CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

Discussion Paper No. 2019-15

Robin Cubitt
Orestis Kopsacheilis
Chris Starmer

An inquiry into the nature and causes of the Description - Experience gap

March 2020

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Suzanne Robey
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 95 14763
suzanne.robey@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

An inquiry into the nature and causes of the Description - Experience gap*

Robin Cubitt[†] Orestis Kopsacheilis[‡] Chris Starmer[§]

Abstract

According to the Description-Experience gap (DE gap), people act as if overweighting rare events when information about those events is derived from descriptions but as if underweighting rare events when they experience them through a sampling process. While the is now clear evidence that the DE gap exists, so far, its exact nature, the causes of it and its implications for economics remain unclear. Due to the variety of experimental designs and measures reported in previous literature, the nature, causes and implications of the phenomenon for economic theory remain unclear. We present a new experiment which examines in a unified design four distinct causal mechanisms that might drive the DE gap, attributing it respectively to information differences (sampling bias), to a feature of preferences (ambiguity sensitivity) or to aspects of cognition (likelihood representation and memory). Our design permits model-free and model-mediated tests for these mechanisms and for the DE gap itself. Using a model-free approach, we elicit a DE gap similar in direction and size to the literature's average and find that, when each factor is considered in isolation, sampling bias stemming from under-represented rare events, is the only significant driver. Yet, model-mediated analysis shows that rare events are overweighted even in experience. Moreover, this level of analysis reveals the potential of a smaller DE gap, existing even without information differences.

Keywords: Decisions from Description, Decisions from Experience, Risk Preferences, Cumulative Prospect Theory, Ambiguity

JEL Codes: D81, C91, D91

*The research was supported by the Economic and Social Research Council [grant numbers ES/K002201/1, ES/P008976/1] under the Network for Integrated Behavioural Science

†School of Economics, University of Nottingham, Nottingham, UK

‡Corresponding author. School of Economics, University of Nottingham, Nottingham, NG7 2RD, UK. email:orestiskopsacheilis@gmail.com

§School of Economics, University of Nottingham, Nottingham, UK

1 Introduction

In this paper, we present an experimental investigation into the nature and causes of the so-called Description-Experience gap (DE gap for short). The DE gap is a widely-documented tendency for people to act as if they have systematically different preferences over risks, depending on whether their information about those risks is derived from explicit descriptions or, alternatively, acquired through sampling or other experience that permits learning (Barron and Erev, 2003; Hertwig et al., 2004; Weber et al., 2004).

The distinction between description and experience is pertinent for a wide range of human decisions because, in everyday life, people tend to acquire information about risks via both description and experience. Practitioners, such as doctors, insurance brokers or investment advisors, often provide clients with written numerical information about different types of risk. Yet, people also continually learn about risks from a multitude of experiences: examples include seeing your investments go up and down; observing people returning from skiing trips with injured limbs; and living through one more day without being burgled or mugged. Hence, if there is a significant DE gap, it may influence many economic decisions. With that in mind, our primary motivation in this paper is to assess what sort of, and how serious, a challenge the DE gap poses for theoretical and applied economics. We do this by using experimental techniques that allow us to investigate the contributions of different possible causes of the DE gap and to measure its footprint in choices and elicited risk-preference functions.

While existing research provides widespread evidence of DE gaps in experimental studies, the exact form of the phenomenon, and the implications it holds for economic analysis, remain controversial. For example, estimates of the size and even the direction of the gap vary across studies. Moreover, while there is ample evidence that misperceptions of objective probabilities in decisions from experience (due to biases in information captured in sampling experiences) explain some component of the DE gap, it is less clear whether – and, if so, how far - other contributory factors related to preferences and/or cognitive processes also play a role. We discuss the relevant evidence in the next section, simply noting here that our experimental design is motivated by two factors underlying diversity in the prior evidence: that the DE gap may be influenced by multiple, importantly distinct, causal mechanisms that have been triggered differentially by competing designs; and that different studies have measured the gap in different ways. For example, some studies use measures of the gap based on choice frequencies alone whereas other studies rely on parameter comparisons within particular preference models. We refer to these approaches as model free and model mediated respectively.

We contribute to the literature, firstly, by presenting a new experiment which tests for the operation of four distinct causal mechanisms in a unified design; and, secondly - building on a novel insight of Kopsacheilis (2018) – by exploiting both measuring approaches when assessing the effects of the mechanisms and the DE Gap itself.

The mechanisms we examine are, respectively: sampling bias; ambiguity attitudes; the form of representation of probability information; and the effects of memory. The first of these mechanisms explains the DE gap in terms of differences in the *information* available to decision makers at the point of choice, comparing description and experience; the second attributes the DE gap to features of *preferences*; the third and fourth effects explain the DE gap as arising from features of human *cognition*.

These channels are not mutually exclusive, as we explain. Yet, identifying which actually operate, and to what degree, is important for economics because the implications for economic theory vary markedly depending upon which of the information, preference or cognition channels are most at play. If the DE gap is simply caused by differences in information about objective risks that result from properties of small samples, that would be a reason to consider information available to agents, but not a fundamental challenge to preference theory. If the DE gap arises from ambiguity sensitive preferences, it would become an important, but so far under-appreciated, part of the rationale for the models of such preferences that have emerged in the last 30 years. However, if the DE gap is caused by cognitive processes and constraints, a full understanding of it may require models of decision processes, rather than pure preference models.

Our main findings are as follows. Our model-free analysis replicates a significant DE gap, similar in magnitude and direction to the literature's average. Also, in line with existing literature, we find that sampling bias contributes importantly to the DE gap. In fact, in our experiment, sampling bias - in the form of under-representation of rare events - is the only one of the four causal factors we consider that generates a statistically significant gap by itself in our model-free analysis. Our model-mediated analysis uses the framework of rank-dependent expected utility theory (RDEU; Quiggin, 1982, Wakker, 2010) to capture effects of our causal factors on probability-weighting, while allowing for any such effects on utility curvature. It supports two findings. First, in all treatments that control for sampling bias, we find inverse-S probability weighting, consistent with overweighting of rare events for both description and experience; by contrast, sampling bias tends to create the appearance of more linear decision weights via the under-representation of rare events in experienced samples. Second, we find some evidence of DE gaps caused by factors besides sampling bias: this arises from treatment comparisons that implicate a mixture of cognitive factors and ambiguity sensitive preference.

In the next section, we discuss existing literature. This provides the background for our experimental design. Section 3 presents our experimental design and details on the methods of analysis. In Section 4 we show results, with discussion and conclusions in Section 5.

2 Background

Much of the evidence for the DE gap derives from lab experiments using variants of the so called 'sampling paradigm' (Hertwig et al., 2004) in which participants make one-off choices between safer and riskier options in one of two different treatments: description or experience.[1]

In description, gamble properties are fully stated, leaving no uncertainty for the chooser regarding possible payoffs or their associated probabilities. In contrast, in experience, participants are not given stated information about consequences and/or their probabilities but must garner it via some form of sampling. In a typical implementation of experience, the two gambles might appear on screen in the form of two buttons. Participants then sample by pressing the buttons in some sequence of their choice and, when a button is pressed, one of the outcomes of the selected gamble appears on screen with outcome likelihoods controlled by the gambles' objective probabilities. Note that, in this framework, relative frequencies of experienced outcomes may not always coincide with the objective probabilities (though in some designs, as in some of our treatments, they may be controlled to do so).

In this framework, a standard test for the DE gap has been to compare choice proportions across the two conditions. The 'canonical finding' is that subjects in the description condition tend to prefer the riskier option when the rare event gives a desirable outcome, and to prefer the safer option when the rare event gives an undesirable outcome; whereas the opposite is observed in the experience condition. Taken together, this pattern has been commonly interpreted as reflecting a tendency to overweight rare events in description but to underweight them in experience (Hertwig et al., 2004).

There is now considerable amount of research investigating the DE gap, with a recent

[1]Notably, however, there are alternative paradigms that have studied the DE gap (e.g. the 'partial-feedback' paradigm, Barron and Erev, 2003).

meta-analysis (Wulff et al., 2018) adding authority to the claim that the DE gap exists. However, this meta-analysis also demonstrates striking heterogeneity with respect to the size of the gap, ranging from very small to very large. In fact, some papers even find a reversed DE gap, with subjects in experience appearing to overweight rare events more than in description (e.g. Glöckner et al., 2016). How can we make sense of these diverse findings?

We suggest that one contributor to the diversity of findings is the wide variation in design features such as the structure of sampling, characteristics of gambles and the ways in which they are evaluated (e.g. choice tasks versus valuation tasks). A second contributor is the fact that different studies have employed different measurement approaches to quantify their findings. We now expand on these points and show how our study contributes to the understanding of the DE gap through its response to them.

The idea that variation in study design accounts for the variation in measured DE gaps is all the more plausible given that existing literature has suggested several potential causes of the DE gap. To the extent that there are multiple causes at work, different designs may have triggered subsets of them to different degrees. We taxonomise causal factors that may drive the DE gap into three categories: sampling bias; preferences; and cognition.

Sampling bias is perhaps the most obvious potential candidate explanation. This attributes the DE gap to individuals acting on the basis of biased information in experience treatments. As already noted, in experience treatments, the relative frequency with which gamble outcomes are observed may not always match their objective probabilities. Moreover, because people usually choose to collect only quite small samples in experience treatments (e.g. the median subject of Hills and Hertwig, 2010 samples each option only 9 times), rare events are systematically under-represented due to a property of the binomial distribution.[2] In such circumstances, we should expect the impact of rare events on choices to be

[2] As a simple demonstration, consider for example drawing a single ball from an urn that contains 90 black and 10 red balls. On average, red balls will be under-represented in 90% of such single-observation (small) samples.

attenuated, in line with the canonical finding.

There is considerable existing evidence that sampling bias contributes towards the DE gap (Fox and Hadar, 2006; Rakow et al., 2008). Were sampling bias the full story, the significance of the DE Gap would largely derive from the potential for sub-optimal search intensity by economic agents and the dangers of environments that generate biased information. But there is evidence that DE gaps can also arise in the absence of sampling bias, from studies that control for sampling bias by engineering experience treatments to ensure that experienced and objective probabilities coincide (e.g. Hau et al., 2010; Ungemach et al., 2009; Barron and Ursino, 2013; Aydogan and Gao, 2019). DE gaps observed in such setups require an explanation that goes beyond biased information, prompting consideration of accounts that attribute some component of the DE gap to features of either preferences or cognitive processes or both.

The most obvious candidate for a preference-based account of the DE gap is some form of attitude toward ambiguity (see Etner et al., 2012, for a review of the theoretical literature on modelling ambiguity sensitive preferences). This is so because, in terms of the classic Knightian distinction (Knight, 1921), decisions in a description treatment are choices among 'risks' whereas those in an experience treatment are more naturally interpreted as involving other forms of uncertainty, in which probabilities are ambiguous or only imprecisely known. If agents are (subjective, where necessary) expected utility maximisers, the distinction would be irrelevant in situations where experienced and objective probabilities coincide. But, the presence of ambiguous information about probabilities may affect behaviour if individuals have non-expected utility attitudes towards ambiguity. For example, as in Ellsberg's famous urn experiments[3] where people are often ambiguity averse in the sense of being more willing to gamble on 'known' than 'unknown' urns, willingness to take risks may be lower in experience (where distributions are unknown) than in description. There is some existing evidence that

[3]See Trautmann and Van De Kuilen (2015) for a recent review of the subsequent literature

ambiguity attitudes play a role in the DE gap. Specifically, Abdellaoui et al. (2011b) find that, in the absence of sampling bias, estimated decision weights for a prospect theory model are systematically smaller (less optimistic) in experience compared with description. While this result might be due to ambiguity aversion, so far, the evidence for such an effect being an important driver of the DE gap is limited. If the finding were to generalise, the DE gap might be an important exhibit of ambiguity sensitivity, ranking along side the Ellsberg paradox in that capacity.

