneman, 1974). We adapted the task used by Tversky and Kahneman (1973) in their Study 8 to measure individual differences in availability bias (for retrieval). On each trial, a list of 20 items was displayed on the screen for 20 s. Each list was made of items from two categories (e.g., brands of cars and drinks). Before each list, participants were informed about the forthcoming two categories and told that they will be asked to recall the number of items of one category after the presentation of the list. Pairs of lists were used such that (a) each list included 10 items of each category (the correct answer was always 10); (b) in one list, items of one category were famous (e.g., Ferrari, Jeep, Toyota) while items of the other category were not famous (e.g., La Casera, Tanqueray, Glenlivet), and vice versa for the other list. Four pairs of lists were used; in three of them the items were brands (food/banking brands, restaurant/smartphone brands, car/drink brands) and the fourth pair involved French cities (North/South of France). In each pair, the category of items to estimate was the same for the two lists and the availability bias score was calculated as the difference between the estimate for the famous list and the estimate for the less famous list. The potential range of the mean availability score was −20 to 20 (M = 3.33, SD = 3.79). The four lists in which the items to estimate were famous were presented first, following by the four lists in which the items to estimate were not famous.[2] Some responses to the lists in which the number of items to estimate was famous were superior to maximum value of 20 (4.82% of the observations for the list with food/banking brands and 1.31% of the observations for the list with restaurant/smartphone brands). These responses were replaced by 20.

2.2.2.2. Anchoring bias. The anchoring-and-adjustment heuristic is the tendency of people to adjust their (numerical) judgments towards the first piece of information (Tversky & Kahneman, 1974). We adapted the task used by Jacowitz and Kahneman (1995) to measure individual differences in susceptibility to anchoring. Participants were first asked to estimate whether a quantity (e.g., the number of gold medals won by Japan at the 2012 Summer Olympics) was more or less than a given value (anchor) and then provided their estimate. Eight pairs of items were used, where items in each pair involved similar quantities (e.g., the number of gold medals won by Japan/Australia at the 2012 Summer Olympics [actual number: 35]) but different anchors, with either a low or a high anchor corresponding to 10% or 190% of the actual number. In each pair, the anchoring bias score was calculated as follows: (estimate[high anchor] − estimate[low anchor]) / (high anchor − low anchor). As participants provided free numerical estimations, we expected outlier responses which were treated as follows. Outliers were defined as observations that fall above Q3 + 1.5*IQR (Tukey, 1977) (4.60% of the observations per item on average). In a winsorizing approach, outliers were replaced by the trimmed maximum (Q3 + 1.5*IQR). The anchoring bias score is not bounded theoretically but scores were expected to range from 0 to 1. In Session 1, the observed range was −0.02 to 1.00 (M = 0.47, SD = 0.22) with one participant falling outside the expected range; in Session 3, the observed range was −0.12 to 1.00 (M = 0.43, SD = 0.25) with three participants falling outside the expected range. The 16 items were presented in a random order, which was the same for all participants.

2.2.2.3. Outcome/hindsight bias. Hindsight bias is the tendency to make different judgments (e.g., judging the probability of an outcome) between hindsight and foresight conditions (Hawkins & Hastie, 1990). Outcome bias is the tendency to make different judgments (e.g., judging the quality of a decision) based on the outcome (e.g., bad vs. good) (Baron & Hershey, 1988). We designed a task aimed to measure both biases in the same setting. In a first phase (foresight condition), participants were presented with various scenarios in which an individual made a particular decision (e.g., "Celine has a college exam in two days. One of her friends invited her to a party at her house tonight. Celine has decided to go, thinking she'll be studying all day tomorrow") and rated the probability of a particular outcome (e.g., "What are the chances that Celine will pass her exam?") on a 11-point scale ranging from 0 to 100%. In a second phase (hindsight condition), participants were presented with the same scenarios but this time with the corresponding outcome (e.g., "Celine passed her exam") and rated the quality of the decision made (e.g., going to the party at her friend's house) on a 11-point scale ranging from 1 ("It was a very poor decision") to 11 ("It was an excellent decision"). A total of 16 items were used, 8 included a positive outcome and 8 a negative one (see Berthet, 2021). For items with a positive outcome, the hindsight bias score was calculated as the difference between the hindsight judgment and the foresight judgment; that difference was reversed for items with a negative outcome (M = 1.82, SD = 1.50). The outcome bias score was calculated as the difference between the mean ratings of decisions with a positive outcome and the mean ratings of decisions with a negative outcome (M = 2.99, SD = 2.88).

