Eliciting Multiple Prior Beliefs

Mohammed Abdellaoui* Philippe Colo† Brian Hill‡

30th October 2024

Abstract

Despite the increasing importance of multiple priors in various domains of economics, choice-based incentive-compatible multiple-prior elicitation remains an open problem. This paper develops a solution, comprising a preference-based identification of a subject's probability interval for an event, and a method for eliciting it. The method applies under weak decision-theoretic assumptions, with no need for probabilistic sophistication. To demonstrate its feasibility, we implement it in three incentivized experiments on artificial and natural sources of uncertainty. Intervals elicited by our method are sensitive to the direction and amount of information, and are typically consistent with 'objective' probabilities where available. We find a predominance of non-degenerate probability intervals, with intervals being wider when there is less information or predictability. The probability intervals elicited with our method are similar to those stated by subjects on aggregate, suggesting that the method can provide behavioral foundations for the use of stated probability-interval techniques in the field.

Key words: Multiple Priors, Probability Intervals, Belief Measurement, α-maxmin EU, Imprecise Probability.

JEL Codes: D9, D81

*HEC Paris & CNRS, 1 rue de la Libération, 78351 Jouy-en-Josas, France. E-mail: abdellaoui@hec.fr.

†ETH Zurich & Universität Bern, Länggassstrasse 49a, 3012 Bern, Switzerland. E-mail: colo.philippe@gmail.com.

‡HEC Paris & CNRS, 1 rue de la Libération, 78351 Jouy-en-Josas, France. E-mail: hill@hec.fr.

1. Introduction

The standard Bayesian model of decision under uncertainty stipulates that a decision maker's beliefs are fully captured by a single probability measure (Savage, 1954; Anscombe and Aumann, 1963). Empirical applications often call for elicitation of subjective beliefs (Manski, 2004), and a wide array of probability elicitation methods have been proposed including, beyond *stated* probabilities, scoring rules and matching-probability based approaches. Importantly, the latter are choice based and incentive compatible, and hence can be used to evaluate simpler methods and ground their use in the field. For instance, studies showing that stated probability elicitation methods often lead to limited performance loss compared to choice-based approaches provide a principled foundation for their use in large-scale field studies (Trautmann and Kuilen, 2015).

Elicitation of subjective probabilities plays a significant role in areas such as macroeconomics—with interest in beliefs concerning future demand or inflation (Guiso and Parigi, 1999; Engelberg et al., 2011)—as well as development and agricultural economics—where important factors include agents' beliefs about future weather, market factors or outcomes of crop, technological or entrepreneurial choices (Delavande et al., 2011; Cerroni, 2020). Such future events often involve significant uncertainties, especially in times of crisis, change or innovation. Uncertainties of this scale, and behavioural evidence concerning them, have motivated the development of multiple prior decision models (Gilboa and Schmeidler, 1989; Ghirardato et al., 2004), which replace the Bayesian single-prior representation of beliefs by a set of priors, generating a probability interval for each event. A rich theoretical literature has documented characteristic differences in insurance and investment decisions taken by multiple prior agents as compared to Bayesian ones (Dow and da Costa Werlang, 1992), with qualitatively distinct consequences in macroeconomics (Ilut and Schneider, 2014), asset pricing (Garlappi et al., 2007; Epstein and Schneider, 2010), mechanism design (Bose and Renou, 2014), health economics (Giustinelli et al., 2022) and climate economics (Hill, 2024). However, despite this evidence that multiple-prior-generated imprecision is a potential driver of various economic phenomena—and indeed, despite its use for communication by several institutions, e.g. the Intergovernmental Panel on Climate Change and central banks (Mastrandrea et al., 2010; Carney et al., 2019)—probability elicitation methods rule it out. They thus cannot allow us to properly ascertain its role and leverage its potential. Elicitation of multiple prior beliefs is needed.

The situation concerning multiple prior *elicitation* is markedly different from that for Bayesian probability, with almost all attempts to date focusing on subjects' *stated* probability intervals (Giustinelli et al., 2022; Kriegler et al., 2009). In particular, the absence of *bona fide* theoretically well-founded, choice-based and incentive-compatible multiple-prior elicitation methods deprives stated intervals of a firm grounding, hence raising potential questions about the findings based on them. This paper proposes such an elicitation method for multiple-prior probability intervals, implements it in a series of laboratory experiments and compares it to stated intervals.