A third class of explanation attributes the DE gap to factors that have their roots in human cognition (as opposed to preference). We consider two such candidates: *likelihood representation* and *memory*. Recall that gamble information is represented in different ways across description and experience. In description, probabilities are communicated through written information often in the form of percentages (e.g. '£16 with 10% chance') but in experience, gamble information is obtained through sequential sampling experiences which must be interpreted by the receiver and may result in perceptions (e.g. 'this option gave me a good prize one out of 10 times'). While there is considerable evidence that representation of chance can affect decisions in different contexts (Gigerenzer and Hoffrage, 1995; Slovic et al., 2000), it is not yet clear how important differences in likelihood representations are as drivers of DE gaps, when the information represented is held constant. A related consideration arises from noticing that, when likelihoods are discovered through sequential sampling, claims about what information subjects have in mind are contingent on assumptions about their recall. As such, imperfect memory of sampling is a further possible driver of the DE gap. While the possible role of imperfect memory in the DE gap has been noted in previous literature, its actual role is hard to assess based on existing evidence (see Wulff et al., 2018, for a discussion).

With respect to differences in the way that DE gaps have been measured, while most studies in this literature have used direct choice comparisons to assess the DE gap (e.g.

comparing choice proportions as described above), studies in a slightly different genre have estimated behavioural models (usually based on cumulative prospect theory; Tversky and Kahneman, 1992) to examine the impact of description versus experience on parameters of estimated preference functions (especially the parameters that control the shape of decision weighting functions). It seems possible that different measurement approaches may support different conclusions. Notwithstanding this possibility, there remains considerable variation across the results of studies even within each of these genres. For example, while DE gap studies that estimate prospect theory weighting functions have generally reported inverse S-shaped probability weighting curves in description, there is considerable variability in the shapes of curves elicited in experience conditions: Abdellaoui et al. (2011b) report inverse-S shaped weighting; (Ungemach et al., 2009) report S-shaped weighting; while Hau et al. (2008) find linear weighting. More recently, Kopsacheilis (2018) put forward the 'Relative Underweighting Hypothesis', according to which, people overweight rare events in experience but less so than in description. This hypothesis is accommodated by an inverse S-shaped weighting function in experience that is closer to the diagonal for probabilities closer to 0 or 1, when compared to description.

Our experiment is designed to facilitate direct tests for the influence of each of the four explanatory factors just discussed (sampling bias, ambiguity attitude, likelihood representation and memory). These factors are tested by pairwise comparisons of treatments in a unified design which, by varying a single factor in each comparison, allows assessment of their isolated influences. Our setup is also designed to facilitate evaluation via model-free tests of effects and via comparisons of the impact of factors on estimated probability weighting functions. By using four different Experience-treatments, our design also permits tests of four distinct forms of DE gap.

3 Design & Methods

3.1 Treatments

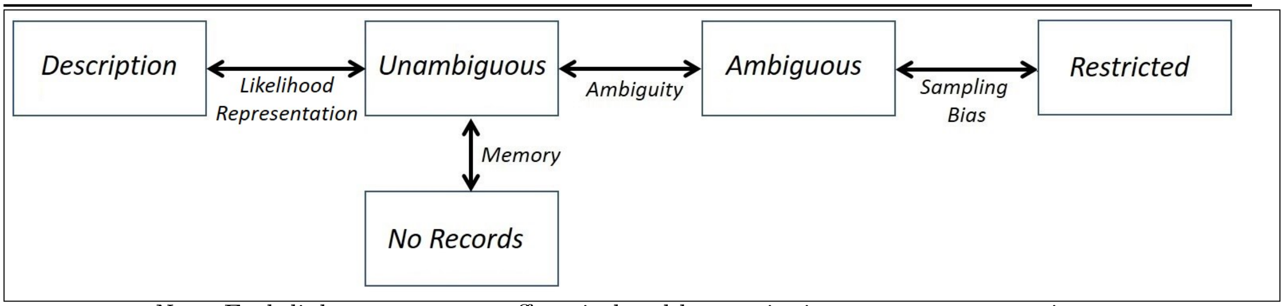
In our experiment, subjects evaluate a series of binary gambles via a process described below. Payoffs are (non-negative) money amounts which are always known to the decision maker at the point of evaluation. Gambles are represented by virtual decks of cards, each containing two types of card, demarcated by colour.[4] Within a gamble, each outcome is associated with one of the two colours in the deck and outcome probabilities are equal to the relative frequencies of the corresponding colours.

The design involves five treatments: one *Description* (Desc) treatment plus four variants of experience which we label *Unambiguous* (E-Unamb), *No Records*, (E-NR) *Ambiguous* (E-Amb), and *Restricted* (E-Res). As we summarise in the top part of Figure 1, the treatments differ in how subjects obtain information about the contents of the deck. We summarize these differences in the top part of Figure 1.

[4]Each lottery's outcomes are demarcated by a different pair of colours – Figure 2 displays one such pair. The correspondence between colours and outcomes for each lottery was randomized for each subject. This was done to avoid systematic influence of connotations associated with particular colours such as 'danger' with red or 'environmental risk' with green. Moreover, the order of cards within each sample from each deck was randomised by subject.

Figure 1: Summary of treatments and treatment - comparisons

Desc	Subjects see numerical (percentage) statements of likelihood
E-Unamb	Subjects knowingly sample the entire deck with a history table
E-NR	Same as **E-Unamb** but without the history table
E-Amb	Same as **E-Unamb** but subjects are unaware that they sample all cards
E-Res	Same as **E-Amb** but sampling amount is restricted introducing sampling bias



Note. Each link represents an effect, isolated by a pairwise treatment-comparison.

In the description treatment (Desc), gamble probabilities are communicated in explicit, numerical form (as percentages) during evaluation (e.g. '*90% of those cards are grey and 10% of those cards are yellow*'). By comparison, in the experience treatments, subjects are not told the relative frequencies of the colours in each deck but have opportunities to discover this information by sampling deck contents. The sampling environment varies by treatment as we now explain.

The E-Unamb treatment is intended to provide a version of experience which is informationally equivalent to Desc. This involves two key ingredients. The first is that, in E-Unamb, subjects sample the entire deck, without replacement, and are told that they see the full deck with each card appearing once and only once. Hence, in principle, subjects in this treatment have access to full information about the chances of the two outcomes which is logically identical to that available to subjects in Desc. However, subjects having seen the full set of cards exactly once is no guarantee that subjects have accurate perceptions of the colour composition of the deck at the point of gamble evaluation: they may not have paid full attention to the sampling experience and they might have forgotten aspects of it, prior to the evaluation phase. To control for the influence of such cognitive constraints in E-Unamb, we introduce, as the second key ingredient, a *history table* that remains on screen during the evaluation phase.[5] This records the colours of cards that were sampled in the order they were sampled. This record is shown on the screen where subjects evaluate gambles and, for this treatment, a message on top of the history table reads: "*This is the entire deck with its cards displayed in the order you sampled them*".[6]

Figure 2: Instances of E-Unamb's interface

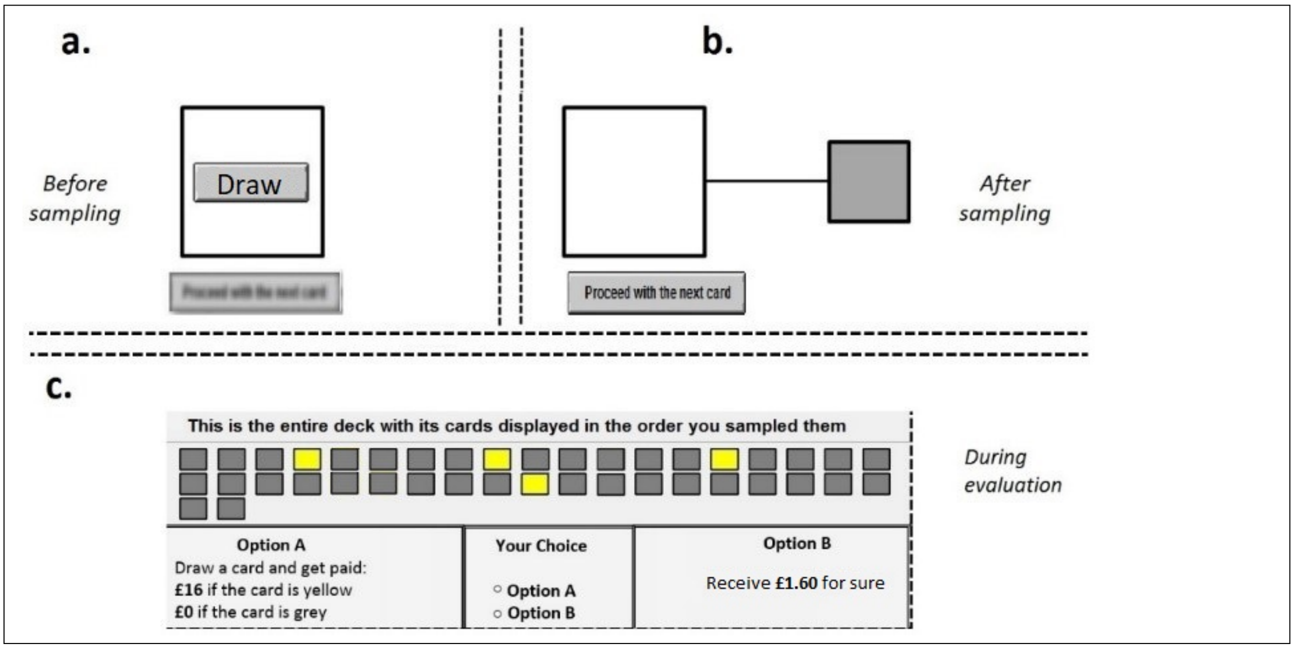


Figure 2 illustrates how the sampling process was displayed to subjects, depicting three instances of the experimental procedure for E-Unamb. Panels a. and b. capture before and after instances of a single sample event, while panel c. demonstrates an example of the evaluation phase. Notice that the history table encodes mathematically identical probability information to that provided, in a different format, in the Desc treatment. Hence, comparing behaviour in the Desc and E-Unamb treatments provides a test of whether behaviour depends on the way in which likelihood information is represented.

[5]The technology of this memory aid was introduced by Kopsacheilis (2018). However, unlike that paper, sampling amount is fixed (not an endogenous decision) in our study. A similar device was used in the past also by Hau et al. (2010) but with a markedly different technology.

[6]See Appendix 6.1 for details of instructions.

