Unpacking Overconfident Behavior When Betting on Oneself

Mohammed Abdellaoui, Han Bleichrodt, Cédric Gutierrez

Unpacking Overconfident Behavior When Betting on Oneself

Mohammed Abdellaoui,[a] Han Bleichrodt,[b] Cédric Gutierrez[c]

[a] Department of Economics and Decision Sciences, GREGHEC-CNRS, HEC-Paris, 78350 Jouy-en-Josas, France; [b] Fundamentos del Análisis Económico (FAE), University of Alicante, Alicante, Spain; [c] Department of Management and Technology, and ICRIOS, Bocconi University, Milan, Italy

Overconfident behavior, the excessive willingness to bet on one's performance, may be driven by optimistic beliefs and ambiguity attitudes. Separating these factors is key for understanding and correcting overconfident behavior, as they may call for different corrective actions. We present a method to do so, which we implement in two incentivized experiments. The first experiment shows the importance of ambiguity attitudes for overconfident behavior. Optimistic ambiguity attitudes (ambiguity seeking) counterbalanced the effect of pessimistic beliefs, leading to neither over- nor underconfident behavior. The second experiment applies our method in contexts where overconfident behavior is expected to vary: easy vs. hard tasks. Our results showed that task difficulty affected both beliefs and ambiguity attitudes. However, while beliefs were more optimistic for relative performance (rank) and more pessimistic for absolute performance (score) on easy tasks compared to hard tasks, ambiguity attitudes were always more optimistic on easy tasks for both absolute and relative performance. Our findings show the subtle interplay between beliefs and ambiguity attitudes: they can reinforce or offset each other, depending on the context, increasing or lowering overconfident behavior.

1. Introduction

Many decisions involve "betting" on one's skills and performance. Examples are a student choosing an educational path (Schulz and Thöni 2016), an entrepreneur deciding to start a new venture (Hooshangi and Loewenstein 2018), or a CEO considering an acquisition (Malmendier and Tate 2008). Such decisions are prone to one of the main cognitive biases: overconfidence (De Bondt and Thaler 1995, Kahneman 2011). In this paper, we argue that overconfident behavior—the excessive willingness to bet on one's own performance—may be driven by two factors: (1) overly optimistic beliefs and (2) ambiguity attitudes, that is, a preference to bet on events with known rather than unknown probabilities (Ellsberg 1961), which determine the willingness to act on those beliefs. Separating these factors is important as they may call for different corrective actions. Our paper proposes a method to do so and implements it in two complementary experimental studies showing that both beliefs and ambiguity attitudes contribute to overconfident behavior and illustrating their interplay.

We compare two methods that have been used to elicit beliefs. The first method uses *matching probabilities* (MP), in which participants choose between bets on their performance on a test and bets on chance. We show that this method measures the combined effect of belief bias and ambiguity attitudes on overconfident behavior. The second method uses *exchangeable events* (EE), in which participants choose only between bets that involve their performance on a test, for instance, if their performance is above or below a certain threshold. In contrast to matching probabilities, the exchangeable events method measures beliefs free from ambiguity attitudes, hence making it possible to isolate the pure impact of belief bias on overconfident behavior. The difference between MP beliefs and EE beliefs reflects the impact of ambiguity attitudes on overconfident behavior.

We measure EE beliefs and MP beliefs in two experiments in which subjects bet on their absolute and relative performance on an ability test. This allows us to study two types of overconfidence: overestimation—"thinking that you are better than you are"—and overplacement—"the exaggerated belief that you are better than others" (Moore and Schatz 2017, p.1). Our experimental setup assumes a model where both subjective probabilities and ambiguity attitudes are inferred from choice. We use real incentives and, in line with recent recommendations, elicit subjective probability distributions rather than single probabilities (Soll et al. 2021). Ambiguous prospects in our setup are evaluated as in Tversky and Fox's (1995) two-stage model, except that we use choice-based probabilities instead of judged and possibly non-additive subjective probabilities. While the two-stage model is easier to implement and has been widely used (Fox and Rottenstreich 2003, Kilka and Weber 2001, Rottenstreich and Tversky 1997), it is less appropriate to separate the role of beliefs and ambiguity attitudes on overconfident behavior as non-additive (judged) probabilities could also reflect attitudes toward ambiguity (Wakker 2004). Instead, our model uses additive subjective probabilities but requires the assumption of probabilistic sophistication within sources of uncertainty (Chew and Sagi 2006, 2008). That is, "events are distinguished only by their subjective probabilities" (Chew and Sagi 2008, p. 2), a condition supported by recent empirical studies in various decision-making contexts (Abdellaoui et al. 2011, 2021, Bleichrodt et al. forthcoming, Cerroni et al. 2012, Jiao 2020, Sonsino et al. 2022).[1]

The first study shows the importance of ambiguity attitudes for overconfident behavior. Using a test of average difficulty level, we found no difference between expected performance using MP beliefs and actual performance, indicating that overall, subjects exhibited neither overconfident nor underconfident behavior. Instead, performance expectations using EE beliefs were significantly lower than actual performance (and expectations using MP beliefs), indicating that subjects' beliefs were actually underconfident but that this underconfidence was offset by their attitudinal optimism (i.e., ambiguity seeking). The attitudinal optimism in our self-evaluation setting was stronger than what has usually been found for exogenous sources of uncertainty, such as Ellsberg's (1961) urns or stock market performance (Baillon and Bleichrodt 2015).

In the second study, we applied our method in a setting where overconfident behavior was expected to vary: easy versus hard questions. The literature suggests a *hard-easy effect*: underestimation and overplacement for easy tasks and overestimation and underplacement for hard tasks (Moore and Healy 2008). Our findings supported the hard-easy effect using choice-based beliefs. Results further showed that the difficulty of the task affected not only the confidence in beliefs but also ambiguity attitudes: our participants exhibited more attitudinal optimism for easy than hard tasks. Unlike beliefs that were affected differently by the task difficulty depending on whether the evaluation was absolute (overestimation) or relative (overplacement), ambiguity attitudes were always more optimistic on easy tasks independent of whether subjects bet on their absolute or relative performance. This attitudinal optimism generated a subtle interplay between beliefs and ambiguity attitudes in producing overconfident behavior. When betting on relative performance on an easy task, optimistic beliefs and ambiguity seeking worked in tandem to produce overconfident behavior. Instead, when betting on absolute performance on an easy task, pessimistic beliefs and ambiguity seeking partially offset each other. The differential effect of task difficulty on beliefs and ambiguity attitudes illustrates the importance of separating the two factors of overconfident behavior.

Our findings have important implications for the empirical literature on overconfidence. Scholars have increasingly used incentive-compatible measures of beliefs, particularly matching probabilities (Coutts 2019, Mobius et al. 2014, Urbig et al. 2009), challenging some of the findings in the overconfidence literature (Grieco and Hogarth 2009, Murad et al. 2016). Using matching probabilities facilitates the elicitation process by not requiring the measurement of the utility function. However, as we show, matching probabilities can lead to biased estimations of beliefs if attitudes toward ambiguity are not considered (Charness et al. 2021, Murad et al. 2016). Separating the role of beliefs and

[1] We further examine the assumptions of our method in the discussion section.

ambiguity attitudes is also essential for the interpretation of overconfidence and the implementation of corrective actions. For example, overconfidence has been linked to entrepreneurial failure (Hooshangi and Loewenstein 2018), undesirable business takeovers (Malmendier and Tate 2008), and biased product selection (Feiler and Tong 2021). Organizations and governments that wish to reduce the possible harmful effects of overconfidence should aim for the right cause. Providing statistical information (McGraw et al. 2004), advice, and mentorship (Bryan et al. 2017) can address the bias in beliefs but may be less effective on ambiguity attitudes.

The paper is organized as follows. Section 2 introduces the theoretical framework we use. Section 3 describes the elicitation methods for beliefs and ambiguity attitudes. Sections 4 and 5 report our two studies' experimental protocols and results. Section 6 discusses our findings, and Section 7 offers conclusions.

2. Theoretical Framework

2.1. Notation and Definitions

In the experiments, the performance of the subjects on an ability test was ambiguous, that is, uncertain with unknown probabilities. As subjects bet on performance on the test, the nature of uncertainty in our setup was epistemic as opposed to aleatory (Fox and Ülkümen 2011). We used a *state space*, which consisted of subjects' possible scores (from 0% to 100% of correct answers) or ranks (from 0% to 100%), to describe this ambiguity.

*Event*s are subsets of the state space. The notation $[s_\star, s^\star]$ means that a score [rank] was between $s_\star$% and $s^\star$%. We use $x_E y$ to denote the *ambiguous prospect* that pays x if event E occurs and y otherwise. If probabilities are known, the prospect is *risky,* and we write it as $x_p y$ with p the probability of x. Outcomes x and y are monetary gains, and x is always the better outcome. The *certainty equivalent* of a prospect is the sure outcome c equivalent to the prospect. The *matching probability* of an ambiguous prospect $x_E y$ is the probability m_E for which $x_E y$ and the risky prospect $x_{m_E} y$ are equivalent.

2.2. Choice Setup

We assume that two-outcome prospects $x_E y$ and $x_p y$ are evaluated as

$$\pi U(x) + (1 - \pi) U(y), \tag{1}$$

where π is a *decision weight* that measures subjects' willingness to bet on the "winning event" and U is a strictly increasing *utility function*. For two-outcome prospects, equation (1) has the main models of decision under ambiguity and risk as special cases, including the ones that do best empirically (e.g., α-maxmin and prospect theory). For risk, $\pi = w(p)$, where w is a strictly increasing *probability weighting function* that maps probabilities onto the unit interval and satisfies $w(0) = 0$ and $w(1) = 1$. Under ambiguity, $\pi = W(E)$ where W is a *weighting function* that maps events to the unit interval $[0,1]$ and satisfies $W(\emptyset) = 0$, $W(S) = 1$ and *monotonicity*: smaller sets (with respect to set inclusion) receive smaller weights. In a Bayesian choice setup, that is, under subjective expected utility (SEU), the willingness to bet on a given event E, $\pi = W(E)$, reduces to its subjective probability. Thus, in the standard Bayesian framework, rational decision-makers can exhibit neither aversion nor proneness to ambiguity, that is, ambiguity neutrality.

We consider various sources of uncertainty in our two studies, including the risky source used as a baseline source. In Study 1, we characterize a source in terms of performance on a given test (whether a score or rank) while keeping the difficulty of the test constant. In Study 2, we distinguish sources of uncertainty on two dimensions: performance (score or rank) and task difficulty (easy or hard). In both studies, we assume that subjects are probabilistically sophisticated within sources (Chew and Sagi 2006, Machina and Schmeidler 1992), but not between sources, which would amount to ambiguity neutrality.

From a decision theory perspective, probabilistic sophistication involves two assumptions. First, subjects could assign choice-based subjective probabilities to events generated by a specific source through a probability measure $P(.)$. Second, choices among bets (e.g., $x_E y$) are exclusively based on the "risky-like" prospects induced by $P(.)$, that is, $x_{P(E)}y$, without assuming a specific model of choice. From a psychological perspective, the assumption underlying Fox and Tversky's (1998) two-stage model is similar to probabilistic sophistication. The two-stage model stipulates that, when facing a bet, the decision-maker first assigns subjective (judged) probabilities to events then evaluates the resulting "risky-like" prospect. This assumption only differs from probabilistic sophistication in that subjective probabilities are judged and possibly non-additive, and induced "risky-like" prospects are evaluated under prospect theory.

Formally, probabilistic sophistication in our setup means that there exists an (additive and choice-based) subjective probability measure $P(.)$ over the events. When $W(E) \neq P(E)$, we can find a strictly increasing *transformation function* f_s from $[0,1]$ to $[0,1]$ satisfying $f_s(0) = 0$ and $f_s(1) = 1$ such that

$$W(E) = f_s(P(E)). \tag{2}$$

The subscript s shows that the transformation function depends on the source of uncertainty. Note that, in the standard Bayesian choice setup, f_s should be the identity function, that is, for all events E, $W(E) = P(E)$.

Apart from the very nature of subjective probability, equation (2) is similar to the equation used in Fox and Tversky (1998), except that these authors used w, the probability weighting for risk, instead of the source-dependent function f_s. In their discussion, Fox and Tversky (1998) recognize, however, that their model could be extended to account for source preference by introducing a probability transformation similar to f_s in our setup.