2.2.2.4. Confirmation bias. Confirmation bias is the tendency to search for, to interpret, to favor, and to recall information that confirms personal beliefs (Nickerson, 1998). Berthet (2021) reported evidence that confirmation bias can be reliably measured using an adapted version of the employment interview task of Snyder and Swann (1978). Participants were provided with a hypothesis regarding an interviewee's personality (e.g., the candidate is extroverted) and then selected among a set of 20 questions eight ones to ask to the interviewee to test the hypothesis. The set of questions included eight questions based on the assumption that the candidate has the personality attribute (e.g., What events make you feel popular with people?), eight questions based on the opposite assumption (e.g., What things do you dislike about loud parties?), and four neutral questions (e.g., What are some of your favorite books?). Asking a question that is based on the assumption that

[2] Ideally, the order of presentation of the lists should be randomized.

the candidate has the personality attribute reflects a confirmatory strategy, since the slightest element mentioned by the interviewee could be viewed as confirming the hypothesis. On the contrary, asking a question based on the opposite assumption is a more challenging test of the hypothesis, since the interviewee would need to contradict the interviewer for the hypothesis to be confirmed. We used four items, each one involving a particular personality trait (agreeableness, conscientiousness, emotional stability, extroversion). In each item, the confirmation bias score was the number of confirming questions selected by the participant (M = 4.40, SD = 1.37). The material was drawn from Sackett (1979, 1982).

2.3. Procedure

After providing consent, participants completed three one-hour online sessions separated by one week. In the first session, participants provided sociodemographic information and then completed six heuristics and biases tasks in the following order: (1a) gain version items of the framing task, (2) availability bias, (3) anchoring bias, (4) overconfidence, (5a) foresight condition, (5b) hindsight condition, (6) confirmation bias, (1b) loss version items of the framing task. In the second session, participants completed the A-DMC tasks (outside framing and overconfidence) in the following order: (1a) the first part of recognizing social norms (one's estimates of social norms), (2) applying decision rules, (3) consistency in risk perception, (4) sunk cost fallacy, (1b) the second part of recognizing social norms (estimating other people's social norms).[3] The third session repeated the six tasks of the first session. Participants received 8 euros if they completed one session only, and 30 euros if they completed the three sessions.

3. Results

All analyses were conducted on the sample who completed the three sessions (N = 223). Participants completed four tasks in Session 2, and the six same tasks in Sessions 1 and 3. For these tasks, we considered the average scores between the two sessions in order to obtain the most reliable estimates.[4]

3.1. Descriptive statistics and reliability

Table 1 shows the descriptive statistics, discriminative properties, and reliability of scores to the tasks (for the six tasks completed in Sessions 1 and 3, we report the summary statistics for each session). All six tasks measuring cognitive biases produced the expected mean effects, which were all large with the exception of overconfidence (this was observed in Session 1 and Session 3, we report here the effect sizes calculated on the data from Session 1). The mean framing effect score (M = 1.00) was significantly higher than the normative value 0, $t(222) = 13.60$, $p < .001$ (all t-tests were one-tailed), indicating that participants were more prone to choose the risky-choice option in items with a loss framing (Cohen's d = 1.03).[5] The mean overconfidence score (M = 8.54) was significantly higher than the normative value 0, $t(222) = 5.55$, $p < .001$, indicating that participants were overconfident (Cohen's d = 0.37). In the availability task, the estimated number of items of the target category was larger for the famous than for the less famous lists, and the mean difference (M = 3.33) was significantly higher than the normative value 0, $t(222) = 13.13$, $p < .001$, indicating that participants' recalls of the target items were based on their availability in long-term memory (Cohen's d = 1.12). The mean anchoring score (M = 0.47) was significantly higher than the normative value 0, $t(222) = 31.26$, $p < .001$, indicating that participants' estimates were pulled towards the anchor (Cohen's d = 1.91). The mean hindsight bias score (M = 1.82) was significantly higher than the normative value 0, $t(222) = 18.11$, $p < .001$, indicating that participants' judgments were affected by the knowledge of the outcome (Cohen's d = 1.61). The mean outcome bias score (M = 2.99) was significantly higher than the normative value 0, $t(222) = 15.51$, $p < .001$, indicating that participants' judgments were affected by the valence of the outcome (Cohen's d = 1.74). In the confirmation bias task, the mean number of confirming questions chosen by the participants (M = 4.40) was significantly higher than the baseline value (3.2) if questions were selected at random, $t(222) = 13.06$, $p < .001$, indicating that they were more prone to confirm than disconfirm the hypothesis at hand (Cohen's d = 3.21). Finally, the mean sunk cost score (M = 2.19) was significantly higher than the normative value 0, $t(222) = 51.10$, $p < .001$, indicating that participants were prone to choose the sunk-cost option (Cohen's d = 3.42).