As illustrated at the bottom of Figure 1, treatments branch along two different routes as a consequence of variations relative to E-Unamb. The E-NR treatment is identical to E-Unamb, except that no history table is presented. Hence, while the information contained in the sampled deck remains equivalent to Desc and E-Unamb, memory or other cognitive limitations (including lack of attention) might lead individuals to be acting on assessments of gambles based on misperceptions of objective probabilities. As a convenient shorthand we refer to any such influences of cognition, as 'memory' effects. The comparison of E-Unamb with E-NR isolates such effects.

E-Amb branches in a different way from E-Unamb. E-Amb retains the history table and all other features of E-Unamb except that, in E-Amb, subjects are not told that the 40 cards they sampled comprise the entire deck. In this treatment, the line of text immediately above the history table just says: "*These are the colours you sampled in the order you sampled them*". Hence, while these subjects do in fact see the full deck and the record of it, they do not know that they see the full deck. So, from their perspective, the situation has a degree of ambiguity because they are not informed that the relative frequencies they experience are accurate. Hence, the comparison between E-Unamb and E-Amb isolates the effect of the presence of ambiguity, keeping constant the actual samples experienced and the presence of the history table record of them. This is our cleanest test for the impact of ambiguity. However, if subjects are ambiguity sensitive and also (aware that they) suffer from imperfect recall, they might experience ambiguity in E-NR too. Our shorthand 'memory' should be interpreted as including this additional effect of withdrawing the history table, if it is present.

Our final treatment, E-Res, is identical to E-Amb except that the number of cards sampled was restricted. Specifically, unlike E-Unamb, E-NR and E-Amb that featured 40-card decks, the decks in E-Restr were restricted to 18 cards. In consequence, a unique feature of this treatment is that the experienced relative frequency of colours sampled cannot match the objective one.[7] Therefore, the E-Res treatment necessarily introduces sampling bias and the comparison of it to E-Amb isolates the effect of that factor. Notice that sampling bias can arise in two directions: a particular event can be either over- or under-represented in a given sample, relative to its objective probability. Since we should expect the effects of under and over-representation to be different, in the analysis we split observations in E-Res into two subsets: E-Over and E-Under. Since rare events are the loci of our interest, we taxonomize observations in Over and Under according to whether the event with the smallest probability to occur was over or under-represented.[8]

We are now in a position to summarise the full logic of the set of treatments introduced in Figure 1. In essence, pairwise comparisons of treatments which are adjacent in the bottom panel of Figure 1 provide a series of tests designed to isolate effects due to each of the factors: *likelihood representation*, *memory*, *ambiguity* and *sampling bias*. We refer to these as tests for 'effects'. Since we have multiple variants of experience implemented in our design, and since it is possible that the factors we seek to isolate might also work in combination, we also conduct a set of tests for DE gaps by comparing behaviour in our description treatment with that in each of our different experience conditions. Lastly, in order to get an estimate of the average DE gap we elicit, we compare Desc with E-All, a compilation of observations across all 4 variations of Experience.

Our design for these 4 variations of experience bears some noticeable differences with that of the Sampling paradigm. For example, during the sampling phase, subjects explore only one source of uncertainty at a time instead of two. The second option is always a money amount offered with certainty. Moreover, subjects make repeated choices for each lottery in our decision set - instead of a one-off choice - so that we can infer an indifference between the risky-option and a certain amount. The advantage of these adaptions is that they allow

[7]As will become clearer in the following subsection, the set of objective probabilities that we chose for this study cannot be accurately represented in samples of 18 observations.

[8]For the one 50-50 gamble, we classify the observation according to observed relative frequency of the event corresponding to the better outcome.

us to elicit a more precise account of subjects' risk-preferences (see Abdellaoui et al., 2011b for further discussion).

Another source of variation is that the decision to stop sampling is not endogenous. Instead, subjects draw a fixed number of cards (40 or 18) without replacement. Crucially, this gives us complete control over the information they obtain by sampling. It is interesting to relate this feature of our sampling technology to the one employed by papers such as Aydogan and Gao (2019) and Barron and Ursino (2013), which use urns (or boxes) containing balls instead of decks of cards to resolve uncertainty. For example, in the Description treatment of Aydogan & Gao, subjects were fully informed of the contents of the urn. In their sampling treatment, subjects sampled every ball from the urn in sequence. They were not provided with a record of the sampling, but were allowed to take their own notes (and in fact subjects varied in how far they did so). A subject who did keep a complete record would be in a position akin to our E-Unamb treatment; whereas one who kept no records would be in a position akin to our E-NR treatment; yet the position of one who kept an incomplete record is harder to characterise in our terms. In this sense, their Sampling treatment is a hybrid of our Experience conditions. Our design gives us more control of the information subjects have in each of these treatments and, by using four different Experience treatments, enables us to isolate separate effects more clearly, in the way we have explained.

3.2 Incentives and other procedures

We now explain how gambles were evaluated by subjects. An example of the evaluation phase is depicted in Figure 2. This illustrates the evaluation of a gamble (denoted option A in this figure) which gives a 10% chance of winning £16 (otherwise zero). Note that while the presentation of probability information on this screen would have differed between treatments (the case shown is for the E-Unamb treatment), once probability information was acquired, the protocol for evaluation of gambles was then essentially the same for all five treatments. Gambles were evaluated by a series of comparisons against various certain sums of money (such as Option B in the figure). We achieve this by implementing a version of the bisection method of Abdellaoui et al. (2011b) in which sums of money are updated according to the subject's previous choice. At the first iteration of each evaluation, the certain amount is set equal to the expected value of Option A. In the second iteration, this amount is revised upwards (downwards) to the mid-point of the gamble's highest (lowest) outcome and the certain amount just rejected (accepted). Via this process, for each gamble a subject evaluates, we impute to them a certainty equivalent which is the midpoint of the certain amount of Option B after 5 such iterations and the certain amount that would have been displayed under Option B if a 6th iteration were to take place.[9]

The set of lotteries evaluated is summarized in Table 1. We selected these lotteries in order to comply with the semi-parametric estimation protocol of Cumulative Prospect Theory that was implemented by Abdellaoui et al. (2011b) (see 3.3.2 for more details). One noticeable adaption from the set of lotteries suggested by Abdellaoui et al. (2011b) is that we increase the number of lotteries involving rare events - a feature that allows us to zoom in further in this region of probability weighting.

The order of these lotteries was randomized within two clusters for each subject. Lotteries in the first cluster $(1.1 - 1.7)$ had varying outcomes but with a winning probability fixed at $p = 0.25$. To make this common structure clear to subjects in the experience treatments, this first cluster of lotteries was associated with only one deck and one sampling process. Seven evaluations were then based on that one sampling process. Lotteries in the second cluster $(2.1 - 2.9)$ had a pair of fixed outcomes and varying probabilities. A subset of this second cluster $(2.4 - 2.9)$ feature 'rare' events which will be important in our analysis. Following the convention in this literature we consider an event rare if its corresponding probability

[9] See Appendix 6.2 for a demonstration of the bisection method.

is less than $p < 0.20$ (Hertwig et al., 2004, see). Notice that all of the lotteries with rare events have just one non-zero payoff and sometimes the rare event is associated with the desirable prize [lotteries $2.4, 2.5, 2.6$] and sometimes the rare event is undesirable [lotteries $2.7, 2.8, 2.9$]. The role of lotteries without rare events will emerge in the next sub-section.

Table 1: Decision problems and characterisation

Decision Problem	Risky	Safe (1^{st} iteration)
1.1	(4, 0.25; 0)	1.0
1.2	(8, 0.25; 0)	2.0
1.3	(12, 0.25; 0)	3.0
1.4	(16, 0.25; 0)	4.0
1.5	(16, 0.25; 4)	8.0
1.6	(16, 0.25; 8)	10.0
1.7	(16, 0.25; 12)	13.0
2.1	(16, 0.25; 0)	4.0
2.2	(16, 0.5; 0)	8.0
2.3	(16, 0.75; 0)	12.0
2.4	(16, 0.025; 0)	0.4
2.5	(16, 0.05; 0	0.8
2.6	(16, 0.1; 0)	1.6
2.7	(16, 0.90; 0)	14.4
2.8	(16, 0.95; 0)	15.2
2.9	(16, 0.975; 0)	15.6

Note. We follow the $(x, p; y)$ lottery notation. This represents lotteries that offer £x with probability p or £y otherwise. Decision problems in grey cells contain a rare event.

In total, 198 participants were recruited through ORSEE (Greiner, 2015)) and randomly assigned to one of the five treatments summarized in Table 1. The experiment was programmed in Z-tree (Fischbacher, 2007) and sessions were conducted in the CeDEx laboratory (University of Nottingham) and lasted for approximately one hour. Subjects' payments depended on their choices and on gamble resolutions. At the end of the experiment, one choice was selected at random for payment.[10] If participants had chosen the Safe option, then they would receive the corresponding certain amount. Otherwise, if they had chosen the Risky option, they would play out the lottery. In this case, an integer number - between 1 and 100 - would be randomly generated and displayed on screen. If the number was smaller

[10] This is a standard procedure see Cubitt et al., 1998; Bardsley et al., 2010, ch 6.5; for a relevant discussion.

than or equal to the specified chance of winning the high amount in this lottery, participants would receive the high amount, otherwise they only received the low amount. On average, subjects were paid £11.50 including a flat £2 participation fee.

3.3 Methods of analysis

3.3.1 Model-free methods

In the model-free analysis, we make cross-treatment comparisons using tests that do not rely on any particular behavioural or preference model, but instead let the raw choice data speak. For this analysis we use only the data from the first iteration of each bisection in the evaluation phase. Recall that these are choices between the gamble (risky choice) and the certain amount (safe choice) equal to the gamble's expected value. This choice structure is similar to that of the early studies in the sampling paradigm. As these early studies focused only on situations involving rare events, for comparability, this part of our analysis will focus only on the subset of decision problems involving lotteries containing a rare event (those highlighted grey in Table 1).

Following comparable approaches in the literature, we summarise each individual's behaviour through an *overweighting score*. The score is constructed, for each individual, based on their evaluations of the six gambles which feature rare events. Consider a binary index: $C_i \in \{0, 1\}$, with i indexing one of the 6 problems in Table 2 that contain a rare event. $C_i = 1$ (0) when the subject's choice in decision problem i is consistent with overweighting (underweighting) of rare events. A choice is consistent with overweighting when the riskier option is selected (over the safer one) when the rare event was desirable or when the safer option was selected when the riskier alternative featured an undesirable rare event. We then calculate the overweighting score as:

$$\%\text{OVRW} = \frac{1}{6} * \sum_{i=1}^{6} C_i * 100$$

We interpret the %OVRW score as a measure of the propensity to overweight, which varies from 0 (no choice was consistent with overweighting) to 100 (all choices were consistent with overweighting).[11] While this index is in some ways simplistic[12], it is intuitive, does not require committing to any specific preference model, avoids issues associated with inference from repeated observations and, allows us to benchmark against behaviour reported in earlier literature, where the DE gap was established, using comparable measures. Using this measure, we test for effects and for DE gaps by comparing the average %OVWR scores across the individuals facing each relevant treatment.