2.3. Willingness to Bet and Ambiguity Attitude

Abdellaoui et al. (2011) and Dimmock et al. (2016) define ambiguity aversion for a source s as the difference between f_s (the willingness to bet on source s) and w (the willingness to bet on chance). Dimmock et al. (2016) used matching probabilities to measure willingness to bet on events on the (objective) probability scale. In the Bayesian choice setup, which imposes ambiguity neutrality, the matching probability m_E of an event E should be equal to the subjective probability $P(E)$ of the event.

Under the more general model (1) with W as in equation (2), the equivalence between x_E0 and $x_{m_E}0$ gives

$$w(m_E) = f_s(P(E)),$$

which simplifies to

$$m_E = (w^{-1} \circ f_s)(P(E)). \tag{3}$$

This equation shows that m_E could be affected by biases related to overconfidence (through a possibly biased subjective probability $P(E)$) on the one hand and attitude toward ambiguity (through the difference between w and f_s at the likelihood level $P(E)$) on the other hand.[2]

The comparison between a matching probability m_E and its corresponding subjective probability $P(E)$ can be interpreted in terms of local ambiguity aversion, that is, at the likelihood level $P(E)$. For any event E, if m_E is less than $P(E)$, then the subject is ambiguity averse. She is willing to give up $P(E) - m_E$ of her winning probability $P(E)$ to know the probability of winning; therefore, chance is more attractive than ambiguity. When m_E exceeds $P(E)$, the subject is ambiguity seeking. Thus, the function $w^{-1} \circ f_s$ reflects ambiguity attitudes. In the plot of m_E against $P(E)$, the *ambiguity function* $w^{-1} \circ f_s$ lies everywhere below the 45-degree line for ambiguity-averse subjects and above the 45-degree line for ambiguity-seeking subjects.

We further split ambiguity attitudes into attitudinal pessimism (also referred to as ambiguity aversion) and likelihood insensitivity (Abdellaoui et al. 2011, Dimmock et al. 2016). More pessimistic attitudes mean less weight to the best outcome, and more likelihood insensitivity means less sensitivity to changes in likelihood.

To measure attitudinal pessimism and likelihood insensitivity, Dimmock et al. (2016, p. 1367)

[2]Note that, in a non-Bayesian setup where w and f_s are not identity functions and are possibly different, as observed in recent experimental studies assuming prospect theory (Abdellaoui et al. 2011), equation (3) results in $m_E + m_{E^c} \neq 1$. In this case, m_E is rather a willingness to bet coefficient (measured on an objective probability scale) reflecting ambiguity attitude, in addition to pure beliefs regarding E (Wakker 2004).

proposed the following simple model:

$$m_E = c + dP(E), 0 < P(E) < 1, \tag{4}$$

where the matching probability m_E of an event E is linearly related to its subjective probability $P(E)$. Attitudinal pessimism and likelihood sensitivity in this model are measured using the indexes $b = 1 - d - 2c$ and $a = 1 - d$, respectively. Index b is inversely related to the average height of the regression line, hence representing a global index of ambiguity aversion. A smaller b means more optimistic attitudes. Index a reflects likelihood sensitivity with values smaller than 1, indicating a lack of sensitivity to probabilities. Figure B.1 (in Appendix B) illustrates the two indexes.

3. Elicitation Method

The present section describes our elicitation method. In particular, we explain how beliefs, ambiguity attitudes, and overconfidence are measured in our two studies.

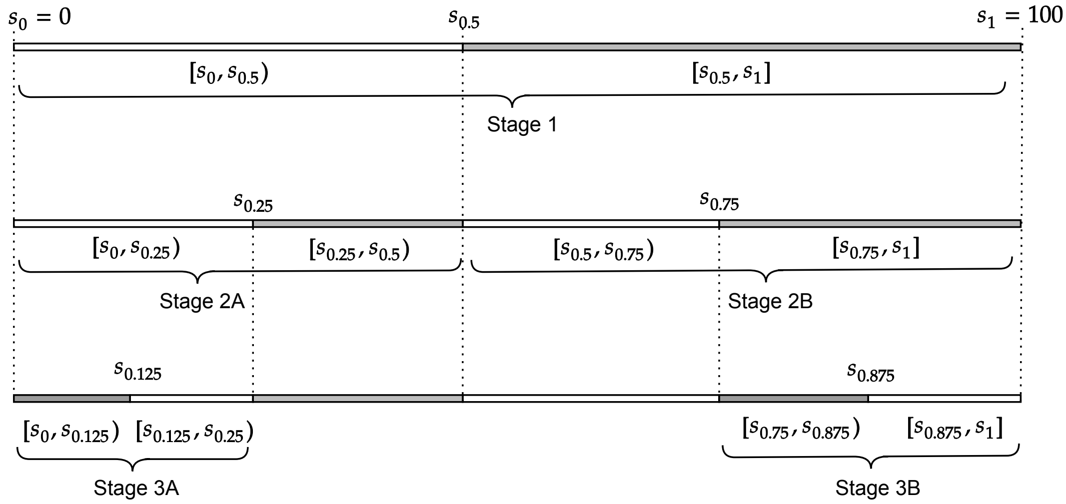
3.1. Beliefs

This paper operationalizes probabilistic sophistication using the exchangeability method (Abdellaoui et al. 2011, Baillon 2008). Formally, the exchangeability method is based on the idea of partitioning a "rich" state space (such as the set $[0,100]$ of possible scores in our experimental setup) into n equally likely (and consequently) exchangeable events for any n (Chew and Sagi 2006).[3] Partitioning the state space into an n-fold partition of exchangeable events allows assigning a probability mass of $1/n$ to each event of the partition. Then, the probability $P(E)$ of an arbitrary event E can be inferred from the overall distribution (Abdellaoui et al. 2021, Soll et al. 2021). This method is different from starting with assigning subjective probabilities to individual events with the risk of ending up with probabilities that do not sum to one. This is the case when probabilities are inferred from judgment (e.g., Fox and Tversky 1998), from the elicitation of certainty equivalents (Murad et al. 2016), or from matching probabilities under SEU (Bruhin et al. 2018, Mobius et al. 2014).

Figure 1 illustrates how the exchangeability method proceeds. For any probability $p \in [0,1]$, let $s_p \in [0,100]$ denote the score such that a subject believed that the probability of obtaining at most score s_p was $p : P(\text{score} < s_p) = p$. By definition, $s_0 = 0$ and $s_1 = 100$. We first elicited the score $s_{0.5}$ that made subjects indifferent between betting on $100_{[s_0,s_{0.5})}0$ and betting on $100_{[s_{0.5},s_1]}0$ (stage 1 in Figure 1). By equation (2), $f_s(P([s_0,s_{0.5}))) = f_s(P([s_{0.5},s_1]))$, i.e., equal willingness to bet on $[s_0,s_{0.5})$ and $[s_{0.5},s_1]$. Because f_s is strictly increasing, we have $P([s_0,s_{0.5})) = P([s_{0.5},s_1]) = 0.5$. Thus, we obtain

[3] The idea of eliciting beliefs by splitting events into equally likely events is similar to eliciting utility functions by eliciting equally spaced outcomes in terms of utility (Wakker 2010).

a clean measurement of subjective probabilities free from ambiguity attitudes. The exchangeability method filters out the source-dependent distortion of subjective probabilities, which reflects attitudes towards ambiguity (cf. equation (2) in Section 2.2). Consequently, the score $s_{0.5}$ splits the whole score domain $[0,100]=[s_0,s_1]$ into two equally likely subevents $[s_{0,},s_{0.5})$ and $[s_{0.5},s_1]$. Figure A.2 in Appendix A shows an example of a question we used to elicit $s_{0.5}$.

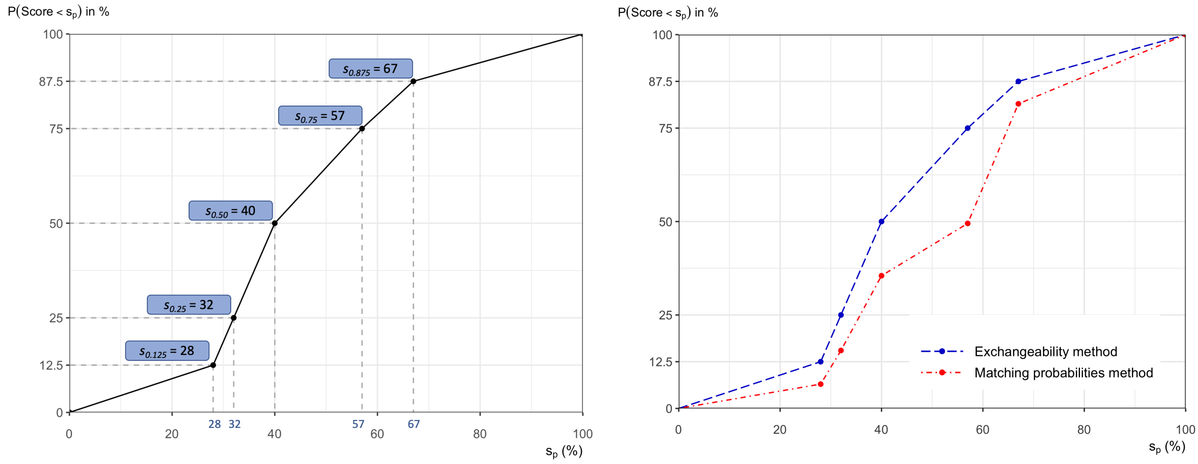


Each line represents a stage of the exchangeable events method to elicit subjective probabilities
(first stage at the top and third stage at the bottom).

Figure 1: Illustration of the Exchangeable Events Method

We then elicited the score $s_{0.25}$ that splits the event $[s_0,s_{0.5})$ into two equally likely subevents by eliciting the indifference between $100_{[s_0,s_{0.25})}0$ and $100_{[s_{0.25},s_{0.5})}0$, and the score $s_{0.75}$ that splits the event $[s_{0.5},s_1]$ into two equally likely subevents (stage 2A/2B in Figure 1). This process led to four events $[s_0,s_{0.25})$, $[s_{0.25},s_{0.5})$, $[s_{0.5},s_{0.75})$ and $[s_{0.75},s_1]$ with the same subjective probability 0.25. We also split the events $[s_0,s_{0.25})$ and $[s_{0.75},s_1]$ into two equally likely subevents (stage 3A/3B in Figure 1), thereby obtaining five fractiles ($s_{0.125}$, $s_{0.25}$, $s_{0.5}$, $s_{0.75}$, and $s_{0.875}$) of the subject's subjective probability distribution. To illustrate, Figure 2 shows the distribution of subject 9 in Study 1. For this subject, the first stage of the exchangeability method yielded a score $s_{0.5}=40$. The second stage, which split events $[0,40)$ and $[40,100]$ into two sets of equally likely subevents, yielded the fractiles $s_{0.25}=32$ and $s_{0.75}=57$. Finally, the third stage of the method yielded fractiles $s_{0.125}=28$ and $s_{0.875}=67$. We used the same method to elicit five fractiles ($r_{0.125}$, $r_{0.25}$, $r_{0.5}$, $r_{0.75}$, and $r_{0.875}$) of subjects' subjective probability distributions about their rank. All fractiles were determined by a bisection process that zoomed in on the desired indifference value and determined it up to a precision of 1% for score and 1 point for rank (see Appendix A3 for an example).

To smooth out response errors, we also estimated the subjective probability distributions over the score and rank of each subject using a beta distribution (for estimation details, see Appendix B1). We chose the beta distribution for its flexibility (Berry 1996).



(a) Distribution of beliefs (i.e., EE-based probabilities) about her score

(b) Distribution of EE-based and MP-based probabilities about her score

Note that the MP-based probabilities in Figure 2b correspond to the complementary events to those used in the elicitation of matching probabilities. Therefore, the position of the red curve under the blue curve suggests more optimistic beliefs with MP-based than EE-based probabilities.