Results of the Shapiro-Wilk test showed that all bias scores were not normally distributed (e.g., framing and availability scores were positively skewed). Scores on each task showed a large variability as evidenced by the SD and the fact that the observed range was relatively closed to the theoretical one.

Scores to the six tasks used in sessions 1 and 3 reached an adequate level of reliability. Regarding internal consistency, all Cronbach's alpha coefficients were equal or above 0.67 in session 1 (test). The values obtained for the framing effect (α = 0.76), outcome bias (α = 0.89) and confirmation bias (α = 0.83) replicated the results of Berthet (2021), supporting the improvements that we provided for the measurement of these biases. In addition, the values obtained for the availability bias (α = 0.67) and the anchoring effect (α = 0.67) confirmed that the new measurement introduced for each of these biases leads to reliable scores (for both of them, Cronbach's alpha rises to 0.70 if just one item is excluded from analysis). Session 3 (retest) produced similar (even higher) Cronbach's alphas, with the exception of overconfidence (mean internal consistency = 0.80). Test-retest correlations (with a two-week interval) ranged from 0.45 to 0.78 (mean r = 0.74), all were significant at the $p < .001$ level. With regard to the standards of Cicchetti (1994), scores to outcome bias and confirmation bias tasks showed excellent test-retest reliability (above 0.75), scores to anchoring and hindsight bias tasks reached a good level (between 0.60 and 0.74), and scores to framing, overconfidence, and availability tasks showed fair values (between 0.4 and 0.59). In the case of framing and availability, a reason might be that some participants may have identified the experimental manipulation in these tasks in Session 3, thereby reducing their validity (but not their internal consistency). This was supported by the observation that the mean framing effect ($t(222) = 4.99$, $p < .001$) and the mean availability effect ($t(222) = 3.70$, $p < .001$) were significantly reduced in Session 3 compared to Session 1. In the case of overconfidence, the lower test-retest reliability might be explained by the lower internal consistency of the measure in Session 3 (0.59). Scores to the A-DMC tasks (other than framing and overconfidence) showed adequate internal consistency (α = 0.62 for Recognizing social norms; α = 0.69 for Applying decision rules; α = 0.73 for Consistency in risk perception) except for Sunk cost fallacy (α = 0.38), a result reported in previous studies (Berthet, 2021; Blacksmith et al., 2019; Scopelliti et al., 2015).

[3] In the second session, participants also completed other measures such as cognitive ability, decision style, and personality. For the sake of clarity and simplicity, we decided to focus our analysis on individual differences in performance on heuristics and biases tasks.

[4] We thank a reviewer for this suggestion.

[5] For CB tasks including two experimental conditions which differed with regard to a normatively irrelevant factor (framing effect, availability bias, anchoring effect, hindsight bias, outcome bias), Cohen's ds were calculated following the standard formula $d = (M1 - M2) / SD$. For CB tasks that did not involve the manipulation of an irrelevant factor (overconfidence, confirmation bias, sunk cost fallacy), Cohen's ds were calculated by the formula $d = (M - 0) / SD$ (0 being normative value).