3.3.2 Model-mediated methods: RDEU

To the extent that the DE gap reflects variation in the weighting of events across different environments, it is natural to consider modelling them using theories which, in the tradition of prospect theory, embody a concept of decision weighting. We follow this approach exploiting a simple and now rather standard RDEU framework. One important benefit of the model mediated analysis is that it allows for more refined inferences, e.g. by separating effects that come through utility curvature and ones that come through probability weighting. In our design, a second advantage of this level of analysis is that it takes into account a richer information set, incorporating all 5 iterations of the bisection method (instead of only the first one) and more probability targets (instead of only the ones containing a rare outcome). A third benefit is that this analysis facilitates comparison with more recent literature on the DE gap which has exploited related approaches.

[11] Glöckner et al. (2016) refer to the same index as p(overweighting) and interpret it as the probability of making a choice consistent with overweighting.

[12] We are not entitled to assume that choices consistent with overweighting are fully explained by overweighting as other factors may be at work.

To formalise our approach, consider binary lotteries of the form $x_{E_p}y$ which give one of two monetary outcomes x, y where $x > y > 0$. Outcome x arises in event E which occurs with probability p; otherwise the outcome is y. The rank dependent utility of any such lottery is given by the expression:

$$W(E_p)u(x) + (1 - W(E_p))u(y) \tag{1}$$

where $u(\cdot)$ is a strictly increasing utility function and $W(\cdot)$ is a weighting function and $W(E_p)$ is the decision weight associated with event E_p. This model reduces to expected utility theory in the special case where $W(E_p) = p$, for all events. For the class of lotteries considered in this study, it also coincides with a range of common non-expected uility models (Tversky and Kahneman, 1992; Ghirardato and Marinacci, 2001; Luce, 1991; Miyamoto, 1988).

To study the DE gap using the RDEU model, we follow the source method (Tversky and Fox, 1995; Abdellaoui et al., 2011a) which was specifically adapted for this purpose by Abdellaoui et al. (2011b). A key feature of this approach is to allow probability weighting functions to depend on the source of uncertainty. So, for example, different weighting functions might apply to decisions under risk than apply under different forms of experienced uncertainty, even if the underlying probability distributions over outcomes are otherwise identical. In our setting, we apply this idea by interpreting our various treatments as potentially different sources.

More formally, for an event E_p, such as drawing a yellow card from a deck in a specific treatment, where $(100 \times p)\%$ of its cards are yellow, the decision weight is given by:

$$W(E_p) = w_\sigma(\pi(E_p)) \tag{2}$$

In Equation 2, w_σ is a source function which transforms probabilities into decision weights according to the source of uncertainty, σ. In this expression, $\pi(\cdot)$ is the individual's belief of the likelihood of E_p. In line with standard practice, we assume that in description environments, $\pi(E_p) = p$.

In experience on the other hand, this belief depends on a variety of other factors, including the relative frequency (f_p) of each event E_p that is observed by the individual. Following common practice, we assume that $\pi(E_p) = f_p$ for decisions from experience. Under this assumption, Equation 2 can be re-written as:

$$W(E_p) = w_\sigma(f_p)) \tag{3}$$

As, in three of our four variations of Experience, we control for sampling bias we can set $f_p = p$ in these cases. This allows us to further simplify Equation 3 into:

$$W(E_p) = w_\sigma(p)) \tag{4}$$

In our treatment E-Res, $f_p \neq p$, by construction. Although in principle, operating under Equation 3 for E-Res could allow us to control out the role of sampling bias, this would defy the purpose of this treatment, i.e., quantifying the effect of sampling bias. Therefore, we choose to operate under Equation 4 for E-Res too, thereby incorporating a sampling bias of magnitude: $|p - f_p|$. This point becomes clearer in the Results section.

In our analysis, we estimate utility curvature and decision weights at the individual level. We use the seven certainty equivalents elicited from evaluations of risky options in problems $1.1 - 1.7$ to fit the utility curvature parameter of a power utility function: $U(x) = x^\alpha$. We do so by minimizing the non-linear least squares: $\sum_{j=1}^{7}(z_j - \hat{z}_j)^2$, where z_j refers to the observed certainty equivalent of the risky option appearing in decision problem j, with j ranging from 1 to 7, and $\hat{z}_j$ is the estimated certainty equivalent. From Equation 1, using $u(x) = x^\alpha$, we obtain the following expression for $\hat{z}_j$ which we can use to estimate α for every subject:

$$\hat{z}_j = [W(E_{p^*})(x_j^\alpha - y_j^\alpha) + y_j^\alpha]^{\frac{1}{\alpha}} \tag{5}$$

An important feature of this estimation protocol is that the event E_{p^*} corresponding to outcome x_j is common for $j = 1, \ldots, 7$ (and therefore associated with the same probability: p^*). Therefore, we can treat the corresponding decision weight: $W(E_{p^*})$ as a free parameter to be estimated together with the utility curvature parameter.

Having obtained an estimate of each subject's utility curvature, we proceed to calculate decision weights (non-parametrically) for the risky options in 2.1 - 2.9. Notice that these options have fixed outcomes: $x^* = 16$ and $y^* = 0$ and varying probability: p_r, with r indexing the risky option in decision problems 2.1 - 2.9. Using Equation 5, we can therefore calculate decision weights for each probability level: p_r, according to:

$$W(E_{p_r}) = \left(\frac{z_r^{'}}{x^*}\right)^\alpha \tag{6}$$

where $z_r^{'}$ is the elicited certainty equivalent for risky option r ($r = 1, 2, \ldots, 9$). Taking the median weight across individuals, we obtain an aggregated source function under each treatment. By studying the shape of the elicited weighting curves and comparing them across treatments, we examine the DE gap and its driving forces from the perspective of a model that allows for probability weighting. This combination, derived from Abdellaoui et al. (2011b) and explained above, allows us to control for utility curvature while letting the data speak on the exact form of probability-weighting in different treatments.

4 Results

4.1 Model-free analysis

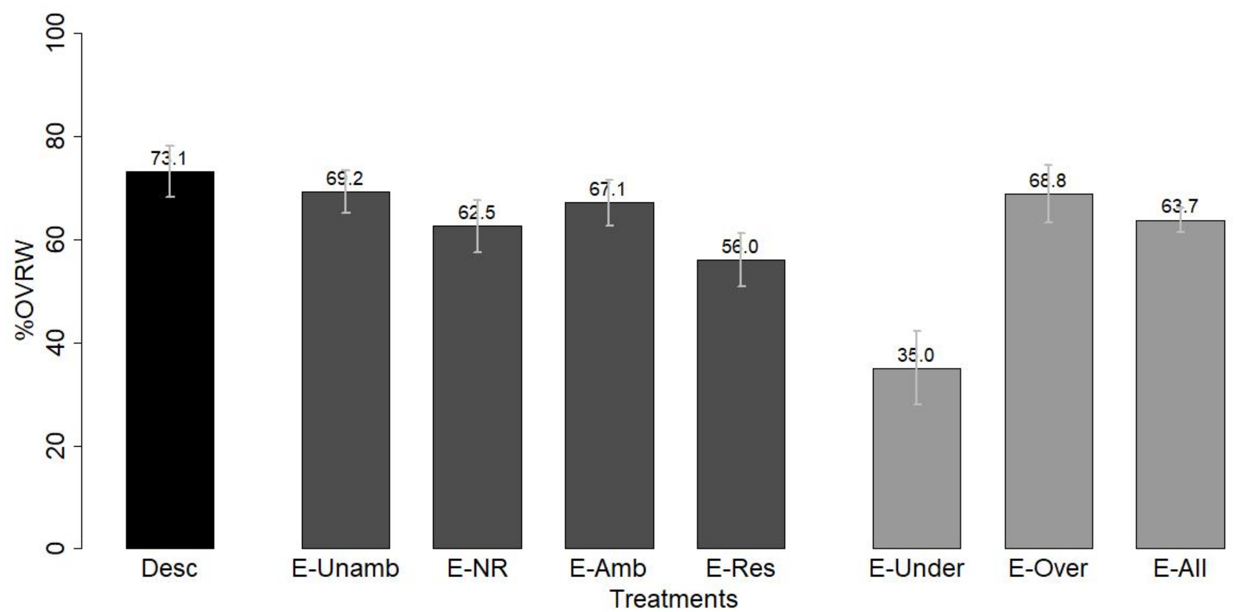
We begin our analysis by examining choice proportions through the lens of the %OVRW scores. Recall that these scores derive from choices made in the first iteration of each bisection process and, in line with the discussion of Section 3.3.1, we interpret the average %OVRW for each treatment as representing a treatment-level propensity to overweight rare events. The results are presented via Figure 3.

We highlight three features of the data evident from the histogram. First, and in line with the canonical finding, the propensity to overweight is higher in the Description treatment than in any variant of Experience. Second, every extra feature of experience that our design introduces, moving left from Description across the four Experience treatments, leads to a lower %OVRW score, suggesting an expanding DE gap (we explore the statistical significance of these changes shortly). Finally, a third salient feature of this plot is the comparatively low score for E-Under: this is consistent with the intuition that rare-events carry less weight in decisions from experience, when they are under-represented in the sample.[13]

The Table in the bottom half of Figure 3 details the size and statistical significance of the various DE gaps and effects that our experiment was designed to measure. Its top row reports the overall average DE gap in our experiment by comparing Desc vs E-All. This measure of the gap is statistically significant (P = 0.041; MW) and its size at 9.35 percentage points is very close to the literature's average of 9.7 percentage points, based on a large meta-analysis from 80 data sets (Wulff et al., 2018). We view this close correspondence to the existing evidence as a reassuring indication that we are capturing a familiar DE gap in our experimental setup. In line with the canonical finding, we find less overweighting of rare

[13]See Appendix 6.4 for more details on sampling bias and its interaction with %OVRW.

Figure 3: Average %OVRW scores across treatments



	Treatment comparisons	Difference in %OVRW	$P - value^{+}$
	Desc vs. E-All	9.35**	0.041
DE gaps	Desc vs. E-Unamb	3.91	0.284
DE gaps	Desc vs. E-NR	10.58	0.121
DE gaps	Desc vs. E-Amb	6.00	0.209
DE gaps	Desc vs. E-Res	17.09***	0.008
	Desc vs. E-Over	4.87	0.555
	Desc vs. E-Under	39.34***	0.000
Effects	Likelihood representation	3.91	0.284
Effects	Memory	6.67	0.468
Effects	Ambiguity	2.08	0.727
Effects	Sampling Bias	11.10	0.126
	Overrepresentation	-1.73	0.579
	Underrepresentation	32.13***	0.000

Note. Top: Average %OVRW scores across treatments (Desc, E-Unamb, E-Amb, E-NR, E-Res) and treatment-derivatives (Over, Under, E-All), with standard errors.
Bottom:+ Reported P-values from 2-tailed, Mann-Whitney (MW) U tests on %OVRW across treatments. $^{***}P < 0.01$;$^{**}P < 0.05$;$^{*}P < 0.1$.

events in Experience than in Description. These observations lead to Result 1.