Figure 2: Subject 9's Distribution of Subjective and Matching Probabilities about her Score

3.2. Ambiguity Attitudes

For each elicited fractile s_{p_k}, we determined the matching probability m_k that led to indifference between the prospects $100_{m_k}0$ and $100_{[s_{p_k},s_1]}0$. That is, subjects bet on their score being at least s_{p_k}. For example, for subject 9 in Study 1, we elicited the matching probabilities of $100_{[28,100]}0$, $100_{[32,100]}0$, $100_{[40,100]}0$, $100_{[57,100]}0$, and $100_{[67,100]}0$, where the events had subjective probabilities of 0.875, 0.75, 0.5, 0.25, and 0.125, respectively. Figure A.3 in Appendix A shows the elicitation of the matching probability of $100_{[33,100]}0$. We measured each matching probability using a bisection process with a precision of 1%.

Figure 2b shows the relation between the subjective probabilities and matching probabilities for subject 9. As introduced in Section 2.3, the difference between the matching probability and the subjective probability reflects the ambiguity attitude. For subject 9, the matching probabilities exceeded the subjective probabilities; consequently, the subject was ambiguity seeking for her score.

To demonstrate the impact of ambiguity attitudes, we also estimated the beta distribution of the matching probabilities for each subject. This distribution is only equal to the subjective probability distribution when the subject is ambiguity neutral. Any difference between the distribution based on exchangeable events (*EE-based*) and matching probabilities (*MP-based*) is due to ambiguity attitudes. Much of the behavioral economics literature assumes a Bayesian setup where matching probabilities are assumed to equal subjective probabilities despite the possibility that they could be affected by ambiguity attitudes in real decision situations. Figure 2b shows that this leads to a bias for subject 9. For example, while the estimate of the subject's probability of getting at least 40% of the questions right is 50% using exchangeable events, it is 62% using matching probabilities. Hence, for this subject, matching probabilities suggest more optimistic beliefs than with the exchangeable events method.

3.3. Overconfident Beliefs and Overconfident Behavior

In our setup, overconfident behavior can be measured in two ways. The first method uses the probability scale and focuses on individual events, while the second uses the performance scale and resorts to the overall elicited probability distributions (score/rank).

At the level of a "performance-based" event E, we define overconfident behavior as the difference between the willingness to bet on event E, m_E, and its true probability, $P_t(E)$. This difference represents a deviation from Bayesian behavior that can be decomposed as follows

$$m_E - P_t(E) = \underbrace{[P(E) - P_t(E)]}_{BB} + \underbrace{[m_E - P(E)]}_{AB}$$

where BB stands for the belief bias and AB represents the ambiguity attitude bias. In the absence of these two biases, namely $BB = AB = 0$, the observed matching probability of event E should be equal to its true probability, resulting in the absence of overconfident behavior. Note that under the Bayesian choice setup, $AB = 0$.

Alternatively, overconfident behavior can also be defined in terms of the difference between expected performance (score/rank) as inferred from the elicited matching probability-based (MP-based) probability distribution over performance on the one hand and the observed actual performance on the other hand. Paralleling the above decomposition in terms of probabilities, this measure of overconfident behavior can further be decomposed into beliefs and attitudes components. The belief component is the measure of overconfidence in beliefs, that is, the difference between the expected performance measured using the exchangeability method (deemed to factor out attitude considerations from beliefs) and the actual performance. The attitude component is the difference between MP-based and EE-based expected performance.

4. Study 1

To address the key assumption resulting from our non-Bayesian choice framework that overconfident behavior is impacted by both belief-based overconfidence and ambiguity attitude, Study 1 investigates beliefs and attitudes about performance.

4.1. Experiment

Subjects and incentives. The subjects were 58 students with diverse academic backgrounds.[4] Data were collected through personal interviews that averaged 75 minutes in length. Subjects received a participation fee of €15. Subjects also had a 10% chance of playing out their ability test questions for real and receiving €0.20 per correct answer. Moreover, ten subjects were selected to play out one of their choices for real.[5] We selected the choice that was played out for real randomly. In the end, 19 subjects were paid according to their answers on the ability test and received an extra €5.20 on average. The ten subjects who played out one of their choices for real received an extra €76 on average. No subject was both paid for their score and got to play out one of their choices for real even though this was theoretically possible.

Part 1: Ability test
- 50 questions taken from the Raven's matrices test
- Estimate the score and rank after 25 questions and at the end of the test

Part 2: Bets on performance and on chance
- 10-minute video that explained the type of questions and the real incentive system

The order of the three blocks was randomized
- Block 1: Bets on score on the ability test
  - Measure of the beliefs: elicitation of five events with probabilities 0.125, 0.25, 0.5, 0.75 and 0.875
  - Matching probabilities of five events with probabilities 0.125, 0.25, 0.5, 0.75 and 0.875
- Block 2: Bets on rank on the ability test
  - Measure of the beliefs: elicitation of five events with probabilities 0.125, 0.25, 0.5, 0.75 and 0.875
  - Matching probabilities of five events with probabilities 0.125, 0.25, 0.5, 0.75 and 0.875
- Block 3: Bets on risky lotteries
  - Certainty equivalents of ten risky prospects

Figure 3: Summary of Study 1

First part: Ability test. Figure 3 summarizes the design of Study 1. We first asked 50 questions from Raven's matrices test (Raven et al. 2003), which measures reasoning ability. We chose Raven's matrices as they have frequently been used in studies on overconfidence (e.g., Bruhin et al. 2018, Burks et al. 2013, Herz et al. 2014). In addition, the questions had a large variance in the level of difficulty, which made them well-suited for our experiment. Before starting, subjects were provided with written instructions and a simple example (see Appendix G). Each question had to be answered within 50 seconds.

Subjects assessed their performance after the first 25 questions and at the end of the test.[6] We asked subjects to judge their score (percentage of correct answers) and their rank (among 100 randomly selected subjects).[7] Subjects used a scrollbar to indicate the most likely intervals in which their score

[4]Study 1 was performed with two additional subsamples of subjects devoted to another study on the impact of positive or negative feedback on beliefs and attitudes. Overall, we interviewed 187 subjects.

[5]Subjects were made aware of this opportunity before the experiment started.

[6]Subject self-assessment was done to study the effect of learning, which is not reported here.

[7]In our experimental setup, subjects compare themselves with a non-individuated target (Alicke et al. 1995).

A_k	x^k	Event E_k	y^k	$P(E_k)$
A_1	100	$[s_{0.875}, s_1]$ $[r_0, r_{0.125}]$	0	0.125
A_2	100	$[s_{0.75}, s_1]$ $[r_0, r_{0.25}]$	0	0.25
A_3	100	$[s_{0.5}, s_1]$ $[r_0, r_{0.5}]$	0	0.5
A_4	100	$[s_{0.25}, s_1]$ $[r_0, r_{0.75}]$	0	0.75
A_5	100	$[s_{0.125}, s_1]$ $[r_0, r_{0.875}]$	0	0.875

Table 1: Ambiguous Prospects in Study 1

and rank would lie (see Figure A.1 in Appendix). Each possible interval had a width of 8. For the score, the intervals were 0%–8%, 2%–10%,..., 90%–98%, and 92%–100%. For rank, the intervals were 1–9, 2–10,..., 91–99, and 92–100. We used the midpoint of the elicited intervals as the subject's judged score and rank. We did not use real incentives for these judgments.

Second part: Subjective probabilities and ambiguity attitudes. The second part consisted of three sets of choices. The first set of choices was devoted to eliciting willingness to bet on chance through certainty equivalents. The collected certainty equivalents are not used in the present paper but are treated as fillers in Study 1. The second and third sets of choices focused on eliciting subjective probabilities and matching probabilities for the score and rank. As explained in Section 3, we first elicited five fractiles of the subject's subjective probability distribution about score/rank. Using these fractiles, we then elicited the matching probabilities of five ambiguous events with subjective probabilities of 0.875, 0.75, 0.5, 0.25, and 0.125 (see Table 1). The order of the three sets was random, but the order of the choices within sets was always the same.

Subjects first watched a 10-minute video about the questions they would obtain and the real incentive system (screenshots in Appendix G). They then received two sets of practice questions. The first set elicited (i) the certainty equivalent of the risky prospect $100_{0.4}0$, (ii) the matching probability of the ambiguous prospect $100_{[1,40]}0$, where $[1,40]$ is the event that the subject's rank is between 1 and 40, and (iii) the score s that led to indifference between the ambiguous prospects $100_{[20,s)}0$ and $100_{[s,80]}0$. The second set of practice questions assessed each subject's perceived minimum and maximum score (s_{min} and s_{max}) and rank (r_{min} and r_{max}) on the ability test. These questions were not used in the analyses.

4.2. Results

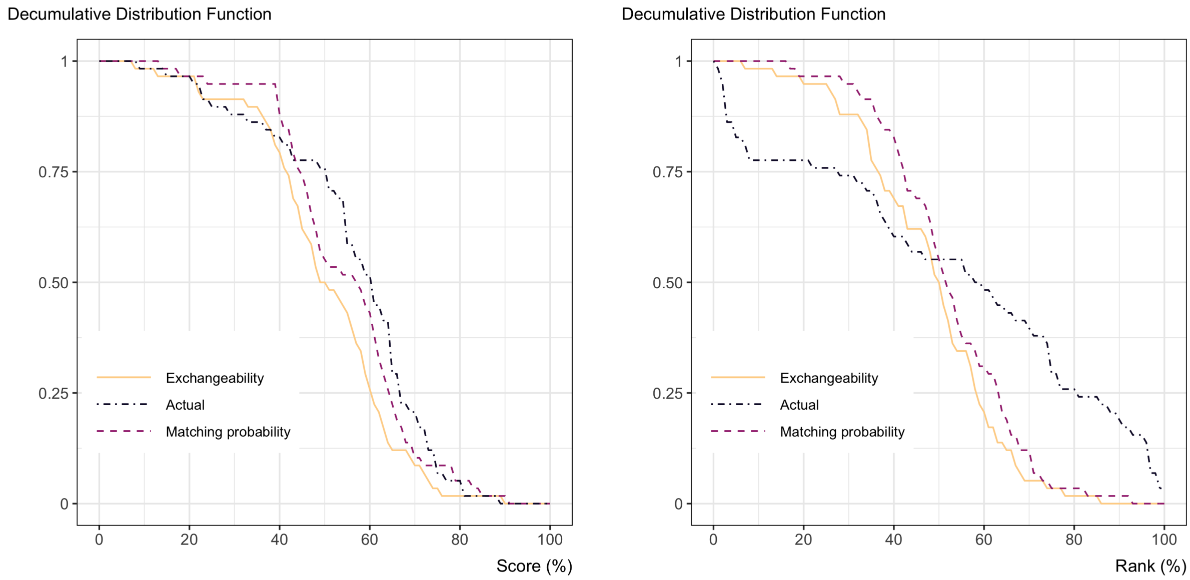
Beliefs about performance. For each subject in Study 1, we elicited EE-based (using exchangeable events) and MP-based (using matching probabilities) subjective probability distributions for the score on one hand and rank on the other hand. This allowed us to infer individual expected performance through the expectations of the corresponding estimated beta distributions. Figure 4 shows the corresponding empirical decumulative distributions defined in terms of the events upon which the subjects bet. For simplicity, and to facilitate comparison with Study 2, we reparametrized the ordinal rank as a percentile rank, that is, ranking number 1 out of 100 corresponds to a percentile rank of 100%, and ranking last at number 100 out of 100 corresponds to a percentile rank of 0%. The left-hand panel of the figure shows for score $x\%$, the probability of getting at least that score. The right-hand panel shows the probability of being ranked higher than the lowest $x\%$.

At the aggregate level, as suggested by Figure 4, we could not reject the null hypothesis that subjects' MP-based expected scores (M = 54.7, SD = 15.2) were equal to their actual scores (M = 55.4, SD = 17.5), $t(57) = -0.352$, p = 0.727. A similar conclusion holds for MP-based expected ranks (M = 52.2, SD = 14.6) as compared to actual ranks (M = 52.9, SD = 34.3), $t(57) = -0.163$, p = 0.871. Binomial tests confirmed these results for score (p = 0.694) and rank (p = 0.358). Using the definition from Section 3.3, the difference between expected performance as inferred from the matching probabilities and the actual performance measures overconfident behavior. Therefore, we did not observe overconfident behavior for score or rank in the sample. This result shows that, in Study 1, an analysis assuming a Bayesian choice framework (i.e., SEU) would have concluded that overall beliefs were well-calibrated, neither over- nor underconfident.