3.2. Exploratory factor analysis (EFA)

As a preliminary note, it turned out that scores on the hindsight bias and outcome bias were highly correlated ($r = 0.93$, $p < .001$). In fact, these two biases are inherently related (Teichman, 2014) and the task we used may have inflated this relationship. Given this overlap, we retained the outcome bias in the following analyses as scores on this bias showed higher reliability. Table 2 displays bivariate correlations among the 10 heuristics and biases tasks. The correlations were low overall (mean $r = 0.07$), thus replicating previous findings showing no positive manifold (Berthet, 2021; Teovanović et al., 2015). Among these 10 tasks, 18 correlations reached statistical significance but only 5 of them were of practical importance (above 0.30 in absolute value).

We conducted EFA (with maximum likelihood estimation) to explore the factorial structure of the tasks. Note that our N/p ratio (223/10 = 22) was above the recommended values, which range from 3 to 6 (Cattell, 1978) to 20 (Hair et al., 2006), suggesting that we had enough power to conduct factor analysis on our data. We first performed EFA on the six original A-DMC measures. Table 2 shows that most correlations between these measures were significant and positive (mean $r = 0.18$), indicating a moderate positive manifold. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.64 (above the accepting value of 0.50 recommended by Hair et al., 2006) and Bartlett's test of sphericity was significant ($\chi^2(15) = 172.32$, $p < .001$), indicating that the data were suitable for factor analysis (Bartlett, 1950).. A one-factor model yielded relatively poor fit statistics, $\chi^2(9) = 25.81$, $p < .01$, RMSEA = 0.09 (90% CI = 0.05–0.13), CFI = 0.91, accounting for 25.0% of the total variance. Examination of the factor loadings revealed that Overconfidence, Applying Decision Rules, and Consistency in Risk Perception had loadings above 0.40 (Table 3). Considering that Recognizing Social Norms had a decent loading on the factor (0.35) and that scores to Sunk cost fallacy showed poor reliability, our findings supported to some extent the claim that scores on the A-DMC tasks can be reasonably aggregated into a composite index. With the exception of BIC, a two-factor model yielded the best fit, $\chi^2(4) = 2.72$, $p = .0.61$, RMSEA = 0.00 (90% CI = 0–0.08), CFI = 1.00, accounting for 37.0% of the variance. Table 3 indicates that Applying Decision Rules and Consistency in Risk Perception, together with Overconfidence and Recognizing Social Norms to a lesser extent, loaded highly onto the first factor, while Recognizing Social Norms loaded highly onto the second factor.

We then performed EFA on the six A-DMC tasks and the four additional heuristics and biases tasks combined. Bartlett's test of sphericity was significant, $\chi^2(45) = 279.49$, $p < .001$, and the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.61. With regard to the fit statistics, a two-factor solution has the lowest BIC, and acceptable values of CFI and RMSEA (see Table 4). The three-factor had comparable values of CFI and RMSEA, but a notably worse BIC value. To further arbitrate between these two solutions, we noted that the Empirical Kaiser Criterion method, the Hull method and the Parallel analysis of eigenvalues (as implemented in the *dimtest* function in the *EFA.dimensions* package in R) all suggested an optimal number of two factors.

With regard to the meaning of the factors, loadings in the two-factor solution seemed consistent with two factors of the taxonomy of Stanovitch et al. (2008), namely "Mindware gaps" (factor 1) and "Contaminated mindware" (for factor 2). Indeed, Table 4 shows that Overconfidence, Applying Decision Rules, and Consistency in Risk Perception had loadings above 0.40 onto the first factor, while Outcome bias and Confirmation bias had loadings above 0.40 onto the second factor. Note that this pattern of loadings was also found when conducting separate analyses on data from Session 1 and Session 3 (see Supplementary Material). This two-factor solution might be related to the taxonomy proposed by Stanovich et al. (2008): Factor 1 could be labeled "Mindware gaps" (the ability to apply relevant rules or knowledge) while Factor 2 could be called "Contaminated mindware", that is, the tendency to make decisions in a faulty way (evaluating a particular decision based on its outcome, evaluating a question based on its ability to confirm a particular hypothesis). Note that the moderate loading of Overconfidence onto the Mindware gaps factor (0.45) actually reflected the variance of scores to the matrix reasoning task (reflecting the ability to infer a rule) as these two scores were highly correlated ($r = -0.63$, $p < .001$) (this was due to the fact that Overconfidence was calculated as the difference between the mean confidence ratings and the mean performance in the matrix reasoning task). This interpretation of the factors following the taxonomy of Stanovich et al. (2008) was not entirely satisfactory though, as one might have expected Recognizing Social Norms to load more on the Contaminated mindware factor, and Availability and Anchoring to load onto a third factor reflecting Cognitive miserliness.[6]