An impossibility result from the statistics literature (Seidenfeld et al., 2012, Prop 5) provides a flavour of the challenge posed by incentive-compatible elicitation of multiple priors: there exist no real-valued continuous strictly proper scoring rule for multiple-prior probability intervals. Related issues affect the matching probability method (Borel, 1939; Anscombe and Aumann, 1963). It elicits the matching probability (MP)—that is, the proportion of red balls in an unambiguous red-and-blue-balled urn at which the subject is indifferent between betting on red from the urn and betting on a target event E. Under Subjective Expected Utility (SEU), the MP of E coincides with the subject's probability of it. For multiple-prior preferences, however, this is no longer the case. For instance, under the popular (Hurwicz) α-maxmin EU model, the MP reflects the bounds of the subject's probability interval for the event, but also her *attitude to uncertainty* or *ambiguity*. Indeed, even eliciting the MPs of E and its complement E^c (which, beyond SEU, need not add to one) does not allow identification of the subject's probability interval in general, due to the confounding ambiguity attitude factor.[1] Existing theoretical and experimental approaches to this well-known issue (e.g. Ghirardato et al. 2004; Eichberger et al. 2011; Section 5) assume that the subject's set of priors is generated by precise probabilistic beliefs, i.e. preferences are probabilistically sophisticated (Chateauneuf et al. 2007; Baillon et al. 2018b, 2021; Gul and Pesendorfer 2015; Section 5). However, such assumptions are least warranted in situations where multiple priors are most relevant—and hence undermine the suitability of such *precision-laden* methods for multiple-prior probability-interval elicitation. Indeed, to meet the challenge of multiple-prior elicitation, an incentive-compatible, fully general and hence *precision-free* method is required.

[1]For instance, under the Hurwicz α-maxmin EU model, the MP of an event E, $MP(E)$, satisfies:

$$MP(E) = \alpha \underline{p}(E) + (1-\alpha)\overline{p}(E)$$

where the subject's probability interval for E is $[\underline{p}(E), \overline{p}(E)]$ and α is typically interpreted as a reflection of the subject's ambiguity attitude (Section 2.3). Since the probability interval for the complement event E^c is $[1-\overline{p}(E), 1-\underline{p}(E)]$, eliciting the MPs for the event and its complement yields two equations in three unknowns—and hence does not allow identification of the subject's probability interval. See Section 2.4.

Drawing on theoretical results that provide a solution to the identification problem for α-maxmin EU and a wide range of generalisations (Hill, 2023), we develop an MP-like elicitation method that uses extraneous random devices with *interval-valued* rather than precise probabilities. To illustrate, consider an urn containing only red and blue balls, where all that is known is that at least proportion r of the balls in the urn are red, and at least proportion b are blue (with $r+b \leq 1$). Here, the probabilities of getting red or blue on the next draw from the urn are summarized by the intervals $[r,1-b]$ and $[b,1-r]$, respectively. To identify the bounds of the subject's probability interval for E, it suffices to find such an urn where the subject is indifferent between betting on E and betting on red from the urn, and between betting against E and betting against red (i.e. on blue). As we show in Section 2.4, under α-maxmin EU and a range of generalisations, the subject's probability interval is given by the interval $[r,1-b]$ corresponding to this urn. Moreover, this identification holds independently of the subject's ambiguity attitudes.[2] This thus yields a choice-based association of an 'interval-valued' urn to each event, which identifies the subject's probability interval for it. This *matching probability interval* (MPI) notion resolves the problem of choice-based incentive-compatible probability-interval elicitation in theory.

Our approach resolves the aforementioned foundational challenges. First of all, it is theoretically robust, insofar as it operates under Hurwicz α-maxmin expected utility as well as an array of generalisations—and hence without assumptions on subjects' ambiguity attitudes. Moreover, it is precision free, requiring no assumption of precise probabilities underpinning subjects' probability intervals. As discussed in Section 5, beyond distinguishing our approach from those mentioned above, this also differentiates it from scoring rules for most-likely intervals for the value of an unknown parameter (Winkler and Murphy, 1979; Schlag, 2015).