In contrast to MP-based beliefs, the EE-based approach to beliefs (which does not postulate SEU) points to underconfidence. Specifically, our observations reject the null hypothesis that expected scores ($M = 50.1$, $SD = 15.5$) were equal to actual scores ($M = 55.4$, $SD = 17.5$), $t(57) = -2.499$, p = 0.015. Additionally, 69% of subjects were underconfident for score (binomial test, p = 0.005). Regarding rank, although the mean of expected ranks ($M = 48$, $SD = 15.5$) was below that of actual ranks ($M = 52.9$, $SD = 34.3$), our observations cannot reject the null hypothesis that they were equal, $t(57) = -1.213$, p = 0.230. That said, most subjects (66%) had an expected rank below their actual rank, which points to underconfidence (binomial test, p = 0.025).

Figure 4 also suggests that EE-based expectations were lower than MP-based expectations for both the score ($M = 50.1$, $SD = 15.5$ vs. $M = 54.7$, $SD = 15.2$) and rank ($M = 48$, $SD = 15.5$ vs. $M = 52.2$, $SD = 14.6$). As explained in Section 2, the observed gap can be explained in our non-Bayesian choice setup by ambiguity seeking. Paired t-tests reject the null hypothesis of equal expectations (EE-based vs. MP-based) for both score ($t(57) = -5.899$, $p < 0.001$) and rank ($t(57) = -3.668$, $p < 0.001$). Moreover, EE-based distributions were more precise than MP-based distributions in terms of standard deviation for both score ($M = 10.3$, $SD = 4.7$ vs. $M = 13.1$, $SD = 7.4$, $t(57) = -3.464$, p = 0.001) and rank ($M = 12.0$, $SD = 5.8$ vs. $M = 15.9$, $SD = 8.0$, $t(57) = -4.522$, $p < 0.001$).[8]

[8]The mean of a beta distribution $B(\alpha,\beta)$ is computed as $\alpha/(\alpha+\beta)$ and the standard deviation as $\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2+(\alpha+\beta+1)}}$.



(a) Score

(b) Rank

Figure 4: Empirical Distributions of Actual Performance and MP-based and EE-based Expectations Assuming Beta Distributions

Ambiguity attitudes. Given that MP-based probabilities exceeded EE-based probabilities, the dominant pattern was ambiguity seeking (Section 2.3). Table 2 shows that, on average, matching probabilities m_Es exceeded subjective probabilities $P(E)$s except for very likely events. At the individual level, we found ambiguity seeking in terms of score for subjective (i.e., EE-based) probabilities up to 0.75, and in terms of rank, for $P(E) \leq 0.50$. We found more ambiguity seeking than studies using exogenous sources of uncertainty (e.g., Kocher et al. 2018). These studies typically also found ambiguity seeking for unlikely events but ambiguity aversion for moderately likely and likely events. This suggests that, within the so-called source effect, endogeneity may be considered a factor that inflates optimistic attitudes. This issue is investigated in Study 2.

$P(E)$	m_E for score		m_E for rank	
	Mean (SD)	$\%(m_E > P(E))$	Mean (SD)	$\%(m_E > P(E))$
0.125	33.4 (17.6)	84.5%***	34 (19.7)	86.2%***
0.25	50.5 (21.6)	84.5%***	46.7 (19.1)	86.2%***
0.50	68.9 (18.7)	86.2%***	63.2 (16.7)	74.1%***
0.75	78.5 (16.3)	63.8%*	72.9 (18)	48.3%
0.875	81.5 (16.7)	46.6%	81.7 (14.8)	43.1%

Binomial tests *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Matching probabilities are expressed in percentages.

Table 2: Attitude Towards Ambiguity: m_E vs. $P(E)$

We further characterize ambiguity attitude in terms of attitudinal pessimism/optimism and insensitivity using the method described in Section 3, i.e., $m_E = c + dP(E)$, where $b = 1 - d - 2c$ and $a = 1 - d$ define the attitudinal pessimism and insensitivity indexes, respectively. Figure 5 displays the empirical Cumulative Distribution Functions of the attitudinal pessimism index for score and rank. The figure shows that 83% of subjects exhibited optimistic attitudes about their score (i.e., $b < 0$) and 78% about their rank (both binomial tests, $p < 0.001$). Additionally, although our observations point to marginally more optimistic attitudes about score than about rank, we could not reject the null hypotheses of equal attitudinal pessimism between score and rank ($t(57) = -1.978,\ p = 0.053$). The conclusion is similar and clearer for insensitivity ($t(57) = -0.317,\ p = 0.753$).

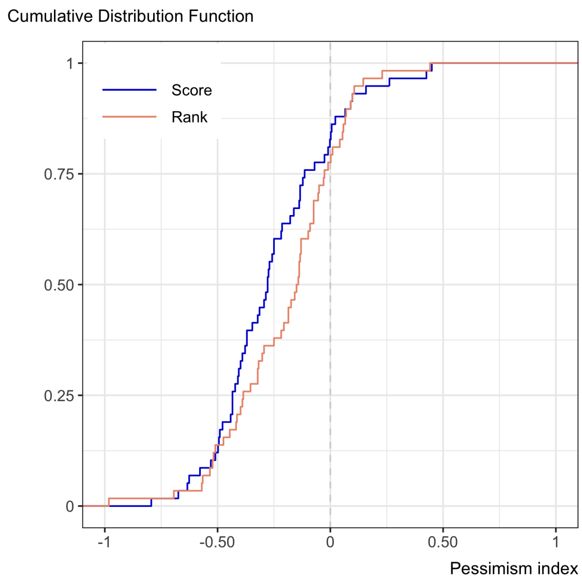


Figure 5: Empirical Distributions of the Attitudinal Pessimism Index

4.3. Conclusion of Study 1

As explained in Section 2.2, the experimental setup in Study 1 follows the literature (Fox and Tversky 1995, 1998) by assigning to events willingness to bet coefficients (decision weights) that account for attitude toward ambiguity in addition to probabilistic beliefs considerations. However, our setup does not rely on probability judgments whose popularity is usually justified by "the common intuition that belief precedes preference" (Fox and Tversky 1998, p. 879) and by violations of the standard model that hinder choice-based elicitation of beliefs. Specifically, the exchangeability method used to elicit beliefs from choice in Study 1 does not commit to SEU and avoids potential distortions that could result from the non-additivity of judged probabilities (Fox and Tversky 1998, Kilka and Weber 2001, de Lara Resende and Wu 2010).

While beliefs elicited with incentivized choice-based methods can perform better than probability judgments in terms of predicting decisions (Atanasov et al. 2017, Trautmann and van de Kuilen 2015b), they require "careful control of additional factors" (Santos-Pinto and de la Rosa 2020, p. 8), such as risk and ambiguity attitudes or the shape of the utility function (Kilka and Weber 2001, Tversky and Fox 1995). Popular choice-based methods for measuring beliefs, such as scoring rules, generally rely on the assumption of risk and ambiguity neutrality, which could bias the elicited beliefs (Armantier and Treich 2013, Offerman et al. 2009). As shown in Section 3, the exchangeability method measures beliefs independent of risk attitude, ambiguity attitude, and utility curvature.

To the best of our knowledge, Study 1 is the first choice-based and incentivized attempt to separate the impact of attitude toward ambiguity and beliefs on overconfident behavior. Overall, assuming SEU, we found neither underconfident nor overconfident behavior. This did not mean that the subjects' beliefs were well-calibrated. In fact, correcting for ambiguity attitudes, we find that subjects were underconfident, particularly when evaluating their score. However, this underconfidence in beliefs was compensated by widespread ambiguity seeking when betting on one's performance, resulting in no under- or overconfident behavior. Study 1 shows that not controlling for non-neutral ambiguity attitudes, such as when using matching probabilities to elicit beliefs (Coutts 2019, Holt and Smith 2009, Urbig et al. 2009), would have led to different conclusions. That said, the results of Study 1 may have been driven by the specific design of the test on performance that included both (very) easy and (very) hard questions, hence resulting in questions that were overall neither hard nor easy. The empirical literature suggests that the degree of confidence may depend on the difficulty of the task (Moore and Healy 2008): on easy tasks, people tend to underestimate their absolute performance but overestimate their relative performance, while on hard tasks, people tend to overestimate their absolute performance but underestimate their relative performance. Therefore, we designed a second study in which we varied, between subjects, the difficulty of the test.

5. Study 2

This study investigates overconfident behavior in a hard-easy context while assuming the same formal choice framework as for Study 1. Among other things, this allows us to check the robustness of the hard-easy effect using an incentivized task. We also included an Ellsberg-like treatment to explore further the difference in ambiguity attitudes between endogenous and exogenous sources of uncertainty.

5.1. Experiment

Subjects and incentives. Due to COVID-19 restrictions, we ran Study 2 online using the Prolific platform. Subjects received a participation fee of \$5.50. In addition, they had the possibility of earning up to \$7 extra based on their performance on the ability test or their choices in the experiment.

Because Study 2 was run online, we used a simplified version of Study 1. Figure 6 summarizes the design of Study 2. Like Study 1, Study 2 had two parts: an ability test using 20 questions and a series of choices using subjects' performance on the ability task as events. Unlike Study 1, however, Study 2 included an Ellsberg task to compare ambiguity attitudes for endogenous and exogenous sources.

After having answered the ability test, subjects were told that one question from the second part of the study would be randomly selected to determine their bonus payment. Before each block of questions, we introduced the type of question and the procedure that would determine the bonus payment (see Appendix H). On average, the study took 26 minutes and subjects earned a $3.30 bonus payment.

Part 1: Ability test
Between subject treatment: Easy vs. Hard test
- 20 questions taken from the Raven's matrices test

Part 2: Bets on performance and on chance
The order of the first two blocks was randomized
- Block 1: Bets on score on the ability test
  - Measure of the beliefs: elicitation of three events with probabilities 0.25, 0.5 and 0.75
  - Matching probabilities of three events with probabilities 0.25, 0.5 and 0.75 (randomized order)
- Block 2: Bets on rank on the ability test
  - Measure of the beliefs: elicitation of three events with probabilities 0.25, 0.5 and 0.75
  - Matching probabilities of three events with probabilities 0.25, 0.5 and 0.75 (randomized order)
- Block 3: Bets on the color of a ball drawn from an urn with unknown composition
  - Matching probabilities of three events with probabilities 0.25, 0.5 and 0.75 (randomized order)

End questions:
- Estimate the percentile score and percentile rank on the test
- Demographics questions

Figure 6: Summary of Study 2

First part: Ability test. The first part consisted of 20 questions from Raven's matrices test (Raven et al. 2003). Subjects had 45 seconds to answer each question. Subjects first answered a filter question, and could only proceed to the actual experiment if they answered this question correctly. Those who did not answer the question correctly received $1 for their participation. Before starting the experiment, subjects rated their perceived level of competence on the ability test on a 10-item Likert scale.

We randomly assigned subjects to one of two conditions: *easy* and *hard*. Subjects in the easy condition got 20 of the easiest questions from the Raven test, and subjects in the hard condition received 20 of the hardest questions. We used information from Study 1 to design the two tests. To check the difficulty of the tests, we ran a pilot study with 45 subjects on Prolific. As expected, the average score in the pilot was higher in the easy condition ($M = 87.5, SD = 15.3$) than in the hard condition ($M = 15.7, SD = 12.5$), $t(40) = 17.2$, $p < 0.001$.[9]

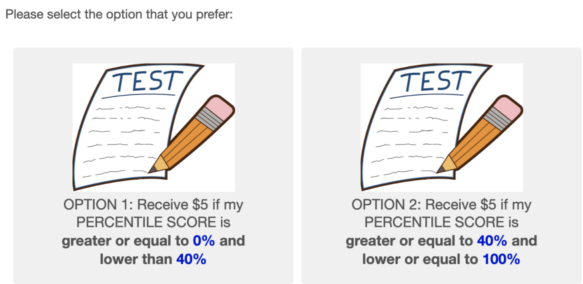
Second part: Subjective probabilities and ambiguity attitudes. The second part consisted of three blocks of choices. The first two blocks elicited subjective probabilities and ambiguity attitudes for score and rank. The order of these blocks was randomized. The third block, which always came last, measured ambiguity attitudes within an Ellsberg task. Within the three blocks, we randomized the order

[9] In case of unequal variance between the two samples, we report the Welch approximation to the degrees of freedom for two-sample t-tests.

of the questions measuring ambiguity attitudes. The questions measuring subjective probabilities were not randomized due to the nature of the exchangeable events method. For all measurements, we used a bisection process with a precision of 5%. Each block started with an explanation of the task, a practice question (see Figure 7), and a series of comprehension questions included to check for data quality (Appendix H).