4. Discussion

The purpose of the present study was to test whether a general DMC factor accounts for the variance of scores to the six original A-DMC tasks (Bruine de Bruin et al., 2007), and in particular whether such a factor still emerges when adding other heuristics and biases tasks to the A-DMC. In fact, while heuristics and biases can be conceived as a part of DMC, the A-DMC assesses just a few of them. We found positive – though limited – evidence of a general DMC factor underlying the variance among the A-DMC tasks. Indeed, our data revealed a slight positive manifold of correlations between the tasks (mean $r = 0.18$) and most of them had moderate to high loadings on the factor with the exception of Sunk cost fallacy, which happened to show poor reliability. The fact that we used our improved measurement of Framing (Berthet, 2021) rather than the original procedure, and the low reliability of scores to Sunk cost fallacy might explain the partial discrepancy between our results and those of Bruine de Bruin.

However, and most importantly, merging the A-DMC with four additional heuristics and biases tasks (availability, anchoring, outcome bias, confirmation bias) moved apart from a general DMC factor as a two-factor model provided the best fit and coherent factor solution. These two factors were partially meaningful with regard to the taxonomy of Stanovich et al. (2008). On the one hand, Applying Decision Rules and Consistency in Risk Perception loaded highly onto the first factor. High scores in Applying Decision Rules reflect a lack of logical knowledge to apply effectively a given decision rule while high scores in Consistency in Risk Perception reflect a lack of knowledge of probability rules, both leading to sub-optimal reasoning. Therefore, these two components might be viewed as Mindware gaps, a lack of knowledge of normative rules. On the other hand, Outcome bias and Confirmation bias loaded highly onto the second factor. As these two biases might be viewed as a form of focal bias (focusing on the outcome of the decision, or on confirming evidence), they might reflect a contaminated mindware, which leads to cognitive distortions and unwarranted beliefs. Note that the taxonomy of Pohl (2017), which distinguishes heuristics and biases with regard to the cognitive processes involved (Thinking, Judgment, and Memory), might also shed light on our two-factor model as Factor 1 could be viewed as a Thinking factor (the ability to apply a certain rule and to follow probability rules) while Factor 2 could be described as a Judgment factor, that is, the tendency to make judgments in a certain direction. In sum, our results are consistent with previous attempts of empirical classification of heuristics and biases, that yielding multiple-factor solutions which can be interpreted in light of various theoretical taxonomies (Ceschi et al., 2019; Teovanović et al., 2015; Weaver & Stewart, 2012).

Noteworthy, we conducted factor analyses while ensuring that scores on all tasks showed adequate reliability (with the exception of Sunk costs fallacy), in contrast with previous studies (e.g., Aczel et al., 2015;

[6] In order to test the robustness of our results, we conducted EFA on residuals, after controlling scores to each task for age, gender and level of education. This analysis provided a similar two-factor solution.

Table 2
Observed correlations between study variables (N = 223).

	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	–											
2. Education	−0.24***	–										
3. Framing	0.04	−0.05	–									
4. Overconfidence	0.29***	−0.10	0.18**	–								
5. Availability	−0.05	0.07	−0.07	−0.17**	–							
6. Anchoring	0.17*	−0.03	−0.09	0.04	0.08	–						
7. Outcome bias	0.23***	−0.15*	0.15	0.11*	−0.07	0.11	–					
8. Confirmation	bias	0.20**	0.05	0.04	−0.10	−0.09	−0.08	0.28***	–			
9. SN	0.34***	−0.15*	0.19**	0.26***	0.09	−0.03	0.23**	0.07	–			
10. DR	0.48***	−0.34***	0.09	0.34***	0.03	0.14*	0.18**	−0.07	0.31***	–		
11. RP	0.31***	−0.21**	0.10	0.36***	−0.10	0.15*	0.13*	−0.30***	0.14*	0.57***	–	
12. Sunk costs	−0.17**	−0.05	−0.03	0.02	0.07	0.07	−0.16*	−0.12	−0.12	0.05	0.11	–

Note. SN: Recognizing Social Norms, DR: Applying Decision Rules, RP: Consistency in Risk Perception. * $p < .05$, ** $p < .01$, *** $p < .001$, two-tailed.