Result 1 *We replicate a DE gap. It has the same direction as the canonical finding and, in terms of the %OVRW measure, a very similar size to the literature's average.*

The next six rows of Figure 1 test for a set of DE gaps via pairwise comparisons of the %OVWR score for Description with each of the individual Experience treatments including the two sampling bias derivatives (E-Over and E-Under). Across these comparisons, statistically significant differences are identified only in cases where sampling bias is present (i.e., the cases involving E-Res and E-Under). Not surprisingly, the gap is widest when rare events are under-represented rather than when they are over-represented. In comparisons that do not involve sampling bias (i.e., those comparing Desc with each of E-Unamb, E-NR or E-Amb), while the direction of each effect is in the typical direction of the DE gap, we find no statistically significant differences, although the comparison between Desc and E-NR is slightly bigger in magnitude than the literature average gap and approaching weak significance ($P = 0.121$; MW).

The bottom section of the table in Figure 3 provides analogous tests, but focusing on treatment comparisons that we interpret as capturing effects associated with specific mechanisms as summarised in Figure 1. Based on this analysis, while each factor again moves the %OVRW in the direction consistent with the canonical finding, the only significant effect is that associated with sampling bias – in the form of underrepresentation - where the effect is both large and highly significant. This leads to our second main result.

Result 2 *The most important and only statistically significant single driver of the DE gap in our data is sampling bias, in the form of under-representation of rare events.*

Result 2 adds to existing evidence identifying sampling bias as an important factor accounting for the DE gap. While the results we have presented so far are consistent with claims made elsewhere (Fox and Hadar, 2006; Rakow et al., 2008, e.g.) to the effect that sampling bias is the only significant driver of the gap, we hold short of such a firm conclusion at this point for two reasons. First, while no factor other than sampling bias was statistically significant in isolation, the measured DE gap nevertheless consistently widens with each factor introduced in our design. That is, the difference in %OVRW is positive for each gap and effect reported in the table of Figure 3, with the exception of over-representation where we expect an opposing effect. Moreover, combinations of other factors sometimes come close to producing a significant change in the gap (see the comparison between Desc and E-NR, which captures the combination of likelihood representation and memory). This suggests the DE gap might be partly driven by a range of other factors beyond sampling bias, even if these are relatively weak when operating in isolation. Second, we have a further analysis to present which involves a different and more detailed examination of the DE gap and its underpinning causes. We turn to this analysis now.

4.2 Model-mediated analysis

In this analysis, we use certainty equivalents derived from the bisection elicitation process to estimate a best fitting RDEU model for each individual. We estimate –parametrically - utility curvature first (as per Equation 5) and then calculate –non-parametrically- decision weights (as per Equation 6).

Table 3 reports median values for the utility curvature parameter (α) across treatments. In aggregate, our estimations suggest near linear utility over money which is a not-uncommon finding.[14] Median values are very similar across treatments and a Kruskal-Wallis test does not reject the null hypothesis of equal utility curvature across treatments ($P = 0.708$). Despite this, the size of the interquartile ranges (IQR) suggests that there was considerable heterogeneity of utility curvatures across individuals - a result that demonstrates the im-

[14]These estimates fall within the typical range of contemporary studies such as Abdellaoui (2000); Booij et al. (2010); Etchart-Vincent (2004); Murad et al. (2016), all of which find α lies between 0.8 and 1.1.

portance of our having controlled for this when assessing probability-weighting. We also consistently fail to reject the null of no difference in utility curvature in pairwise comparisons of treatments.[15] This suggests that potential treatment effects are more likely to occur due to differences in probability weighting rather than due to differences in preferences over money.

Table 2: Utility curvature estimates (α) across treatments (medians)

	Desc	E-Unamb	E-Amb	E-NR	E-Res
Median	1.06	1.08	1.10	1.08	0.98
IQR	0.84-1.38	0.82-1.60	0.83-1.35	0.81-1.69	0.66-1.18

Note. Parametric estimations of utility curvature: α from x^{α}. These estimates derive from a non-linear least squares algorithm (Bates and Watts, 1988; Bates et al., 1992; Moré, 1978), commonly specified for all 198 subjects: we estimate α for every subject 20 times with a randomly chosen starting value and select the iteration with thee best fit.

Next, we calculate decision weights $W(E_{p_r})$ for each subject at each probability level r, following Equation 6. Median values for these decision weights are reported in Table 3.

In this table, we use upward sloping arrows to indicate cases where estimated decision weights are statistically significantly above the diagonal line (i.e. the line consistent with linear decision weights where $W(E_{p_r}) = p_r$ for all events), and we use downward arrows to indicate cases where weights are significantly below the diagonal (in each case, the number of arrows indicates the critical value). The shaded cells highlight cases where weights do not deviate significantly from the diagonal. Although unconventional, this labelling makes it easy to see that, if we confine attention to cases that control for sampling bias (i.e. the first 4 of the 8 data columns in Table 1), then probability weighting generally takes an inverse S-shape: weights tend to be above the diagonal for small probabilities and below them for high probabilities with a cross-over at, or in the vicinity of, $p = 0.25$.[16] This inverse S-shaped probability weighting is consistent with a general tendency to overweight rare events.[17] This

[15]See Appendix 6.3 for more details on these tests and for a plot of all subjects' utility curves.

[16]Note that only one of the 36 cells in these four columns of data – that associated with E-NR for $p = 0.025$ - is inconsistent with this inverse-s pattern.

[17]Recall that rare events lie in two regions: for $p \in 0.025, 0.05, 0.10$ and for $p \in 0.90, 0.95, 0.975$. As p is

leads to Result 3.

Result 3 *Controlling for sampling bias and utility curvature, probability weighting is generally inverse S-shaped, consistent with overweighting of rare events in both Description and Experience treatments.*

Things look different when we consider cases with potential for sampling bias. In particular, E-Under stands out by having a pattern of median weights consistent with an S-shaped weighting function: for low probabilities, median weights are nominally below the diagonal, rising above it for high probabilities. If we confine attention to the analysis of statistical significance, however, almost none of the weights associated with rare events depart significantly from the diagonal. Hence, an interpretation of this analysis is that sampling bias is counteracting an underlying behavioural tendency to overweight rare events. That is, the overweighting is a property of the preferences which is disguised by the sampling bias.

the probability of the better outcomes, rare events in the first interval have desirable outcomes; rare events in the second region have undesirable outcomes.

Table 3: Median decision weights and comparison with the diagonal

p	Desc	E-Unamb	E-Amb	E-NR	E-Res$^+$	Over$^+$	Under$^+$	E-All$^+$
0.025	$0.096^{\nearrow\nearrow\nearrow}$	$0.041^{\nearrow\nearrow\nearrow}$	$0.048^{\nearrow\nearrow\nearrow}$	0.019^{ns}	$0.052^{\nearrow\nearrow\nearrow}$	$0.064^{\nearrow\nearrow\nearrow}$	0.012^{ns}	$0.039^{\nearrow\nearrow\nearrow}$
0.050	$0.121^{\nearrow\nearrow\nearrow}$	$0.100^{\nearrow\nearrow}$	$0.070^{\nearrow\nearrow}$	$0.061^{\nearrow\nearrow}$	$0.070^{\nearrow\nearrow}$	$0.075^{\nearrow}$	0.036^{ns}	$0.072^{\nearrow\nearrow\nearrow}$
0.100	$0.186^{\nearrow\nearrow\nearrow}$	$0.145^{\nearrow\nearrow}$	0.090^{ns}	0.065^{ns}	0.105^{ns}	$0.132^{\nearrow}$	0.092^{ns}	$0.098^{\nearrow\nearrow\nearrow}$
0.250	0.183^{ns}	0.239^{ns}	$0.190^{\searrow\searrow}$	0.188^{ns}	0.189^{ns}	0.200^{ns}	0.183^{ns}	0.200^{ns}
0.500	$0.360^{\searrow\searrow\searrow}$	$0.334^{\searrow\searrow}$	$0.329^{\searrow\searrow\searrow}$	$0.307^{\searrow\searrow\searrow}$	$0.354^{\searrow\searrow\searrow}$	$0.304^{\searrow\searrow}$	$0.385^{\searrow\searrow}$	$0.326^{\searrow\searrow\searrow}$
0.750	$0.472^{\searrow\searrow\searrow}$	$0.495^{\searrow\searrow\searrow}$	$0.456^{\searrow\searrow\searrow}$	$0.367^{\searrow\searrow\searrow}$	$0.472^{\searrow\searrow\searrow}$	$0.512^{\searrow\searrow\searrow}$	$0.452^{\searrow\searrow\searrow}$	$0.467^{\searrow\searrow\searrow}$
0.900	$0.754^{\searrow\searrow\searrow}$	$0.726^{\searrow\searrow\searrow}$	$0.625^{\searrow\searrow\searrow}$	$0.553^{\searrow\searrow\searrow}$	$0.663^{\searrow\searrow\searrow}$	$0.630^{\searrow\searrow\searrow}$	0.996^{ns}	$0.663^{\searrow\searrow\searrow}$
0.950	$0.628^{\searrow\searrow\searrow}$	$0.702^{\searrow\searrow\searrow}$	$0.736^{\searrow\searrow\searrow}$	$0.832^{\searrow\searrow\searrow}$	$0.787^{\searrow\searrow\searrow}$	$0.666^{\searrow\searrow\searrow}$	$0.997^{\searrow}$	$0.775^{\searrow\searrow\searrow}$
0.975	$0.843^{\searrow\searrow\searrow}$	$0.898^{\searrow\searrow\searrow}$	$0.849^{\searrow\searrow\searrow}$	$0.893^{\searrow\searrow\searrow}$	$0.932^{\searrow\searrow\searrow}$	$0.853^{\searrow\searrow\searrow}$	0.999^{ns}	$0.890^{\searrow\searrow\searrow}$

Note. Differences with the diagonal derive from MW tests for: $H_0 : W(E_{p_r}) = p_j$ vs. $H1 : W(E_{p_r}) \neq p_r$

"$\nearrow\nearrow\nearrow$ ($\searrow\searrow\searrow$)": significantly different from the diagonal at $P < 0.01$

"$\nearrow\nearrow$ ($\searrow\searrow$)": significantly different from the diagonal at $P < 0.05$

"$\nearrow$ ($\searrow$)": significantly different from the diagonal at $P < 0.10$

"ns": not significantly different from the diagonal at $P < 0.10$

$^+$Containing cases where objective and observed relative frequencies do not (always) coincide.