We used the exchangeable events method to measure three fractiles of the subjective probability distribution for both score ($s_{0.25}$, $s_{0.5}$, $s_{0.75}$) and rank ($r_{0.25}$, $r_{0.5}$, $r_{0.75}$). For each of these fractiles, we measured the matching probabilities of obtaining at least that score or rank: $5_{[j_{0.25}, j_1]}0$, $5_{[j_{0.5}, j_1]}0$ and $5_{[j_{0.75}, j_1]}0$, $j = s, r$.

In the Ellsberg task, we elicited the matching probabilities of three ambiguous prospects using an urn with 100 balls, either blue, red, yellow, or green. The exact composition of the urn was unknown. Before the task, we showed four urns with different compositions to illustrate the concept of an urn with unknown composition. The events depended on the color of a ball randomly drawn from the urn. We elicited the matching probabilities of three events: $5_{[color \subset \{blue\}]}0$, $5_{[color \subset \{blue,\ red\}]}0$, and $5_{[color \subset \{blue,\ red,\ yellow\}]}0$. As in Dimmock et al. (2016), we assumed that subjects would consider each color equally likely a priori (see also Abdellaoui et al. 2011).



(a) Beliefs for Score

(b) Matching Probabilities for Rank

Figure 7: Example of Questions used in Study 2

5.2. Results

Of the 408 subjects who participated in the experiment, 47 did not answer the filter question correctly and could not enter the experiment. Eighty percent of the remaining 361 subjects who entered the experiment answered at least four of the five comprehension questions correctly.[10] We conducted the main analyses with these 289 subjects (155 participants in the easy condition and 134 in the hard condition). The results for the entire sample are similar and are reported in Appendix D.

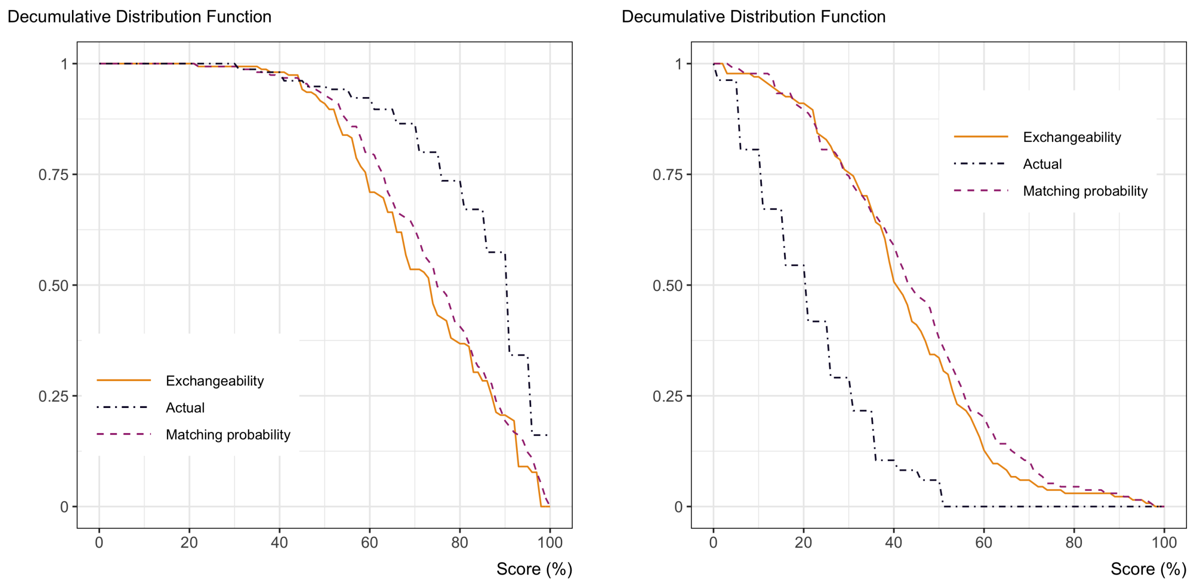
[10]We chose this threshold as four is the median number of correct answers in the sample.

The hard and easy groups did not differ in gender composition, age, education, and prior perception of their competence in the ability test (all t-tests, $p > 0.1$. Details are in Table D.1 in the Appendix). As expected, actual scores were higher in the easy ($M = 83.9,\ SD = 15.9$) than in the hard task ($M = 20.8,\ SD = 13.4$), $t(287) = 36.63,\ p < 0.001$.

We start by presenting the analyses of beliefs about performance and ambiguity attitudes following the structure we used in Study 1. We then move to analyses specific to Study 2 related to the effect of task difficulty on ambiguity attitudes and the difference in ambiguity attitudes between self-evaluation contexts and Ellsberg urns.

Beliefs about performance. Figure 8 shows the empirical distribution functions of actual performance, EE-based expected performance, and MP-based expected performance. As in Study 1, we estimated the beliefs distributions over the score and rank of each subject assuming a beta distribution.

The comparison between actual performance and MP-based expected performance shows that we could replicate the hard-easy effect for (overconfident) behavior (column 4 in Table 3). In the easy condition, subjects had underconfident behavior (i.e., lower MP-based expected performance than actual performance) for their score ($M_{\text{MP}} = 74.4$ vs. $M_{\text{actual}} = 83.9$, $t(154) = -7.32$, $p < 0.001$) and overconfident behavior (i.e., higher MP-based expected performance than actual performance) for their rank ($M_{\text{MP}} = 72.4$ vs. $M_{\text{actual}} = 60.4$, $t(154) = 5.55$, $p < 0.001$). In the hard condition, we observed the opposite pattern: subjects had overconfident behavior for their score ($M_{\text{MP}} = 44.2$ vs. $M_{\text{actual}} = 20.8$, $t(133) = 11.86$, $p < 0.001$) and underconfident behavior for their rank ($M_{\text{MP}} = 48.3$ vs. $M_{\text{actual}} = 58$, $t(133) = -3.47$, $p = 0.001$).



(a) Easy test (score)

(b) Hard test (score)

Figure 8: Empirical Distributions of Actual Performance, MP-based Expectations, and EE-based Expectations Assuming Beta Distributions

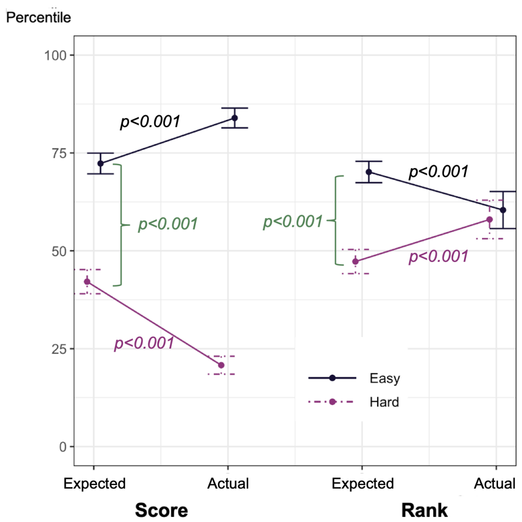
Condition	Expectations/Performance: Mean (SD)			Comparisons: Paired t-tests		
	MP-Beliefs	EE-Beliefs	Actual	MP-Beliefs vs. Actual over/underconfident behavior	EE-Beliefs vs. Actual over/underconfident beliefs	MP vs. EE-Beliefs attitudinal optimism/pessimism
	(1)	(2)	(3)	(4)	(5)	(6)
Score Easy	74.4 (16.5)	72.3 (16.7)	83.9 (15.9)	$t_{154} = -7.32$ p < 0.001	$t_{154} = -8.59$ p < 0.001	$t_{154} = 2.61$ p = 0.010
Rank Easy	72.4 (17.4)	70.1 (17.2)	60.4 (29.8)	$t_{154} = 5.55$ p < 0.001	$t_{154} = 4.52$ p < 0.001	$t_{154} = 2.95$ p = 0.004
Score Hard	44.2 (19.5)	42.1 (18.1)	20.8 (13.4)	$t_{133} = 11.86$ p < 0.001	$t_{133} = 11.27$ p < 0.001	$t_{133} = 2.48$ p = 0.014
Rank Hard	48.3 (18.9)	47.3 (18.0)	58.0 (28.8)	$t_{133} = -3.47$ p = 0.001	$t_{133} = -3.83$ p < 0.001	$t_{133} = 1.16$ p = 0.247

Columns 1-3: means (M) and standard deviations (SD). Column 4: comparisons capturing overconfident/underconfident behavior. Column 5: comparisons capturing overconfident/underconfident beliefs. Column 6: comparisons capturing ambiguity attitudes.

Table 3: Summary of Expectations (EE-based and MP-based Beliefs) and Actual Performance

The comparison between EE-based expected performance and actual performance shows that the hard-easy effect also held for beliefs corrected for ambiguity attitudes (see column 5 in Table 3). In the easy condition, expected scores were lower than actual scores ($M_{\text{EE}} = 72.3$ vs. $M_{\text{actual}} = 83.9$, $t(154) = -8.59$, $p < 0.001$), while expected ranks were higher than actual ranks ($M_{\text{EE}} = 70.1$ vs. $M_{\text{actual}} = 60.4$, $t(154) = 4.52$, $p < 0.001$). In the hard condition, expected scores were higher than actual scores ($M_{\text{EE}} = 42.1$ vs. $M_{\text{actual}} = 20.8$, $t(133) = 11.27$, $p < 0.001$) and expected ranks were lower than actual ranks ($M_{\text{EE}} = 47.3$ vs. $M_{\text{actual}} = 58.0$, $t(133) = -3.83$, $p < 0.001$). That is, we observed underestimation and overplacement for the easy task and overestimation and underplacement for the hard task (see Figure 9).

As in Study 1, the paired t-tests reported in Table 3 (column 6) show that MP-based beliefs differed from EE-based beliefs for both score ($t(154) = 2.61$, $p = 0.010$) and rank ($t(154) = 2.95$, $p = 0.004$) in the easy condition and score ($t(133) = 2.48$, $p = 0.014$) in the hard condition. No significant difference was evident between MP- and EE-based expectations for rank in the hard condition ($t(133) = 1.16$, $p = 0.247$). When expectations differed, MP-based expectations were higher than EE-based expectations, indicating ambiguity seeking. We further explore this pattern in the paragraph "*Ambiguity attitudes*".



Note: The vertical bars show the 95% confidence intervals. The figure illustrates the hard-easy effect on beliefs: underestimation and overplacement for the easy task and overestimation and underplacement for the hard task.

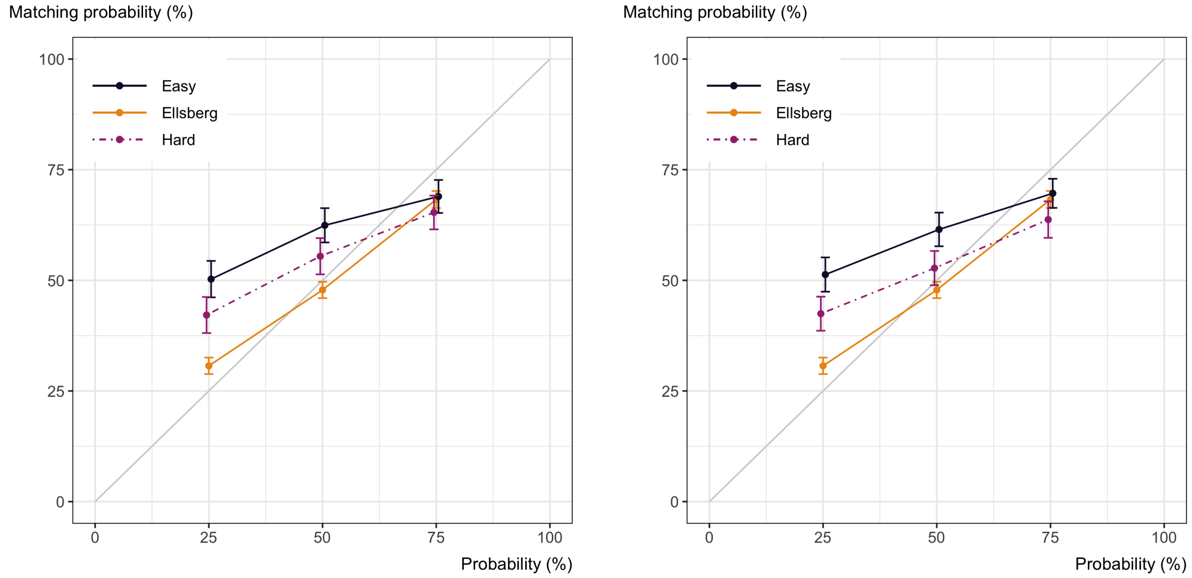
Figure 9: Expected (EE-based) vs. Actual Performance for Score and Rank

These results, illustrated for score in Figure 8 (see Appendix D for rank), reveal a subtle interplay between beliefs and ambiguity attitudes in producing overconfident or underconfident behavior. For the score on an easy task, ambiguity attitudes softened the underconfidence of beliefs. In contrast, ambiguity attitudes reinforced the effect of overconfident beliefs for rank on an easy task and score on a hard task. For rank on a hard task, we observed no effect of ambiguity attitudes at the aggregate level. In general, we observed that ambiguity attitudes softened the effect of beliefs when they were underconfident, leading to less underconfident behavior, but reinforced the effect of beliefs when they were overconfident, leading to more overconfident behavior. The results based on median beliefs are similar (see Appendix D.2).