Table 3
Exploratory factor analysis of the A-DMC measures.

Fit statistics	One-factor model	Two-factor model	
Chi-square	$\chi^2(9) = 25.81$, $p < .01$	$\chi^2(4) = 2.72$, $p = .61$	
RMSEA	0.09 (90% CI = 0.05–0.13)	0.00 (90% CI = 0.00–0.08)	
CFI	0.91	1.00	
BIC	−22.85	−18.91	
Component	Factor 1	Factor 1	Factor 2
Framing	0.17	0.14	0.19
Overconfidence	**0.48**	**0.46**	0.13
Recognizing social norms	0.35	0.28	**0.72**
Applying decision rules	**0.79**	**0.70**	0.06
Consistency in risk perception	**0.70**	**0.84**	−0.23
Sunk costs fallacy	0.07	0.11	−0.21
Eigenvalue	2.08	2.08	1.12
Variance explained	25.0%	26.6%	10.4%

Note. Oblimin rotation was used as the method of factor rotation in the two-factor solution. The correlation between the two factors was 0.13.

Ceschi et al., 2019; Weaver & Stewart, 2012). In particular, we provided evidence that the measurement of six classic heuristics and biases (framing, overconfidence, availability, anchoring, outcome bias, confirmation bias) can reach satisfactory levels of reliability with regard both to internal consistency and temporal stability. These results confirmed that the measures we improved (framing, outcome bias; Berthet, 2021) or developed (availability, anchoring, confirmation bias) produce reliable scores.

It should be noted that DMC might be viewed as a component of a more general decision-making skill as described by Skilled Decision Theory (Cokely et al., 2018). Integrating the Skilled Memory Theory (Ericsson et al., 1980), rational thinking (Baron, 1985, 2008) and adaptive heuristic decision making (Gigerenzer et al., 1999), Skilled Decision Theory depicts general decision-making skill as an effective use of adaptive heuristics (e.g., resampling, reframing, disconfirming), numeracy (statistical numeracy and risk literacy), and metacognition in decision problems. Interestingly, Cokely et al. (2018) reported that an overall general decision-making skill estimate, derived from an assessment of four components (A-DMC, Risky Prospect Evaluation, Reference Class and Class-Inclusion Neglect, and Ecological Risk Literacy), was mainly predicted by statistical numeracy. This suggests that both decision-making skill and statistical numeracy overlap in terms of practical probabilistic reasoning and metacognitive accuracy (Ghazal et al., 2014). More generally, this line of research suggests that a broader assessment of decision-making skills than the A-DMC would be relevant.

4.1. Limitations

The present study has several limitations. Firstly, our sample was not gender-balanced (females being over-represented in comparison to males) and not representative with regard to age either ($M = 32.5$, $SD = 14.04$). As gender and age are known to influence decision-making (e.g., Boyer, 2006; Lauriola & Levin, 2001), further research should be conducted on more representative samples in order to test the robustness of the results. Secondly, we expanded the set of A-DMC tasks to a limited extent, by adding four heuristics and biases tasks. Clearly, a challenge

Table 4
Exploratory factor analysis of the 10 tasks.

Fit statistics	One-factor model	Two-factor model		Three-factor model		
Chi-square	$\chi^2(35) = 115.12$, $p < .001$	$\chi^2(26) = 51.35$, $p < .01$		$\chi^2(18) = 29.53$, $p < .05$		
RMSEA	0.10 (90% CI = 0.08–0.12)	0.07 (90% CI = 0.04–0.09)		0.05 (90% CI = 0.01–0.09)		
CFI	0.69	0.90		0.96		
BIC	−74.13	−89.24		−67.80		
Component	Factor 1	Factor 1	Factor 2	Factor 1	Factor 2	Factor 3
Framing	0.17	0.12	0.22	0.14	−0.03	0.21
Overconfidence	**0.49**	**0.45**	0.12	**0.48**	−0.15	0.08
Availability	−0.06	−0.06	−0.08	−0.12	**1.00**	0.00
Anchoring	0.18	0.19	−0.02	0.19	0.09	−0.03
Outcome bias	0.22	0.12	**0.54**	0.17	0.00	**0.51**
Confirmation bias	−0.19	−0.34	**0.54**	−0.29	−0.03	**0.55**
Recognizing social norms	0.33	0.24	0.38	0.27	0.15	**0.40**
Applying decision rules	**0.75**	**0.68**	0.16	**0.71**	0.06	0.13
Consistency in risk perception	**0.74**	**0.84**	−0.12	**0.83**	−0.11	−0.19
Sunk costs fallacy	0.08	0.13	−0.26	0.11	0.04	−0.27
Eigenvalue	2.23	2.23	1.57	2.23	1.57	1.18
Variance explained	16.2%	16.6%	8.9%	16.5%	10.6%	8.9%