To operationalize elicitation of matching probability intervals, we develop an MPI version of the two-step MP elicitation method adopted by Abdellaoui et al. (2021, 2023). Under their method, a subject undergoes a 'bisection' binary-choice procedure followed by a 'confirmation' choice list; we develop an analogue binary choice procedure and 'two-dimensional' choice list, tailored for eliciting (two-dimensional) probability intervals instead of (one-dimensional) probability values. We implement our method in three laboratory studies. EXP A involves an artificial source of uncertainty—the colour of the next chip drawn from a bag—where prior information was provided through sampling. This controlled environment allows validation testing of the method, via the observed relationship between the elicited intervals and the exogenous

[2] Technically, under α-maxmin EU, these indifferences yield a pair of equations where the ambiguity attitude factor α cancels out, hence leading to a unique solution for the subject's probability interval; see Section 2.4.

information. Moreover, by eliciting stated probability intervals as well, it permits a comparison of the two elicitation approaches. EXP N1 and EXP N2 focus on natural sources of uncertainty, based on continuous variables. There, the method is used to elicit the interval-valued cumulative distribution functions (CDFs) generated by subjects' multiple priors.[3] Interval-valued CDFs are commonly used in applications to go beyond the assumption of precise subjective probabilities (Karanki et al., 2009); our elicitation of CDFs provides a test of our approach, showing that it can operate in such contexts.

Our method passes the validation tests in EXP A, providing intervals that are sensitive to both the direction (e.g. sample frequency) and quantity (e.g. sample size) of information, and that are typically consistent with 'objective' probabilities. On natural sources (EXP N1 and EXP N2), it elicits, for the vast majority of subjects, non-degenerate interval-valued CDFs. All experiments suggest that imprecise beliefs—i.e. intervals of non-zero width—are widespread, providing a choice-based confirmation of the finding of Giustinelli et al. (2022) using stated probability intervals. We also find that the width of elicited intervals decreases when there is more information, familiarity or predictability—a correlation that could be taken to corroborate the solidity of our method. On aggregate, the intervals elicited by our incentive-compatible method in EXP A are generally similar to stated intervals, suggesting that our method provides foundations for some uses of the latter methods in large-scale field studies. Some interesting differences do however emerge, with stated intervals tending to be larger than choice-based ones in information-rich contexts.

The paper is structured as follows. Section 2 sets out the theoretical background and presents the central planks of our approach (the 'matching probability interval' notion and the elicitation method), with the relevant theoretical results. Section 3 describes our experimental implementations, in the form of three studies. Section 4 contains our results and supporting analyses, whereas in Section 5 we discuss connected issues, related literature and future directions. Proofs, further details, data analyses and experimental details are contained in the Appendices.

[3] Many elicitation applications in economics and beyond require subjects' probability distributions or CDFs over a continuous variable of interest (e.g. US inflation in 2025, Eurozone GDP in 2024, average global temperature in 2030).

2. Theoretical Background

In this section, we first set out the general setup, the objects of elicitation and the underlying decision model (Sections 2.1–2.3). Then we present the elements of our method. First, we propose an analogue of MPs for probability intervals and show that they are sufficient to yield the subject's probability interval for an event, in theory (Section 2.4). Then we turn to implementation, presenting, in Section 2.5, an MPI analogue of the two-step of the MP elicitation method developed by Abdellaoui et al. (2021, 2023).

2.1. Bets on events and interval-valued urns

We consider decision-making situations where the objects of choice are two-outcome prospects that pay a fixed monetary outcome z if an event occurs, and nothing otherwise. Prospects with general winning event E and winning amount z are denoted $(z,E,0)$ and called *bets*. The *complementary* bet, which pays out when the event E does not occur, is denoted $(0,E,z)$. Prospects where the probability of winning is exogenously provided in the form of an interval $[\underline{p},\overline{p}]$ are denoted $(z,[\underline{p},\overline{p}],0)$, and are called *interval lotteries* (IL).[4] As for bets, the complementary IL, where the probability of losing is an objectively given interval $[\underline{p},\overline{p}]$, is denoted $(0,[\underline{p},\overline{p}],z)$.