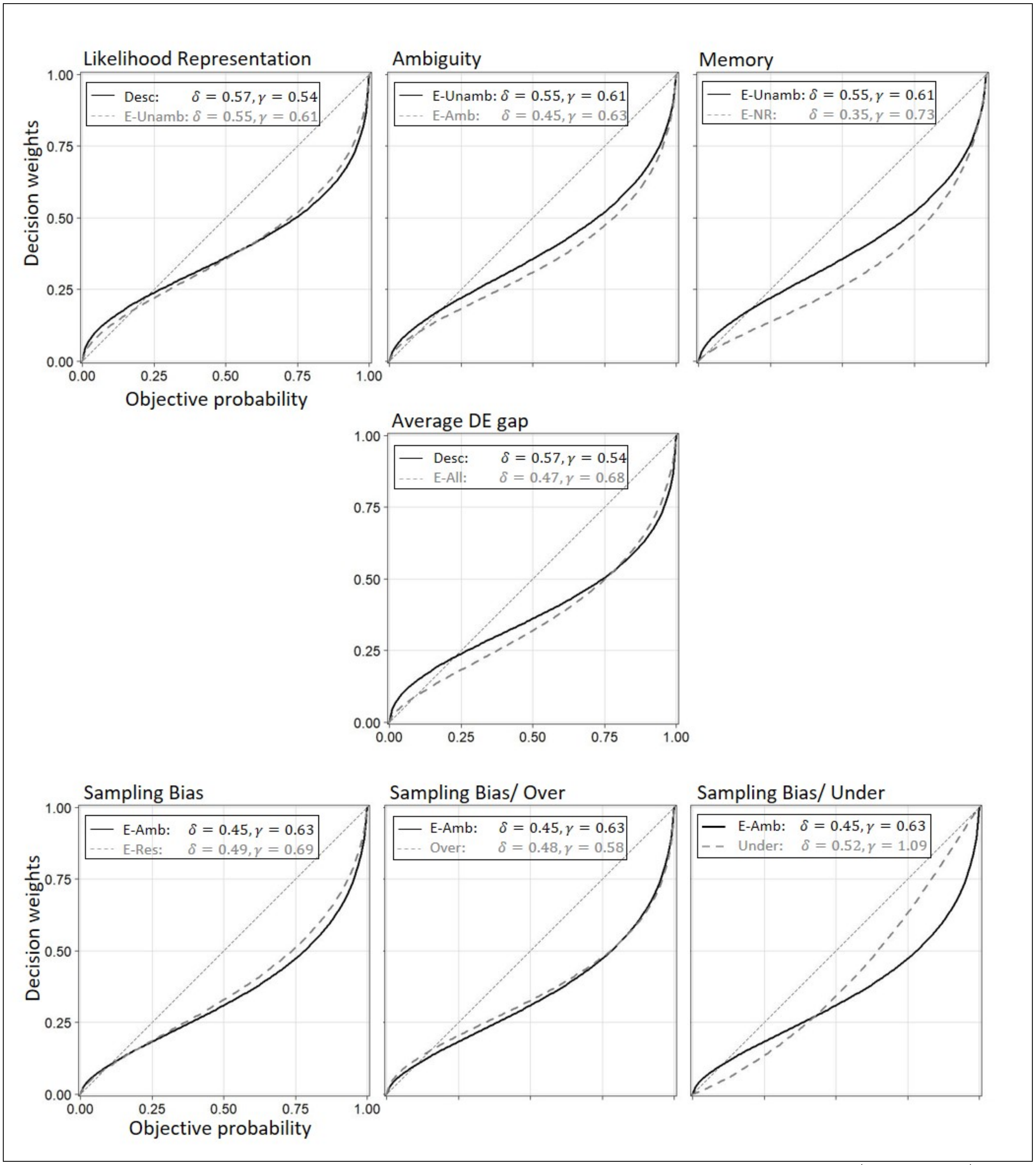
We complement the analysis of Table 3 by providing a visualisation of the weighting functions implied by the weights reported there by fitting a parametric weighting function to the set of median decision weights for each treatment. We estimate parameters δ and γ for each subject, using the linear-in-log-odds specification of the weighting curve (Goldstein and Einhorn, 1987; Gonzalez and Wu, 1999). The functional form of this weighting curve is:

$$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}}$$

The parameter δ is largely responsible for the elevation of the curve and γ for its curvature (Gonzalez and Wu, 1999). When $\gamma < 1$, the weighting function takes its characteristic inverse S-shaped that suggests overweighting of rare events. As γ approaches the value 1, the weighting curve becomes increasingly linear. Finally, values where $\gamma > 1$ suggest an S-shaped curve that is consistent with underweighting of rare events.

The fitted functions are presented in Figure 4 where each of the seven panels provides a comparison of a pair of functions, thereby giving a qualitative impression of the impact of an individual factor manipulated in our design. We also include a comparison of Desc versus E-all for completeness. The top three panels use data from treatments which control for sampling bias and show the impacts of, respectively, likelihood representation; ambiguity and memory. The top left panel reveals that the treatments Desc and E-Unamb generate almost identical inverse-S functions; hence our treatment manipulation capturing the impact of likelihood representation has no discernible impact on the fitted function. The middle and rightmost panels of the top row both use E-Unamb as a benchmark: the introduction of ambiguity (middle panel) slightly depresses the revealed weighting function throughout much of its range. The impact of removing the memory aid (history table) in our design, depresses weights more markedly (top right panel).

Figure 4: Parametric decision weight functions



Note. Parametric decision-weighting curves fitted at the level of median decision weights (see Table 3) for each treatment.

Top row: Comparisons without sampling bias. Bottom row: Comparisons with sampling bias. Centre: Desc vs. E-All. Legends report the parameters of the weighting function that was used to fit the curves.

The bottom three panels provide a similar exercise but focused on the impact of sampling bias. Here we highlight, in particular, the impact of under-representation of rare events: in line with the discussion of Table 3, relative to the E-Amb treatment, we see the under-representation of rare events reducing the weights associated with both desirable and undesirable rare events (i.e. the weighting function for E-Under lies below the E-Amb function for low probabilities and above it for high ones). Finally, the comparison in the central panel, capturing the visual effect of the average DE gap, provides support for the 'relative underweighting hypothesis' (Kopsacheilis, 2018). Although both weighting functions are inverse S-shaped, that of E-All exhibits less overweighting – i.e, it is closer to the diagonal - when compared to Desc for small and high probabilities.

An obvious question is how far the suggestion of treatment differences is supported by more formal statistical analysis. We address this via Table 4 which reports the P-values from a series of 2-tailed Mann–Whitney (MW) U tests. The set of tests mirrors the structure of the treatment-level comparisons presented in Figure 3: we compare the same pairs of treatments testing for gaps and effects, but we use the RDEU approach to conduct a series of tests at each probability level for every treatment comparison.

We highlight three main observations based on this analysis. First, considering the bottom half of the table which tests the impact of individual factors operating in isolation, we confirm the finding of our model-free analysis that sampling bias, in the form of under-representation of rare events, has a major impact on the revealed weights. Second, and again in line with the model-free analysis, the bottom half of the table shows no evidence (at 5% significance or better) to support the impact of any factor beyond sampling bias.

Third, the RDEU analysis presented above offers some new insights too. Based on the results presented in the top half of Table 4, we now can detect a significant DE gap in some cases where there was no sampling bias. Specifically, we find evidence of a significant gap in the comparison of Desc vs E-NR for small values of p (desirable rare events). This confirms the visual impression (from top right panel of Figure 4) that desirable rare events receive lower weights in treatment E-NR relative to E-Unamb and Desc (the latter two functions being almost identical). We take these results as indicating that there may be a replicable effect worthy of further investigation. On that assumption, it is helpful to reflect on the differences in weighting between Desc and E-NR.

Referring back to Figure 1, note that we get from Desc to E-NR in two steps: one changes the likelihood representation; the other removes the history table. The evidence presented in Figure 4 and Table 4 provides tentative support for thinking that the removal of the history table may be the more important of the two manipulations: memory has a more marked effect on the shape of the weighting functions (comparing top left and top right panels of Figure 4) and in the bottom of Table 4, the memory effect in isolation does reach significance at 5% at one probability level ($p = 0.1$). As noted earlier, removal of the history table may be interpreted as not purely cognitive if subjects are aware of their forgetfulness and react to the resulting ambiguity. Hence, to the extent that removing the history table has a genuinely distinct effect, we are not entitled to interpret it as a purely cognitive one, as there may be some preference component too. This leads to Result 4.

Result 4 *We find evidence that there are factors other than sampling bias that contribute to the DE gap. These factors are most clearly seen when our memory aide is removed and, thus, involve cognitive factors and responses to them.*

Table 4: Statistical significance (P-values of MW tests)

DE gaps							
p	**Desc vs E-All**	**Desc vs E-Unamb**	**Desc vs E-Amb**	**Desc vs E-NR**	**Desc vs E-Res**	**Desc vs Over**	**Desc vs Under**
0.025	0.037**	0.120	0.302	0.009***	0.165	0.818	0.039**
0.050	0.076*	0.467	0.280	0.053*	0.062*	0.283	0.037**
0.100	0.279	0.849	0.307	0.060*	0.486	0.907	0.186
0.250	0.987	0.449	0.561	0.707	0.796	0.528	0.927
0.500	0.967	0.574	0.879	0.542	0.735	0.938	0.592
0.750	0.602	0.949	0.834	0.135	0.669	1.000	0.993
0.900	0.733	0.864	0.678	0.371	0.518	0.680	0.024**
0.950	0.112	0.432	0.574	0.148	0.025**	0.452	0.000***
0.975	0.330	0.321	0.664	0.420	0.085*	0.887	0.003***

Effects						
p	**Likelihood Repr.**	**Memory**	**Ambiguity**	**Sampling Bias**	**Over-representation**	**Under-representation**
0.025	0.120	0.193	0.785	0.743	0.292	0.099*
0.050	0.467	0.221	0.829	0.280	0.728	0.121
0.100	0.849	0.044*	0.355	0.856	0.571	0.584
0.250	0.449	0.302	0.148	0.432	0.334	0.694
0.500	0.574	0.293	0.642	0.581	0.928	0.353
0.750	0.949	0.156	0.719	0.751	0.705	0.894
0.900	0.864	0.385	0.568	0.356	0.903	0.018**
0.950	0.432	0.380	0.996	0.104	0.965	0.000***
0.975	0.321	0.942	0.109	0.300**	0.534	0.000***

Note. Column 1 reports the list of objective probabilities for which decision weights were calculated. In the remaining columns we report P-values for the MW tests.
$^{***}P < 0.01$; $^{**}P < 0.05$; $^{*}P < 0.1$.

5 Conclusion

We have reported the results of a lab experiment investigating the Description - Experience (DE) gap, an empirical phenomenon pointing to a marked sensitivity of risk attitudes according to whether they are elicited from description or from experience.

According to the most popular interpretation of the canonical finding in this literature, when uncertainty is communicated through descriptions, people behave as if overweighting rare events relative to their probability while, conversely, when this uncertainty is experienced they tend to make decisions consistent with underweighting rare events.

We taxonomized the key factors that might drive this empirical discrepancy into three broad categories by distinguishing between factors pertaining to: informational (sampling bias), preferential (ambiguity) or cognitive (likelihood representation and memory) aspects of decision making. Then, we implemented a novel 5-treatment design comprising one standard version of description and 4 variations of experience. Our treatment protocol was designed to isolate these factors through a series of pairwise comparisons. At the same time, this design allowed us to elicit a series of DE gap variants; one for each comparison of description with a variant of experience. Moreover, to account for different measuring traditions, our design allows us to employ two measuring approaches.

First, we study the gap without relying on any behavioural model by focusing on choice proportions from paired gambles. This allowed us to compare our findings with those in earlier literature, where the DE gap was first established. We find that despite our departing quite markedly from the 'sampling paradigm' - arguably the best-known experimental framework for studying the phenomenon - our average elicited DE gap coincides in direction and size with the literature's average (Result 1) suggesting that the phenomenon is robust. Moreover, we find that each causal factor that we isolate contributes positively to the phenomenon and therefore, comparisons entailing more than one factor, induce bigger effects. Among those factors, the most important and the only statistically significant one in isolation, is sampling bias due to under-representation of rare events (Result 2).