Note. Oblimin rotation was used as the method of factor rotation in the two-factor and three-factor solution. In the two-factor solution, the correlation between the two factors was 0.13. In the three-factor solution, Factor 1 was correlated at 0.09 with Factor 2 and Factor 3, and the correlation between Factor 2 and Factor 3 was −0.12.

for future factor analytic studies will be to include more tasks, while ensuring adequate score reliability for each task. Finally, while we provided evidence that the measurement of various heuristics and biases led to reliable scores, the validity of these measures should also be addressed. In particular, among our set of 10 tasks, the introduced measure of Availability, directly adapted from the seminal work of Tversky and Kahneman (1973), showed virtually no significant correlation with any of the other measures and accordingly no notable factor loadings. This might reflect a high domain-specificity of this bias but also a lack of validity of the measurement introduced. In fact, the non-famous items used were actually very unfamiliar (e.g., La Casera, Tanqueray, Glenlivet for drink brands) so that the participants' difficulty to recall them could simply reflect a lack of knowledge of these items. Accordingly, our availability effect might also reflect a mere effect of knowledge.

The two last points remind that the results of factor analytic studies (number and interpretation of the factors) are highly dependent on the set of tasks analyzed and the measurement of the variables. Therefore, future individual differences research might benefit from the development of a common database of heuristics and biases tasks that provide reliable scores, as a complement to the Decision Making Individual Differences Inventory (Appelt et al., 2011).

4.2. Conclusion

Recent research on individual differences in rational decision-making has led to two competing views. The first one claims that a single DMC factor accounts for the positive manifold among decision-making tasks and predicts general outcomes (Bruine de Bruin et al., 2007; Parker & Fischhoff, 2005; Stanovich & West, 1998, 2000). On the other hand, heuristics and biases studies have evidenced multi-factor models, supporting the idea that each mental shortcut captures domain-specific processes and thus predicts narrow decision-making outcomes (Aczel et al., 2015; Ceschi et al., 2019; Teovanović et al., 2015). In summary, the present study supports the second view, and more precisely that the results of factor analysis depend on the sample of tasks used. A direct implication is that to further investigate the structure of individual differences in decision-making and rationality, factor analyses should be conducted on more exhaustive samples of tasks. This endeavor requires the development of reliable measures of other biases.

CRediT authorship contribution statement

Vincent Berthet: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **David Autissier:** Funding acquisition, Supervision, Project administration. **Vincent de Gardelle:** Methodology, Writing – original draft, Writing – review & editing.

Funding

This research was funded by Chaire Innovation Managériale et Excellence Opérationnelle (IMEO) at ESSEC Business School.