To determine the relative importance of ambiguity attitudes and beliefs in producing overconfident behavior in our empirical setting, we computed for each subject the ratio between the absolute values of the ambiguity attitude bias (the difference between MP- and EE-based expectations) and the bias in beliefs (the difference between EE-based expectations and actual performance). In the easy condition, the median ratio was 33.5% (IQR=[15.6%; 70.5%]) for score and 28.2% (IQR=[13.1%; 66.3%]) for rank. In the hard condition, the ratio was 22.0% (IQR=[7.8%; 69.3%]) for score and 19.1% (IQR=[6.0%, 44.3%]) for rank. These results suggest that while over- and underconfident behavior was mostly determined by biased beliefs, the effect of ambiguity attitudes on behavior was consequential.

Ambiguity attitudes. Given that MP-based expectations exceeded EE-based expectations, we again observed ambiguity seeking. We further explore this pattern by comparing matching probabilities to their corresponding subjective probabilities. A series of 2 (probability type: matching vs. subjective) $\times 3$ (likelihood level) repeated measures ANOVA revealed significant differences between matching and subjective probabilities for the score $(F(1,154) = 34.36,\ \mathrm{p} < 0.001)$ and rank $(F(1,154) = 42.17,\ \mathrm{p} < 0.001)$ in the easy condition and for the score in the hard condition $(F(1,133) = 6.16, \mathrm{p} = 0.014)$. The difference was not significant at standard levels for rank in the hard condition $(F(1,133) = 2.96,\ \mathrm{p} = 0.088)$.

The interaction effect between probability type and likelihood level was significant for the score $(F(2,308) = 243.59,\ \mathrm{p} < 0.001)$ and rank $(F(2,308) = 253.48,\ \mathrm{p} < 0.001)$ in the easy condition. It was also significant for the score $(F(2,266) = 113.72,\ \mathrm{p} < 0.001)$ and rank $(F(2,266) = 140.06,\ \mathrm{p} < 0.001)$ in the hard condition. These results confirm the patterns observed in Study 1 of ambiguity seeking (except for the rank in the hard condition) and further suggest that the degree of ambiguity seeking is likelihood-dependent.



(a) MPs for Score (b) MPs for Rank

Note: Vertical bars show the 95% confidence intervals. For Ellsberg curves, we grouped the observations of the easy and hard conditions as there was no statistical difference between the two conditions. Subjects are ambiguity seeking when the matching probability is greater than the corresponding subjective probability (i.e., above the diagonal).

Figure 10: Matching Probabilities for the Score and Rank

Figure 10 confirms that ambiguity seeking is likelihood-dependent. Following Study 1, we report the results of a series of binomial tests on the percentage of ambiguity seekers (Table 4). In the easy condition, a majority of subjects were ambiguity seeking for subjective (i.e., EE-based) probabilities up to 0.5 when betting on their score ($p < 0.001$) and rank ($p < 0.010$). In the hard condition, a majority of subjects were ambiguity seeking for their score ($p < 0.001$) and rank ($p < 0.001$), but only at $P(E) = 0.25$. These findings of ambiguity seeking for unlikely and moderately likely events in the easy condition are in line with the findings of the first experiment.

		Easy condition		Hard condition	
	$P(E)$	m_E	$\%(m_E > P(E))$	m_E	$\%(m_E > P(E))$
		(1a)	(1b)	(2a)	(2b)
Score	0.25	50.27 (25.96)	81.9^{***} $p < 0.001$	42.16 (23.84)	76.1^{***} $p < 0.001$
	0.50	62.44 (24.43)	63.9^{***} $p < 0.001$	55.45 (23.97)	58.2^{+} $p = 0.069$
	0.75	68.95 (23.5)	49 $p = 0.872$	65.34 (22.33)	42.5 $p = 0.100$
Rank	0.25	51.31 (24.41)	80.6^{***} $p < 0.001$	42.46 (22.5)	76.1^{***} $p < 0.001$
	0.50	61.5 (23.97)	61.9^{**} $p = 0.004$	52.76 (22.68)	49.3 $p = 0.931$
	0.75	69.66 (20.69)	47.7 $p = 0.630$	63.73 (24.16)	37.3^{**} $p = 0.004$

Columns (a) report the mean and standard deviation (in parentheses) of the matching probabilities, expressed in percentages.
Columns (b) report the results of binomial tests on the percentage of subjects for whom $m_E > P(E)$.
*** $p < 0.001$, ** $p < 0.01$, $^{*}p < 0.05$, $^{+}$ $p < 0.1$.

Table 4: Comparison of Matching Probabilities and Subjective Probabilities, Revealing Ambiguity Seeking/Aversion

Ambiguity attitudes and task difficulty. The different pattern that we observed for the easy and hard conditions suggests that ambiguity attitudes could vary depending on the difficulty of the task. While the previous section analyzed ambiguity attitudes *within* each condition separately, this section explores the differences in ambiguity attitudes *between* the easy and hard conditions. For simplicity, we focus on Dimmock et al.'s (2016) ambiguity attitudes indexes (analyses performed on matching probabilities yield the same conclusions, see Appendix D.3).

A series of two-sample t-tests (see column 3 in Table 5) on Dimmock et al.'s (2016) ambiguity attitudes indexes of the two treatment groups for the different sources (score, rank, and Ellsberg urns) confirm a lower level of attitudinal pessimism in the easy condition than in the hard condition for both score ($M_{\text{easy}} = -0.21$ vs. $M_{\text{hard}} = -0.09$, $t(287) = -2.47$, $\text{p} = 0.014$) and rank ($M_{\text{easy}} = -0.22$ vs. $M_{\text{hard}} = -0.06$, $t(287) = -3.25$, $\text{p} = 0.001$). This difference in the attitudinal pessimism index between the hard and easy conditions is illustrated in Figure 11.

This difference was not due to an intrinsic difference in ambiguity attitudes for exogenous sources between the two groups, as there was no significant difference in attitudinal pessimism for Ellsberg urns between the easy and hard conditions ($M_{\text{easy}} = 0.04$ vs. $M_{\text{hard}} = -0.00$, $t(287) = 1.47$, $\text{p} = 0.142$). There was also no significant difference in the insensitivity index of the easy and the hard conditions for the score ($t(287) = 1.76$, $\text{p} = 0.080$), rank ($t(287) = 1.16$, $\text{p} = 0.248$) or Ellsberg urns ($t(287) = 1.17$, $\text{p} = 0.244$).[11] In sum, task difficulty affected the level of attitudinal pessimism, for both score and rank, but not the sensitivity to changes in likelihood.

Index	Easy Mean (SD) (1)	Hard Mean (SD) (2)	Easy vs. Hard Two-sample t-tests (3)	Paired t-test "Score/Rank" vs. "Ellsberg" Easy (4)	Paired t-test "Score/Rank" vs. "Ellsberg" Hard (5)
Insensitivity Score	0.63 (0.43)	0.54 (0.44)	$t_{287} = 1.76$ $\text{p} = 0.080$	$t_{154} = 7.42$ $\text{p} < 0.001$	$t_{133} = 6.82$ $\text{p} < 0.001$
(Attitudinal) Pessimism Score	-0.21 (0.45)	-0.09 (0.4)	$t_{287} = -2.47$ $\text{p} = 0.014$	$t_{154} = -7.20$ $\text{p} < 0.001$	$t_{133} = -2.61$ $\text{p} = 0.010$
Insensitivity Rank	0.63 (0.39)	0.57 (0.46)	$t_{287} = 1.16$ $\text{p} = 0.248$	$t_{154} = 8.21$ $\text{p} < 0.001$	$t_{133} = 7.27$ $\text{p} < 0.001$
(Attitudinal) Pessimism Rank	-0.22 (0.41)	-0.06 (0.4)	$t_{287} = -3.25$ $\text{p} = 0.001$	$t_{154} = -8.47$ $\text{p} < 0.001$	$t_{133} = -1.74$ $\text{p} = 0.083$
Insensitivity Ellsberg	0.27 (0.38)	0.22 (0.39)	$t_{287} = 1.17$ $\text{p} = 0.244$	— —	— —
(Attitudinal) Pessimism Ellsberg	0.04 (0.27)	-0.00 (0.26)	$t_{287} = 1.47$ $\text{p} = 0.142$	— —	— —

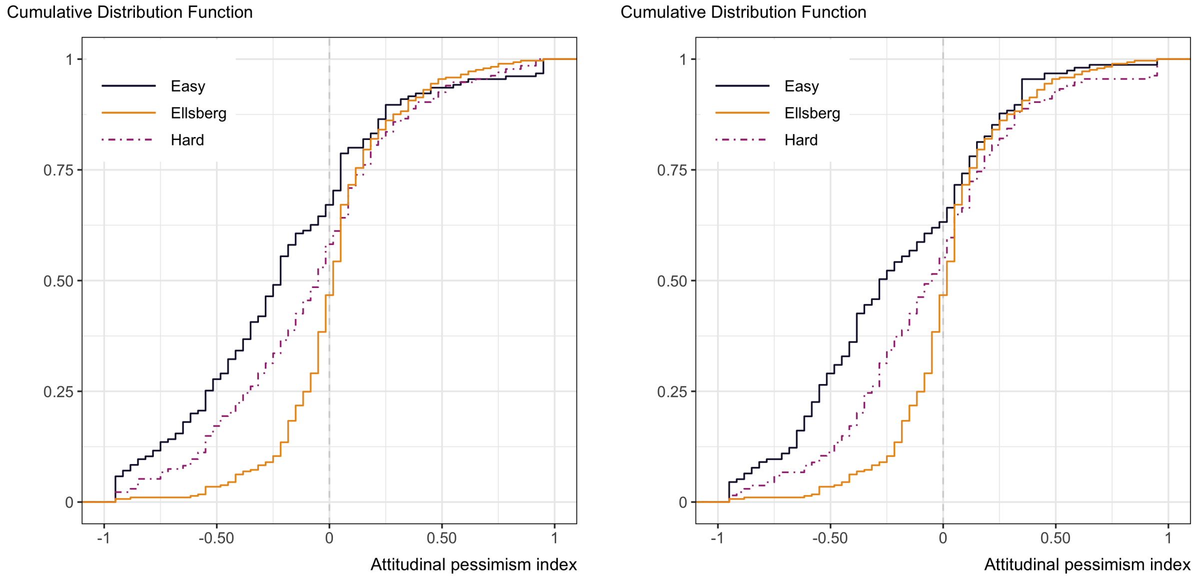
Columns (1) and (2) report the mean and standard deviation of the ambiguity attitudes indexes for the two treatment groups.

Column (3) reports the results of two-sample t-tests on the ambiguity attitudes indexes between the two treatment groups.

Columns (4) and (5) report the results of paired t-tests, within each treatment group, on the ambiguity attitude indexes on the score/rank vs. Ellsberg urns.

Table 5: Ambiguity Attitude Indexes

[11] See Figure D.1. in the online Appendix.