As mentioned previously, interval lotteries are operationalized by urns containing red and blue balls with partial information about the composition. For instance, consider a red-and-blue-balled urn with at least a proportion r of red balls, at least a proportion b of blue balls (with $r+b \leq 1$), but where there is no information about the colour composition of the remaining balls. For such an urn, the information only allows assignment of the interval $[r,1-b]$ for the probability of the next ball drawn from the urn being red; similarly, there is the interval $[b,1-r]$ for the next ball being blue. For the sake of simplicity, we denote the urn with at least proportion r of red balls and at least proportion b of blue balls by $[r,1-b]$. We refer to the set of such *interval-valued urns* by $\mathscr{I}$.[5]

Each urn $[r,1-b]$ in $\mathscr{I}$ can be related to two (sorts of) prospects. One is the prospect that pays z if the next ball drawn from the urn is red, and nothing otherwise. For such a prospect, the probability of winning is characterized by the interval $[r,1-b]$; this thus realises

[4] Our notion of interval lottery is distinct from that used by Gul and Pesendorfer (2014). They use 'interval lottery' to denote (precise) probability measures over the set of intervals of (monetary) prizes; here, 'interval lottery' denotes assignments of probability intervals to (fully determined, precise) outcomes. In particular, the interval lotteries $(z,[r,1-b],0)$ used here clearly do not belong to the concept used by Gul and Pesendorfer (zero probability is assigned to each outcome in the interior of the interval $[0,z]$).

[5] Formally: $\mathscr{I} = \{[x,y] : (x,y) \in [0,1]^2,\ 0 \leq x \leq y \leq 1\}$.

the interval lottery $(z,[r,1-b],0)$. The other prospect involves the complementary bet on this urn—that is, the bet on the next ball drawn from it being blue. Note that the probability of *losing* here is characterised by the interval $[r,1-b]$, so the probability of winning is given by $[b,1-r]$. Hence this prospect realises the complementary IL $(0,[r,1-b],z)$ or equivalently, $(z,[b,1-r],0)$. Standard lotteries correspond to the special case where the composition of the urn is fully known—i.e. $r=1-b$. So, for instance, the *matching probability* (MP) of an event E can be defined in this setup as the r such that $(z,[r,r],0)\sim(z,E,0)$.

2.2. Probability intervals and interval-valued CDFs

Multiple prior belief representations involve a convex, closed set $\mathscr{C}$ of probability measures. For each event E, the set of priors generates a *probability interval* $\{p(E): p\in\mathscr{C}\}=[\underline{p}(E),\overline{p}(E)]$, where $\underline{p}(E)=\min\{p(E): p\in\mathscr{C}\}$ and $\overline{p}(E)=\max\{p(E): p\in\mathscr{C}\}$ are the *lower* and *upper probabilities for E* respectively. In our experiment on artificial sources of uncertainty, the aim is to elicit probability intervals of the relevant events.

The natural sources of uncertainty in our other experiments are real-valued variables, e.g. the daily minimum temperature in Paris between November and March. In the precise probability case, elicitation aims at revealing the subjective probability over the variable, which can be represented as a subjective cumulative distribution function (CDF). One common way of doing so, for a variable taking values in a real interval T, is by eliciting subjective probabilities of events of the form $E_t=\{t'\in T: t'\leq t\}$, i.e. corresponding to the variable lying below certain fixed values. Indeed, for a probability measure $p\in\Delta(T)$, the CDF is defined as $F_p(t)=p(E_t)$. Analogously, a set of priors $\mathscr{C}\subseteq\Delta(T)$ generates the *interval-valued CDF* $F_{\mathscr{C}}(t)=\{p(E_t): p\in\mathscr{C}\}$, which takes the probability interval corresponding to E_t as value, for each t. This can be visually represented in terms of two (real-valued) functions: the *lower CDF*, $\underline{F_{\mathscr{C}}}(t)=\min\{p(E_t): p\in\mathscr{C}\}=\underline{p}(E_t)$, and the *upper CDF*, $\overline{F_{\mathscr{C}}}(t)=\max\{p(E_t): p\in\mathscr{C}\}=\overline{p}(E_t)$. In these experiments, the aim is to elicit subjects' interval-valued CDFs. Although probability intervals and interval-valued CDFs involve an information loss as compared to sets of priors, they are often sufficient for applications, and sometimes preferable insofar as they are easier to communicate. Indeed, interval-valued CDFs are widely used for representing, communicating and studying sets of priors over continuous variables, where they often go under the name of distribution bands or p-boxes (Berger et al., 2000; Karanki et al., 2009).