References

Aczel, B., Bago, B., Szollosi, A., Foldes, A., & Lukacs, B. (2015). Measuring individual differences in decision biases: Methodological considerations. *Frontiers in Psychology, 6*, 1770.
Appelt, K. C., Milch, K. F., Handgraaf, M. J. J., & Weber, E. U. (2011). The decision making individual differences inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision making, 6*(3), 252–262.
Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*(1), 124–140.
Baron, J. (1985). *Rationality and intelligence*. New York: Cambridge University Press.
Baron, J. (2008). *Thinking and deciding* (4th ed.). New York: Cambridge University Press.
Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*(4), 569–579.
Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology, 3*(2), 77–85.
Berthet, V. (2021). The measurement of individual differences in cognitive biases: A review and improvement. *Frontiers in Psychology, 12*, Article 630177.
Berthet, V. (2022). The impact of cognitive biases on professionals' decision-making: A review of four occupational areas. *Frontiers in Psychology*. , Article 80243. https://doi.org/10.3389/fpsyg.2021.802439
Blacksmith, N., Behrend, T. S., Dalal, R. S., & Hayes, T. L. (2019). General mental ability and decision-making competence: Theoretically distinct but empirically redundant. *Personality and Individual Differences, 138*, 305–311.
Boyer, T. (2006). The development of risk-taking: A multi-perspective review. *Developmental Review, 26*, 291–345.
Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology, 92*(5), 938–956.
Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2020). Decision-making competence: More than intelligence? *Current Directions in Psychological Science, 29*(2), 186–192.
Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum.
Ceschi, A., Costantini, A., Sartori, R., Weller, J., & Di Fabio, A. (2019). Dimensions of decision-making: An evidence-based classification of heuristics and biases. *Personality and Individual Differences, 146*, 188–200.
Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290.
Cokely, E. T., Feltz, A., Ghazal, S., Allan, J. N., Petrova, D. G., & Garcia-Retamero, R. (2018). Decision making skill: From intelligence to numeracy and expertise. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *Cambridge handbook of expertise and expert performance* (2nd ed., pp. 476–505). New York, NY: Cambridge University Press.
Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence, 43*, 52–64.
Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*(4), 380–417.
Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208*(4448), 1181–1182.
Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances, 3*(10), Article e1701381.
Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision making, 9*(1), 15–34.
Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*(3), 179–203.
Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). New Jersey: Pearson Prentice Hall.
Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*(3), 311–327.
Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*(11), 1161–1166.
Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341–350.
Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
Lauriola, M., & Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences, 31*(2), 215–226.
Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior & Human Performance, 20*(2), 159–183.
Millroth, P., Juslin, P., Winman, A., Nilsson, H., & Lindskog, M. (2020). Preference or ability: Exploring the relations between risk preference, personality, and cognitive abilities. *Journal of Behavioral Decision Making, 33*(4), 477–491.
Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220.
Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making, 18*(1), 1–27.
Pohl, R. F. (2017). Cognitive illusions. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (pp. 3–21). New York, NY: Routledge/Taylor & Francis Group.
Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty*. Reading, MA: Addison-Wesley.
Sackett, P. R. (1979). *The interviewer as hypothesis tester: The effects of impressions of an applicant on subsequent interviewer behavior*. Ohio State University.
Sackett, P. R. (1982). The interviewer as hypothesis tester: The effects of impressions of an applicant on interviewer questioning strategy. *Personnel Psychology, 35*(4), 789–804.
Scopelliti, I., Morewedge, C. K., McCormick, E., Min, H. L., Lebrecht, S., & Kassam, K. S. (2015). Bias blind spot: Structure, measurement, and consequences. *Management Science, 61*(10), 2468–2486.
Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology, 36*(11), 1202–1212.

Stankov, L. (2017). Overemphasized “g”. *Journal of Intelligence, 5*(33), 1–10.
Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*(2), 161–188.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(5), 645–665.
Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: A taxonomy of heuristics and biases. In R. V. Kail (Ed.), *Advances in child development and behavior* (pp. 251–285). San Diego, CA: Elsevier Academic Press.
Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). Individual differences as essential components of heuristics and biases research. In K. Manktelow, D. Over, & S. Elqayam (Eds.), *The science of reason: A festschrift for Jonathan St B. T. Evans* (pp. 355–396). New York, NY: Psychology Press.
Teichman, D. (2014). The hindsight bias and the law in hindsight. In E. Zamir, & D. Teichman (Eds.), *Oxford handbooks. The Oxford handbook of behavioral economics and the law* (pp. 354–373). New York, NY: Oxford University Press.
Teovanović, P., Knežević, G., & Stankov, L. (2015). Individual differences in cognitive biases: Evidence against one-factor theory of rationality. *Intelligence, 50*, 75–86.
Toplak, Maggie, et al. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*. https://doi.org/10.3758/s13421-011-0104-1
Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.
Weaver, E. A., & Stewart, T. R. (2012). Dimensions of judgment: Factor analysis of individual differences. *Journal of Behavioral Decision Making, 25*(4), 402–413.