(a) Attitudinal Pessimism – Score (b) Attitudinal Pessimism – Rank

Note: For the Ellsberg curves, we grouped the observations of the easy and hard conditions as there was no statistical difference between the two conditions.

Figure 11: Attitudinal Pessimism Indexes for Score and Rank

Ambiguity attitudes for endogenous and exogenous sources. Study 2 measured participants' ambiguity attitudes for both endogenous (score/rank) and exogenous sources (Ellsberg urns). This allows us to test whether ambiguity attitudes differed within-subject between the score/rank and Ellsberg urns. Our results for Ellsberg urns were similar to prior results in the literature. Using binomial tests, we found that the majority of subjects were ambiguity seeking at probability 0.25 (59.2%, $p = 0.002$) and ambiguity averse at probabilities 0.5 (61.2%, $p < 0.001$) and 0.75 (61.9%, $p < 0.001$).

Columns 4 and 5 of Table 5 report paired t-tests on the differences of ambiguity attitudes within each treatment group, between the score/rank and Ellsberg urns. The results indicate a lower attitudinal pessimism for both the score ($t(154) = -7.20$, $p < 0.001$) and rank ($t(154) = -8.47$, $p < 0.001$) than for Ellsberg urns in the easy condition. In the hard condition, attitudinal pessimism was lower for the score than for Ellsberg urns ($t(133) = -2.61$, $p = 0.010$). Instead, this was not the case for the rank in the hard condition ($t(133) = -1.74$, $p = 0.083$).

There were also significant differences between the insensitivity index for Ellsberg urns and score/rank in both easy and hard conditions (all $p < 0.001$, columns 4 and 5 in Table 5). On average, we observed more insensitivity to probabilities for both score and rank than for Ellsberg urns in both easy and hard conditions. The higher insensitivity for our two endogenous sources than for Ellsberg urns implies a stronger willingness to bet on low-likelihood events when betting on oneself than on an exogenous artificial source of uncertainty.

These results suggest that the endogeneity of the source affected not only the attitudinal pessimism but also likelihood insensitivity. Comparing exogenous natural sources with Ellsberg urns, Abdellaoui et al. (2011) concluded that they differed only in terms of the attitudinal pessimism index. This seems untrue when comparing exogenous and endogenous sources.

Competence effect. Prior literature suggests that perceived competence can also affect the willingness to bet on an uncertain event. People tend to exhibit more optimistic attitudes when they feel competent about a source of uncertainty (Heath and Tversky 1991, Keppe and Weber 1995, Kilka and Weber 2000, 2001, Tversky and Fox 1995). To explore this mechanism, we ran additional analyses separating the subjects based on their *perceived competence* in the task. For simplicity, we report the statistical tests in Appendix D.5. While our analyses confirm that perceived competence is related to more attitudinal optimism in our setting, we find that the results we report above are robust to controlling for perceived competence.

5.3. Conclusion of Study 2

Study 2 confirmed the patterns observed in Study 1 of higher ambiguity seeking in self-evaluation decisions than what has been observed in the literature using exogenous sources of uncertainty (see Trautmann and van de Kuilen 2015a). The study also allowed for a direct comparison of ambiguity attitudes between self-evaluation (score/rank) and Ellsberg urns. Overall, participants in the study exhibited less attitudinal pessimism (except for the rank on a hard task) and more insensitivity to probabilities for betting on their performance than for Ellsberg urns.

Study 2 also showed the usefulness of our method in understanding the effect of task difficulty on both beliefs and ambiguity attitudes (see Table 6 for principal findings). Specifically, we replicated the hard-easy effect for beliefs with incentivized choice-based subjective probabilities without assuming subjective expected utility: task difficulty affected beliefs differently, whether the evaluation was on absolute (score) or relative (rank) performance. Importantly, Study 2 uncovers a new finding: task difficulty also affected ambiguity attitudes. Unlike beliefs, however, ambiguity attitudes for both score and rank were always more optimistic for easy than hard tasks. This finding leads to a subtle interplay between beliefs and ambiguity attitudes in producing overconfident behavior. On easy tasks, attitudinal optimism reinforces the optimistic beliefs in relative evaluation (i.e., rank) but partially offsets the effect of pessimistic beliefs in absolute evaluation (i.e., score). This differential effect of task difficulty on beliefs and ambiguity attitudes illustrates the importance of separating the two factors of overconfident behavior.

Condition	Source	Beliefs	Ambiguity Attitudes
Hard	Score	Overconfidence	Neutrality^{+}
	Rank	Underconfidence	Neutrality
Easy	Score	Underconfidence	Attitudinal optimism
	Rank	Overconfidence	Attitudinal optimism

Table based on binomial tests on the percentage of subjects having overconfident beliefs/ exhibiting attitudinal optimism. Most subjects exhibited ambiguity seeking (i.e., a negative attitudinal pessimism index) in the easy condition for both score (67.1%, $p < 0.001$) and rank (63.2%, $p = 0.001$).

$^{+}$While most subjects (58.2%) had optimistic attitudes toward their score on the hard task, the binomial test was not significant at standard levels ($p = 0.069$). For the rank in the hard task, 55.2% of subjects had optimistic attitudes ($p = 0.261$).

Table 6: Summary of the Study 2 Results: Beliefs and Ambiguity Attitudes

6. Discussion

We present a method to split overconfident behavior when betting on one's own performance into a part that can be attributed to a bias in beliefs and a part that can be attributed to ambiguity attitudes. The essence of our method is to distinguish between measuring beliefs using exchangeable events (which account for the pure effect of belief bias) and matching probabilities (which account for the joint effect of belief bias and ambiguity attitudes). We believe this separation can determine the most adequate (policy) reaction to overconfidence.

In two experiments, we show that both beliefs and ambiguity attitudes contribute to overconfident behavior and that their interplay is subtle. In the first experiment, beliefs were underconfident, but overall, there was neither under- nor overconfident behavior because attitudinal optimism (ambiguity seeking) offset underconfident beliefs. In the second experiment, which studied the well-known hard-easy effect, we found that overconfident beliefs and attitudinal optimism sometimes reinforced each other (relative performance on an easy task). In other cases, underconfident beliefs and attitudinal optimism offset each other (absolute performance on an easy task). This study also sheds new light on the hard-easy effect. Task difficulty affected not only beliefs but also ambiguity attitudes, with more attitudinal optimism observed on easy than hard tasks for both absolute and relative performance.

Research in behavioral economics has suggested that using incentive-compatible measures to assess beliefs can lead to different conclusions about overconfidence than using probability judgments (Blavatskyy 2009, Bruhin et al. 2018, Hoelzl and Rustichini 2005). For instance, Blavatskyy (2009) found that subjects exhibited underconfidence about their absolute performance when using financial incentives. Similarly, Clark and Friesen (2009, p. 229) reported a predominance of underconfidence (in particular for absolute performance) when "participants have incentives to forecast accurately." These studies used a variety of choice-based incentivized tasks to measure beliefs, for instance, scoring rules (Clark and Friesen 2009), eliciting certainty equivalents of bets (Murad et al. 2016), matching probabilities (Bruhin et al. 2018, Mobius et al. 2014), or voting games between lotteries and ambiguous events (Blavatskyy 2009, Grieco and Hogarth 2009, Hoelzl and Rustichini 2005).[12]

[12]Voting games between lotteries and ambiguous events are similar to the elicitation of matching probabilities of ambiguous events.

While it is essential "to assess and control the potential impact of ambiguity attitudes in the context of incentivized belief elicitation" (Murad et al. 2016, p. 39), none of these methods is robust to non-neutral ambiguity attitudes. Our paper shows how this can be done and that the interplay between beliefs and ambiguity attitudes is subtle. This work also shows that previous findings on overconfidence, particularly the hard-easy effect, remain valid when using real incentives. We maintain that it is not the use of real incentives that leads to distortions but how beliefs are measured. Unlike earlier studies (Grieco and Hogarth 2009, Murad et al. 2016), we find evidence of the hard-easy effect using choice-based incentivized measures.

Our finding of ambiguity seeking in self-evaluation decisions adds to the literature showing that ambiguity attitudes are much richer than ambiguity aversion alone. Several studies have found that ambiguity attitudes depend on the source of uncertainty. For example, people prefer to bet on more familiar sources of uncertainty (Abdellaoui et al. 2011) and are more ambiguity seeking regarding sources for which they feel competent (Heath and Tversky 1991, Keppe and Weber 1995, Tversky and Fox 1995). Ambiguity seeking has also been observed in competitive (Gutierrez et al. 2020, Klein et al. 2010) and game-theoretic settings (Li et al. 2020). Overall, we find that people are more ambiguity seeking when betting on their own performance than when betting on an exogenous source of uncertainty (i.e., Ellsberg urns), particularly for easy tasks.

While many of the subjects in our two experiments exhibited ambiguity seeking for the various endogenous sources of uncertainty, we cannot generalize these patterns to other types of uncertainty. In our experiments, the uncertainty surrounding participants' performance was mostly epistemic, that is, the events were potentially knowable, but participants lacked knowledge concerning the true value of the event (Fox and Ülkümen 2011). The literature suggests that people's beliefs (Tannenbaum et al. 2017) and attitudes (Fox et al. 2021, Trautmann et al. 2008) depend on whether uncertainty is epistemic or aleatory (i.e., the events are unknowable). Studying how the two components of overconfident behavior are affected by the nature of uncertainty (epistemic vs. aleatory) is an exciting venue for future research.

In our two studies, we used the exchangeability method to elicit participants' beliefs about their score and rank. This method is choice-based, robust to SEU violations, and has been used in recent experimental elicitations of beliefs (Abdellaoui et al. 2011, 2021, Bleichrodt et al. forthcoming, Jiao 2020, Menapace et al. 2015, Sonsino et al. 2022). The method assumes that subjects are probabilistically sophisticated within sources of uncertainty—a condition supported by recent empirical studies (e.g., Abdellaoui et al. 2021, Bleichrodt et al. forthcoming). As explained in Section 2, this assumption could be seen as the choice-based analog of the main idea behind the two-stage model of Fox and Tversky (1998). The operationalization of probabilistic sophistication through exchangeability allows, however, circumventing the problem that judged probabilities assigned to individual events could be subadditive (as in the two-stage model), hence complicating the separation of the two components of overconfident behavior. By construction, the exchangeability method we used is chained, that is, the answer to a question at a given probability level (e.g., 0.5) is used to split events of lower levels of probability (e.g., 0.25). While this is a limitation of the method, Cerroni et al. (2012) showed that the chained nature of the exchangeability method does not affect its incentive compatibility and that the resulting elicitations are coherent with fundamental probability axioms (Abdellaoui et al. 2011, Baillon 2008). Furthermore, results reported by Baillon (2008) and Abdellaoui et al. (2011) show that the possible error propagation that can result from the chained nature of the exchangeability method does not impact the consistency of the elicited beliefs. To show the robustness of our findings, we performed analyses using only the first stage of the exchangeability method (hence unaffected by the chained nature of the method) that confirmed the results of the two studies (see Appendix D.2).

We measured EE- and MP-based beliefs using different scales. In the first method, we varied the size of the uncertain events, while in the second, we varied the probability of a risky outcome. Different elicitation methods may involve distinct cognitive processes. For example, by varying the objective probability in the matching probability task, subjects may attach additional weight to the probability of success of the risky option, thereby leading to more ambiguity aversion (Chow and Sarin 2001, Fox and Tversky 1995). If so, our results may underestimate the true degree of ambiguity seeking, and ambiguity attitudes may be more important in producing overconfident behavior than we found.[13]

We only used one type of task: logic puzzles measuring reasoning ability. While such tasks have been widely used in the overconfidence literature, future research could explore how both components of overconfident behavior are affected by the nature of the task. Moreover, in the rank task, subjects did not know with whom they were competing. This lack of personal contact with the comparison target could thus have amplified the "better than others effect" in our experiment (Alicke et al. 1995) and affected participants' perception of and attitude toward uncertainty. It would be interesting to repeat our study with a personalized comparison group.

Finally, although we observed that the difficulty of a task affected not only individuals' beliefs but also their attitude toward the source of uncertainty, we did not identify the mechanisms responsible for this empirical pattern. A large literature has explored the mechanisms of the hard-easy effect on beliefs (e.g., Suantak et al. 1996). Exploring the mechanisms that drive the hard-easy effect on ambiguity attitudes offers another exciting avenue for future research.