Second, assuming a rank dependent expected utility model, we compare decision-weighting functions across treatments. These are elicited semi-parametrically and at the individual level with the use of the bisection variant of the certainty equivalents method. This level of analysis allowed us to examine probability weighting while controlling for utility curvature, as well as to explore a variety of probability regions separately. We find the two levels of analysis to be complimentary. The model-mediated approach replicates the findings of the model-free analysis. However, the use of weighting functions and their shape allows us to shed some more light on aspects of behaviour that would otherwise be inaccessible.

One such aspect derives from observing the shape of weighting functions. There, we find that when we control for sampling bias, probability weighting is generally inverse S-shaped, consistent with overweighting of rare events in both description and experience treatments (Result 3). This adds to the evidence that overweighting of rare events can be found beyond the narrow confines of described risk. Indeed, our average DE gap is best summarized by a 'relative underweighting hypothesis' (Kopsacheilis, 2018), whereby rare events are overweighted in experience, only less so than in description.

Taking stock of our findings, we confirm the existence of a DE gap, one whereby people appear to be overweighting rare events significantly less in experience than in description. When we consider factors in isolation, sampling bias due to the under-representation of rare events was the only significant driver of this gap. Does this mean that the DE gap can be simply reduced to an information asymmetry between description and experience? If that were the case, the implications for the theory of risk preferences would be limited. The core of risk preference models would not require any revisions, though it might have some implications for how we apply such models. We would need to consider for example whether a particular application is one where sampling bias might arise. Predicted behaviour would then depend on both the theory of preference and the extent of such information biases.

Notwithstanding the convenience of this interpretation, we do not endorse it undeservedly. There are two reasons we suggest that searching for the gap beyond information asymmetries is worthwhile, even if -at least in our study- this pursuit required that the researcher be equipped with a magnifying glass - for the gap was not as big - and a compass - for it was not ubiquitous.

First, although we maintain that our treatment design was apt in capturing the essence of the underlying factors, it is still possible that some of these factors might be more prominent in different environments. Second, even though when considered in isolation, factors that did not involve sampling bias left only faint traces, when more than one of those factors were considered in tandem, their effect was significantly bigger - especially when such mixtures involved memory limitations.

An interpretation that combines elements of sampling bias and memory, inspired by the similarities in typical weighting functions for subjects with no memory aides and for subjects with samples that under-represent rare evens is that the same statistical property of the binomial distribution is at work in both cases. If so, this could reflect the evolutionary idea of synergistic biases leading to second-best adaptations (see Waldman, 1994, for an example of how aversion to effort and overestimation of one's own abilities can act synergistically). Specifically, the behavioural bias to overweight rare events (as captured by several non-expected utility models such as Prospect Theory) may be countering the statistical (and/or cognitive) bias to under-represent them in small - collected or recollected - samples. As a result, the corresponding weighting curves can be indistinguishable from the diagonal - as it is often the case in our data - and therefore from the normative[18] expected utility benchmark.

[18]See Wakker (2010) for a discussion about the normative character of expected utility theory.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, 46(11):1497–1512.

Abdellaoui, M., Baillon, A., Placido, L., and Wakker, P. P. (2011a). The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation. *American Economic Review*, 101(2):695–723.

Abdellaoui, M., L'Haridon, O., and Paraschiv, C. (2011b). Experienced vs. described uncertainty: Do we need two prospect theory specifications? *Management Science*, 57(10):1879–1895.

Aydogan, I. and Gao, Y. (2019). Experience and rationality under risk: Re-examining the impact of sampling experience. *Manuscript in preparation*.

Bardsley, N., Cubitt, R., Loomes, G., Moffat, P., Starmer, C., and Sugden, R. (2010). *Experimental economics: Rethinking the rules*. Princeton University Press.

Barron, G. and Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, 16(3):215–233.

Barron, G. and Ursino, G. (2013). Underweighting rare events in experience based decisions: Beyond sample error. *Journal of Economic Psychology*, 39:278–286.

Bates, D., Chambers, J., and Hastie, T. (1992). Statistical models in s. In *Comp. Sci. and Stat., Proc. 19th Symp. on the Interface*. Wadsworth & Brooks California.

Bates, D. M. and Watts, D. G. (1988). *Nonlinear regression analysis and its applications*, volume 2. Wiley New York.

Booij, A. S., Van Praag, B. M., and Van De Kuilen, G. (2010). A parametric analysis of prospect theory's functionals for the general population. *Theory and Decision*, 68(1-2):115–148.

Cubitt, R. P., Starmer, C., and Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics*, 1(2):115–131.

Etchart-Vincent, N. (2004). Is Probability Weighting Sensitive to the Magnitude of Consequences? an Experimental Investigation on Losses. *Journal of Risk and Uncertainty*, 28(3):217–235.

Etner, J., Jeleva, M., and Tallon, J.-M. (2012). Decision theory under ambiguity. *Journal of Economic Surveys*, 26(2):234–270.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fox, C. R. and Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, 1(2):159–161.

Ghirardato, P. and Marinacci, M. (2001). Risk, ambiguity, and the separation of utility and beliefs. *Mathematics of Operations Research*, 26(4):864–890.

Gigerenzer, G. and Hoffrage, U. (1995). How to Improve bayesian Reasoning Without Instruction: Frequency Formats. *Psychological Review*, 102(4):684–704.

Glöckner, A., Hilbig, B. E., Henninger, F., and Fiedler, S. (2016). The reversed description-experience gap: Disentangling sources of presentation format effects in risky choice. *Journal of Experimental Psychology: General*, 145(4):486–508.

Goldstein, W. M. and Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, 94(2):236–254.

Gonzalez, R. and Wu, G. (1999). On the Shape of the Probability Weighting Function. *Cognitive Psychology*, 38(1):129–166.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.

Hau, R., Pleskac, T. J., and Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making*, 23(1):48–68.

Hau, R., Pleskac, T. J., Kiefer, J., and Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, 21(5):493–518.

Hertwig, R., Barron, G., Weber, E. U., and Erev, I. (2004). Decisions from Experience and the Effect of Rare Events in Risky Choice. *Psychological Science*, 15(8):534–539.

Hills, T. T. and Hertwig, R. (2010). Information Search in Decisions From Experience: Do Our Patterns of Sampling Foreshadow Our Decisions? *Psychological Science*, 21(12):1787–1792.

Knight, F. H. (1921). Cost of production and price over long and short periods. *Journal of Political Economy*, 29(4):304–335.

Kopsacheilis, O. (2018). The role of information search and its influence on risk preferences. *Theory and Decision*, 84(3):311–339.

Luce, R. D. (1991). Rank- and Sign-Dependent Linear Utility Models for Binary Gambles. *Journal of Economic Theory*, 53(1):75–100.

Miyamoto, J. M. (1988). Generic utility theory: Measurement foundations and applications in multiattribute utility theory. *Journal of Mathematical Psychology*, 32(4):357–404.

Moré, J. J. (1978). The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer.

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

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

Rakow, T., Demes, K. A., and Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes*, 106(2):168–179.

Slovic, P., Monahan, J., and MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and human behavior*, 24(3):271–296.

Trautmann, S. T. and Van De Kuilen, G. (2015). Ambiguity attitudes. *The Wiley Blackwell handbook of judgment and decision making*, 1:89–116.

Tversky, A. and Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, 102(2):269–283.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Ungemach, C., Chater, N., and Stewart, N. (2009). Are Probabilities Overweighted or Underweighted When Rare Outcomes Are Experienced (Rarely)? *Psychological Science*, 20(4):473–479.

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity.* Cambridge University Press.

Waldman, M. (1994). Systematic errors and the theory of natural selection. *American Economic Review*, 84(3):482–497.

Weber, E. U., Shafir, S., and Blais, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation. *Psychological Review*, 111(2):430–445.

Wulff, D. U., Mergenthaler-Canseco, M., and Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, 144(2):140–176.

6 Appendix

6.1 Instructions

Instructions were handed to participants in printed form and were read out loud by the experimenter prior to the start of the experiment. Before the start of the experiment and after the instructions had been read out loud, subjects played one trial round.

6.1.1 Instructions for Description

In this study you are asked to make choices that involve lotteries. For each choice, just pick the option you prefer as there are no 'right' or 'wrong' answers. Overall you are going to consider a total of 19 lotteries which are described by virtual decks of cards. Each deck contains exactly two types of cards represented by two different colours. Each deck has its own mix of these two types of cards.

The information about the relative frequency and the monetary value of each type of card will be provided to you (in the form of percentages) prior to making a choice. This information is seen on the bottom of the screen.

The first 7 lotteries are all associated with the same deck of cards. This guarantees that the relative frequency of each colour is the same for Lotteries 1 to 7. Notice however that the rewards associated with each outcome will differ from one lottery to another.

Later in the experiment, you may have the opportunity to 'play' a lottery. That would mean drawing once more from a deck you have sampled and receiving the sum of money assigned to the colour of the drawn card.

Your task is to choose each time between playing the Lottery and receiving the Certain Outcome. Each Lottery entails 5 such choices between the Lottery (Option A) which remains constant across these 5 Choice-Rounds and a Certain Outcome (Option B) that will be changing from each choice to the next.

Payoff Stage

At the end of the experiment one choice is going to be randomly selected to be played out for real. All choices are equally likely to be drawn so each choice you make has equal chances of affecting your final payment. There are two cases:

Case 1: If in the randomly selected choice you chose Option B (the Certain Outcome) then the monetary value of this choice is going to be added directly to your final payment.

Case 2: If in the randomly selected choice you chose Option A (the Lottery) then the deck of cards corresponding to that choice will reappear on the screen. You will then be asked to draw one card from it. Then the monetary value assigned to the colour of the card you just drew will be added to your final payment.

6.1.2 Instructions for the four versions of Experience

In this study you are asked to make choices that involve lotteries. For each choice, just pick the option you prefer as there are no 'right' or 'wrong' answers. Overall you are going to consider a total of 19 lotteries which are described by virtual decks of cards. Each deck contains **exactly two** types of cards, represented by two different colours. Each deck has its own mix of these two types of cards.

For every lottery you go through two stages:

Stage 1: the 'Sampling Stage'

Stage 2: the 'Choice Stage'

Exception: The first 7 Lotteries all share the same 'Sampling Stage' because they relate to the same deck. This means that you will only sample once for the first seven lotteries. Each of the lotteries 8 - 24 has its own Sampling Stage (because it relates to its own deck).

Stage 1: Sampling Stage

In each Sampling Stage you go through a particular computerized deck and explore one by one all of their cards. The information about the relative frequency of each type of card is unknown to you prior to the start of the sampling process. *However by the end of the process, this information will be completely revealed to you as you will have seen every card in the deck exactly once.*[19] As mentioned earlier, the first 7 lotteries relate to the same deck. This guarantees that the relative frequency of each colour is the same for Lotteries 1 to 7. We recommend that you pay attention during this sampling process as this information is relevant for your decisions later on and hence your final payment.

Every time you click on the '*Draw*' button you will observe a new card from the deck.

[19] The italicised text was present on in E-Unamb and E-NR, where there was no ambiguity regarding the representativeness of the sampled cards. In E-Amb and E-Res this message was replaced by the following: '*However by the end of this process you will have discovered something more about this mix because you will have seen a selection of draws from that deck*'.

Once you observe its colour click on '*Proceed with the next card*' for the '*Draw*' button to reappear. You will repeat this process until you go exactly once through all the cards in each deck. Once you have done so, a message will appear on the screen verifying that you have seen all the cards in this deck and a button that reads: '*Go to the Choice Stage*' will become accessible at the bottom of the screen. Once you click on that button you will move on to the 'Choice Stage'.

Stage 2: 'Choice Stage'

At this stage a monetary value is assigned to the colour of each card. This information is seen on the bottom of the screen. Later in the experiment, you may have the opportunity to 'play' a lottery. That would mean drawing once more from a deck you have sampled and receiving the sum of money assigned to the colour of the drawn card.

On the top of the screen you will observe a 'History Table' where you can track your sampling history from each lottery's 'Sampling Stage'. As mentioned earlier, the first 7 lotteries are all associated with the same deck of cards and hence share the same 'History Table'.[20] Notice however that although the relative frequency of each colour of card is the same for lotteries 1 to 7, the rewards associated with each outcome will differ from one lottery to another.

Your task in this stage is to choose each time between playing the Lottery and receiving the Certain Outcome. Each Lottery entails 5 such choices between the Lottery (Option A) which remains constant across these 5 Choice-Rounds and a Certain Outcome (Option B) that will be changing from each choice to the next.

Payoff Stage

At the end of the experiment one choice is going to be randomly selected to be played

[20]References to this 'History Table' did not feature in E-NR where there was no such visual aid.

out for real. All choices are equally likely to be drawn so each choice you make has equal chances of affecting your final payment. There are two cases:

Case 1: If in the randomly selected choice you chose Option B (the Certain Outcome) then the monetary value of this choice is going to be added directly to your final payment.

Case 2: If in the randomly selected choice you chose Option A (the Lottery) then the deck of cards corresponding to that choice will reappear on the screen. You will then be asked to draw one card from it. Then the monetary value assigned to the colour of the card you just drew will be added to your final payment.

6.2 Bisection method

The following table demonstrates an example of the bisection process for the lottery: $(16, 0.1; 0)$. Choices are represented in bold. The elicited CE for the process of this example will be the mid-point between 0.7 (the last certain outcome preferred over the lottery) and 0.6 (the last certain outcome that the lottery was preferred over). This yields a CE equal to 0.65.

Table 5: Illustration of the bisection method for (4, 0.8;0)

Iterations	CE elicitation questions
1	(16, 0.1;0) vs. **1.6**
2	(16, 0.1;0) vs. **0.8**
3	**(16, 0.1;0)** vs. 0.4
4	**(16, 0.1;0)** vs. 0.6
5	(16, 0.1;0) vs. **0.7**

6.3 Utility curvature

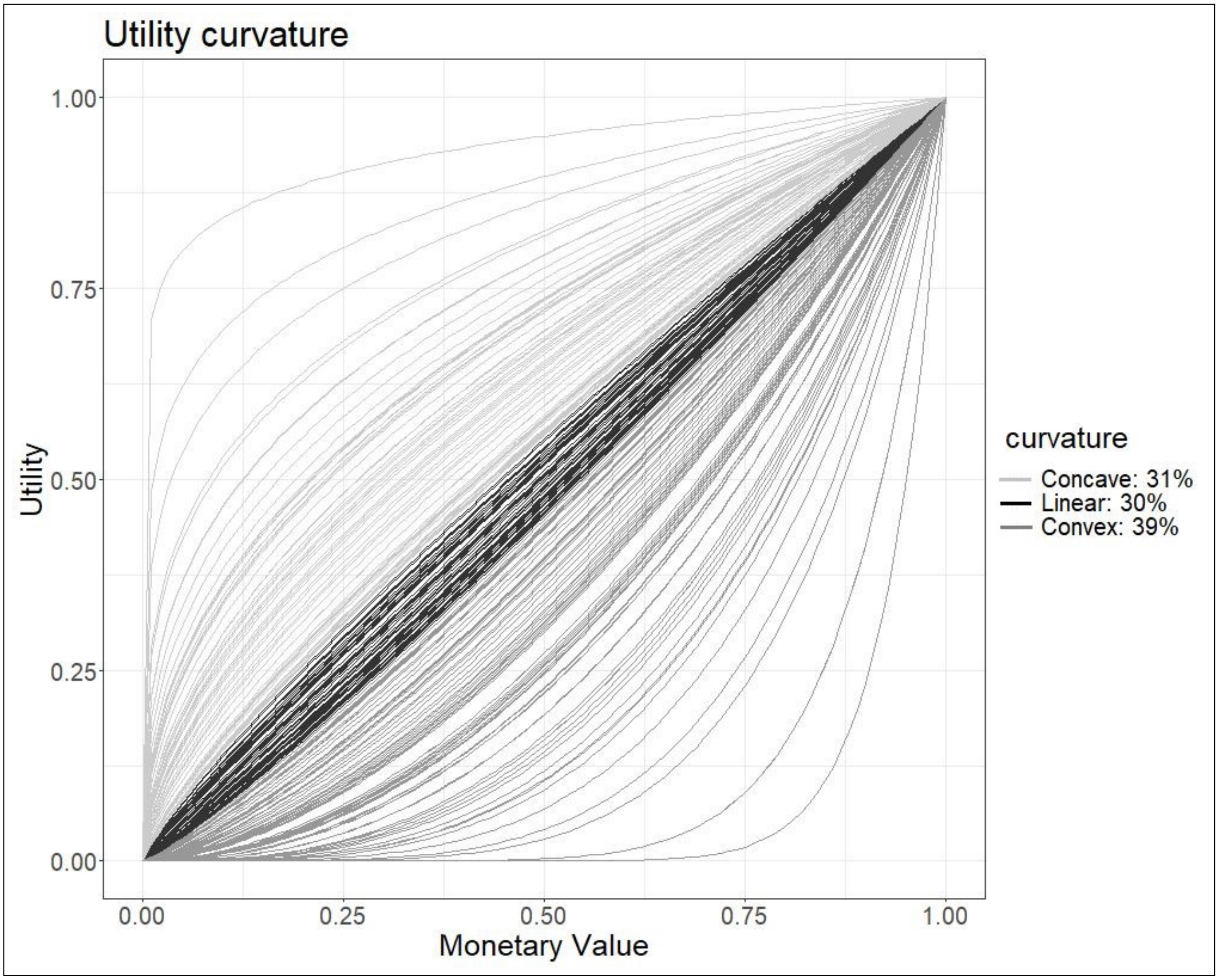
Table 6: P-values from bilateral, MW tests for differences in utility curvature between Description and Experience treatments

Desc vs E-Unamb	Desc vs E-NR	Desc vs E-Amb	Desc vs E-Res	Desc vs Over	Desc vs E-Und Under	Desc vs E-All E-All
0.996	0.811	0.887	0.277	0.277	0.275	0.827

Table 7: P-values from bilateral, MW tests for differences in utility curvature between treatments that isolate effects

Lik. Repr. Desc vs E-Unamb	**Memory** E-Unamb vs E-NR	**Ambiguity** E-Unamb vs E-Amb	**Samp. Bias** E-Amb vs E-Res	**SB-Over** E-Amb vs Over	**SB-Under** E-Amb Under
0.996	0.852	0.950	0.250	0.250	0.249

Figure 5: Utility curves



Note. Plotted utility curves for all 198 subjects across treatments. The estimates derive from the power utility function: $u(x) = x^{\alpha}$. Convex curvature: those with $\alpha > 1.15$, linear curvature: those with $0.85 \leq \alpha \leq 1.15$ and concave curvature: those with $\alpha < 0.85$. Types were approximately uniformly distributed across these three categories (approx. 1/3 in each).

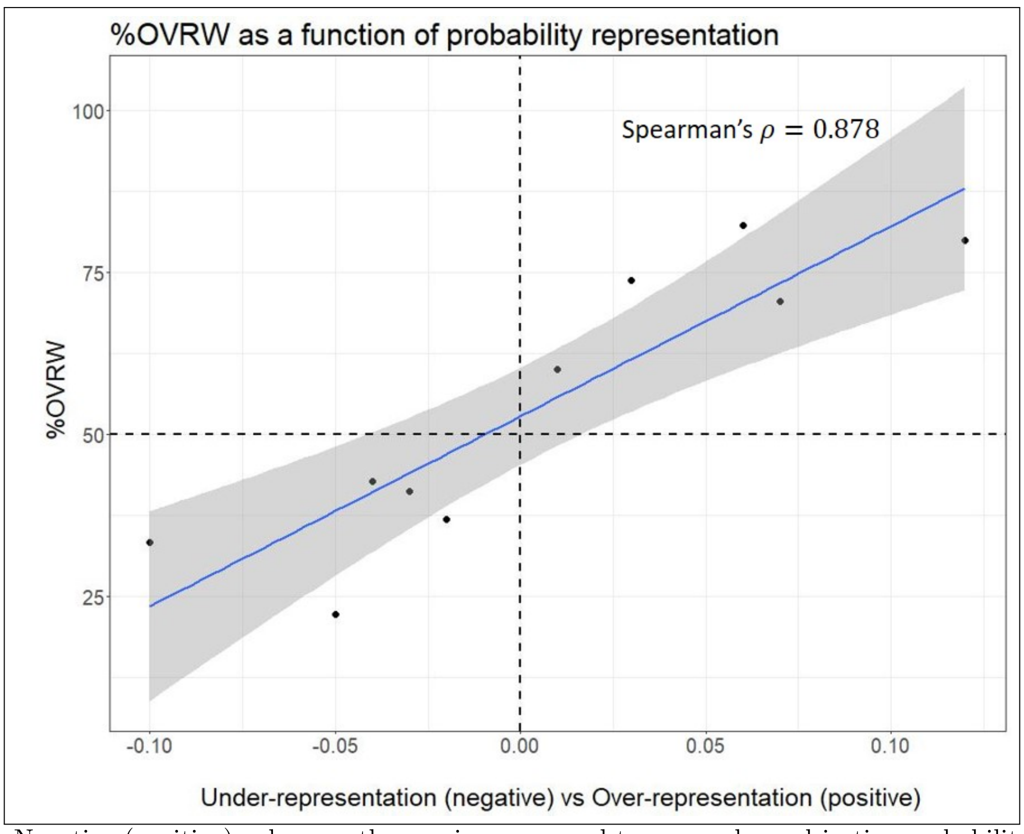
6.4 Sampling Bias

Table 8: Objective and experienced probabilities in E-Res

p	f_p	se
0.025	0.028	0.005
0.050	0.038	0.006
0.100	0.105	0.008
0.250	0.238	0.012
0.500	0.487	0.012
0.750	0.754	0.012
0.900	0.882	0.010
0.950	0.942	0.007
0.975	0.969	0.005

Note.'p': Objective probability. 'f_p': experienced probability. 'se': standard error of f_p.

Figure 6: %OVRW as a function of sampling bias



Note. Negative (positive) values on the x-axis correspond to cases where objective probability was under-represented (over-represented). The fitted line is calculated from a linear model and the shaded area represents the standard errors. Spearman's $\rho = 0.878$, which is significant (P=0.001).