7. Conclusions

We proposed a method to separate the effects of beliefs and ambiguity attitudes on overconfident behavior and show the key role played by ambiguity attitudes. In our first experiment, beliefs were underconfident, but due to ambiguity seeking, behavior was not. Traditional methods of measuring

[13]In Appendix E, we report the results of a pilot study using certainty equivalents instead of matching probabilities to measure ambiguity attitudes. Overall, we find similar patterns of ambiguity seeking as with matching probabilities.

beliefs based on scoring rules or matching probabilities would not capture such a pattern. Our second experiment confirms that attitude towards ambiguity is affected by whether the source of uncertainty is exogenous (Ellsberg urns) or endogenous (score and rank). Subjects exhibited more ambiguity seeking when betting on their own performance than in decisions involving exogenous uncertainty. We also show that task difficulty affected not only beliefs but also ambiguity attitudes. Unlike beliefs that were affected differently by task difficulty whether subjects evaluated their absolute or relative performance (the hard-easy effect), subjects exhibited more attitudinal optimism for both absolute and relative performance for easy than hard tasks. This generates a subtle interplay between beliefs and ambiguity attitudes in producing overconfident or underconfident behavior.

Identifying the causes of overconfident behavior helps determine appropriate corrective actions.[14] For example, studies on entrepreneurship suggest that training can help reduce biases, particularly overconfidence (Camuffo et al. 2020, Zhang et al. 2021). These methods only work if overconfident behavior is due to overconfident beliefs. If overconfident behavior is due to optimistic attitudes, training to correct beliefs would be less effective. If the aim is to reduce overconfident behavior, training may even be counterproductive if biased beliefs and ambiguity attitudes offset each other, as in our first study. Disentangling beliefs and attitude components of overconfident behavior, as we propose, may help determine when training is effective or unproductive. We hope our method proves useful in future research and policy.

References

Abdellaoui, M., Baillon, A., Placido, L., and Wakker, P. P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *American Economic Review*, 101(2):695–723.

Abdellaoui, M., Bleichrodt, H., Kemel, E., and l'Haridon, O. (2021). Measuring beliefs under ambiguity. *Operations Research*, 69(2):599–612.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., and Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, 68(5):804.

Armantier, O. and Treich, N. (2013). Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging. *European Economic Review*, 62:17–40.

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., Ungar, L., and Mellers, B. (2017). Distilling the wisdom of crowds: Prediction markets vs. prediction polls. *Management science*, 63(3):691–706.

Baillon, A. (2008). Eliciting subjective probabilities through exchangeable events: An advantage and a limitation. *Decision Analysis*, 5(2):76–87.

Baillon, A. and Bleichrodt, H. (2015). Testing ambiguity models through the measurement of probabilities for gains and losses. *American Economic Journal: Microeconomics*, 7(2):77–100.

[14]We note that, in some contexts, overconfidence may have positive effects and need not be reduced (e.g., Bengtsson et al. 2005, Kennedy et al. 2013).

Bengtsson, C., Persson, M., and Willenhag, P. (2005). Gender and overconfidence. *Economics letters*, 86(2):199–203.

Berry, D. A. (1996). *Statistics: a Bayesian perspective*. Number 04; QA279. 5, B4. Duxbury Press Belmont, CA.

Blavatskyy, P. R. (2009). Betting on own knowledge: Experimental test of overconfidence. *Journal of Risk and Uncertainty*, 38(1):39–49.

Bleichrodt, H., Grant, S., and Yang, J. (forthcoming). Testing hurwicz expected utility. *Econometrica*.

Bruhin, A., Santos-Pinto, L., and Staubli, D. (2018). How do beliefs about skill affect risky decisions? *Journal of Economic Behavior & Organization*, 150:350–371.

Bryan, K. A., Tilcsik, A., and Zhu, B. (2017). Which entrepreneurs are coachable and why? *American Economic Review*, 107(5):312–16.

Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2013). Overconfidence and social signalling. *Review of Economic Studies*, 80(3):949–983.

Camuffo, A., Cordova, A., Gambardella, A., and Spina, C. (2020). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Science*, 66(2):564–586.

Cerroni, S., Notaro, S., and Shaw, W. D. (2012). Eliciting and estimating valid subjective probabilities: An experimental investigation of the exchangeability method. *Journal of Economic Behavior & Organization*, 84(1):201–215.

Charness, G., Gneezy, U., and Rasocha, V. (2021). Experimental methods: Eliciting beliefs. *Journal of Economic Behavior & Organization*, 189:234–256.

Chew, S. H. and Sagi, J. S. (2006). Event exchangeability: Probabilistic sophistication without continuity or monotonicity. *Econometrica*, 74(3):771–786.

Chew, S. H. and Sagi, J. S. (2008). Small worlds: Modeling attitudes toward sources of uncertainty. *Journal of Economic Theory*, 139(1):1–24.

Chow, C. C. and Sarin, R. K. (2001). Comparative ignorance and the ellsberg paradox. *Journal of risk and Uncertainty*, 22(2):129–139.

Clark, J. and Friesen, L. (2009). Overconfidence in forecasts of own performance: An experimental study. *Economic Journal*, 119(534):229–251.

Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, 22(2):369–395.

De Bondt, W. F. and Thaler, R. H. (1995). Financial decision-making in markets and firms: A behavioral perspective. *Handbooks in operations research and management science*, 9:385–410.

de Lara Resende, J. G. and Wu, G. (2010). Competence effects for choices involving gains and losses. *Journal of Risk and Uncertainty*, 40(2):109–132.

Dimmock, S. G., Kouwenberg, R., and Wakker, P. P. (2016). Ambiguity attitudes in a large representative sample. *Management Science*, 62(5):1363–1380.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, pages 643–669.

Feiler, D. and Tong, J. (2021). From noise to bias: Overconfidence in new product forecasting. *Management Science*.

Fox, C. R., Goedde-Menke, M., and Tannenbaum, D. (2021). Ambiguity aversion and epistemic uncertainty. *Available at SSRN 3922716*.

Fox, C. R. and Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psychological Science*, 14(3):195–200.

Fox, C. R. and Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics*, 110(3):585–603.

Fox, C. R. and Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science*, 44(7):879–895.

Fox, C. R. and Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. *in Essays in Judgment and Decision Making, Brun, W., Kirkebøen, G. and Montgomery, H., eds. Oslo:*

Universitetsforlaget.

Grieco, D. and Hogarth, R. M. (2009). Overconfidence in absolute and relative performance: The regression hypothesis and bayesian updating. *Journal of Economic Psychology*, 30(5):756–771.

Gutierrez, C., Åstebro, T., and Obloj, T. (2020). The impact of overconfidence and ambiguity attitude on market entry. *Organization Science*, 31(2):308–329.

Heath, C. and Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, 4(1):5–28.

Herz, H., Schunk, D., and Zehnder, C. (2014). How do judgmental overconfidence and overoptimism shape innovative activity? *Games and Economic Behavior*, 83:1–23.

Hoelzl, E. and Rustichini, A. (2005). Overconfident: Do you put your money on it? *Economic Journal*, 115(503):305–318.

Holt, C. A. and Smith, A. M. (2009). An update on bayesian updating. *Journal of Economic Behavior & Organization*, 69(2):125–134.

Hooshangi, S. and Loewenstein, G. (2018). The impact of idea generation and potential appropriation on entrepreneurship: An experimental study. *Management Science*, 64(1):64–82.

Jiao, P. (2020). Payoff-based belief distortion. *The Economic Journal*, 130(629):1416–1444.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kennedy, J. A., Anderson, C., and Moore, D. A. (2013). When overconfidence is revealed to others: Testing the status-enhancement theory of overconfidence. *Organizational Behavior and Human Decision Processes*, 122(2):266–279.

Keppe, H.-J. and Weber, M. (1995). Judged knowledge and ambiguity aversion. *Theory and Decision*, 39(1):51–77.

Kilka, M. and Weber, M. (2000). Home bias in international stock return expectations. *The Journal of Psychology and Financial Markets*, 1(3-4):176–192.

Kilka, M. and Weber, M. (2001). What determines the shape of the probability weighting function under uncertainty? *Management Science*, 47(12):1712–1726.

Klein, W. M., Cerully, J. L., Monin, M. M., and Moore, D. A. (2010). Ability, chance, and ambiguity aversion: Revisiting the competence hypothesis. *Judgment and Decision Making*, 5(3):192.

Kocher, M. G., Lahno, A. M., and Trautmann, S. T. (2018). Ambiguity aversion is not universal. *European Economic Review*, 101:268–283.

Li, C., Turmunkh, U., and Wakker, P. P. (2020). Social and strategic ambiguity versus betrayal aversion. *Games and Economic Behavior*, 123:272–287.

Machina, M. J. and Schmeidler, D. (1992). A more robust definition of subjective probability. *Econometrica*, pages 745–780.

Malmendier, U. and Tate, G. (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics*, 89(1):20–43.

McGraw, A. P., Mellers, B. A., and Ritov, I. (2004). The affective costs of overconfidence. *Journal of Behavioral Decision Making*, 17(4):281–295.

Menapace, L., Colson, G., and Raffaelli, R. (2015). Climate change beliefs and perceptions of agricultural risks: An application of the exchangeability method. *Global Environmental Change*, 35:70–81.

Mobius, M., Niederle, M., Niehaus, P., and Rosenblat, T. (2014). Managing self-confidence: Theory and experimental evidence. Technical report, UC San Diego.

Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2):502–517.

Moore, D. A. and Schatz, D. (2017). The three faces of overconfidence. *Social and Personality Psychology Compass*, 11(8):e12331.

Murad, Z., Sefton, M., and Starmer, C. (2016). How do risk attitudes affect measured confidence? *Journal of Risk and Uncertainty*, 52(1):21–46.

Offerman, T., Sonnemans, J., Van de Kuilen, G., and Wakker, P. P. (2009). A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes. *The Review of Economic Studies*,

76(4):1461–1489.
Raven, J., Raven, J. C., and Court, J. H. (2003). Manual for Raven's progressive matrices and vocabulary scales. Section 4, The advanced progressive matrices. *Harcourt Assessment*.
Rottenstreich, Y. and Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological review*, 104(2):406.
Santos-Pinto, L. and de la Rosa, L. E. (2020). Overconfidence in labor markets. *Handbook of Labor, Human Resources and Population Economics*, pages 1–42.
Schulz, J. F. and Thöni, C. (2016). Overconfidence and career choice. *PloS one*, 11(1):e0145126.
Soll, J. B., Palley, A. B., and Rader, C. A. (2021). The bad thing about good advice: Understanding when and how advice exacerbates overconfidence. *Management Science*.
Sonsino, D., Lahav, Y., and Roth, Y. (2022). Reaching for returns in retail structured investment. *Management Science*, 68(1):466–486.
Suantak, L., Bolger, F., and Ferrell, W. R. (1996). The hard–easy effect in subjective probability calibration. *Organizational Behavior and Human Decision Processes*, 67(2):201–221.
Tannenbaum, D., Fox, C. R., and Ülkümen, G. (2017). Judgment extremity and accuracy under epistemic vs. aleatory uncertainty. *Management Science*, 63(2):497–518.
Trautmann, S. T. and van de Kuilen, G. (2015a). Ambiguity attitudes. In *The Wiley Blackwell handbook of judgment and decision making*, chapter 1, pages 89–116. John Wiley & Sons.
Trautmann, S. T. and van de Kuilen, G. (2015b). Belief elicitation: A horse race among truth serums. *The Economic Journal*, 125(589):2116–2135.
Trautmann, S. T., Vieider, F. M., and Wakker, P. P. (2008). Causes of ambiguity aversion: Known versus unknown preferences. *Journal of Risk and Uncertainty*, 36(3):225–243.
Tversky, A. and Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, 102(2):269.
Urbig, D., Stauf, J., and Weitzel, U. (2009). What is your level of overconfidence? a strictly incentive compatible measurement of absolute and relative overconfidence. *Discussion Paper Series/Tjalling C. Koopmans Research Institute*, 9(20).
Wakker, P. P. (2004). On the composition of risk preference and belief. *Psychological review*, 111(1):236.
Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge university press.
Zhang, S. X., Foo, M.-D., and Vassolo, R. S. (2021). The ramifications of effectuation on biases in entrepreneurship–evidence from a mixed-method approach. *Journal of Business Venturing Insights*, 15:e00238.