2.3. Decision model

We only assume that subjects have preferences over bets and interval lotteries. Large parts of our method hold under the representation where a bet $(z, E, 0)$ or interval lottery $(z, [r, 1-b], 0)$ is evaluated according to:

$$W([\underline{p}, \overline{p}])u(z) \tag{1}$$

where $[\underline{p}, \overline{p}] = [\underline{p}(E), \overline{p}(E)]$ (the probability interval for E generated by the subjects' set of priors; Section 2.2) in the case of the bet, and $[\underline{p}, \overline{p}] = [r, 1-b]$ in the case of the IL. In (1), u is a utility function normalized so that $u(0) = 0$, and W is a (real-valued) 'willingness-to-bet function' that is continuous and increasing in both bounds, normalised (i.e. $W([x,x]) = x$ for all x) and strictly increasing in the lower bound. For presentation purposes, we will focus on the special case where W is linear, i.e. where (1) reduces to the Hurwicz α-maxmin EU evaluation of bets and ILs according to:

$$\alpha \underline{p}u(z) + (1-\alpha)\overline{p}u(z) \tag{2}$$

with $\underline{p}, \overline{p}$ and u as above. The mixture coefficient $0 < \alpha \leq 1$ reflects ambiguity attitude in this model, with higher values being associated with more aversion.[6] For instance, $\alpha = 1$ yields the 'maximally ambiguity averse' Gilboa-Schmeidler (1989) maxmin-EU model. For $1 > \alpha > \frac{1}{2}$, (1) accommodates the standard Ellsberg (1961) ambiguity averse preference for a bet on the color of a ball drawn from an urn of known 50-50 composition over a bet on the color of a ball drawn from a 2-color urn of unknown composition, as well as ambiguity seeking behavior at low probabilities.[7] By contrast, such behavior cannot be accommodated when $\alpha < \frac{1}{2}$. Since typical findings suggest some ambiguity seeking behavior at low probabilities, but ambiguity aversion at larger ones (Abdellaoui et al., 2011; Kocher et al., 2018), we take $\alpha > \frac{1}{2}$ to be typical; at a certain point in the presentation, we shall assume that preferences are represented according to (2) with $\alpha > \frac{1}{2}$ (see Sections 2.5 and 5).

Note that the general form (1), which underpins most of the method developed here, can accommodate non-linear, Prospect-Theory-style weighting of the lower and upper probabilities, for instance taking $W([\underline{p}, \overline{p}]) = \alpha w(\underline{p}) + (1-\alpha)w(\overline{p})$, where w is a weighting function and α

[6] The assumption that W is strictly increasing in the lower bound—i.e. decision makers are sensitive to the lower winning probability—rules out the $\alpha = 0$ case of this model, maxmax-EU. However, there is basically no evidence for such preferences in the population.

[7] For instance, when the probability of red from the 2-color red-and-blue unknown urn is characterized by the interval $[0, 1]$, a bet on red from this urn is evaluated as $(1-\alpha)u(z)$ under (2), which is less than the evaluation of a bet on red from the known urn, $\frac{1}{2}u(z)$, when $\alpha > \frac{1}{2}$. However, the evaluation of a bet on the color of a ball drawn from a 10-color urn of unknown composition, $(1-\alpha)u(z)$, is higher than that of a bet on the color of a ball drawn from a 10-equiprobable-color known urn, $0.1u(z)$, whenever $\alpha < 0.9$.

is as in the α-maxmin EU model (2). It can also accommodate transformations of the probability interval $[\underline{p},\overline{p}]$, taking $W([\underline{p},\overline{p}]) = \alpha\underline{\varphi([\underline{p},\overline{p}])} + (1-\alpha)\overline{\varphi([\underline{p},\overline{p}])}$ for α as above and some transformation φ taking probability intervals to probability intervals. Hence it covers cases where the subject's 'real' probability interval is transformed, for instance to incorporate certain ambiguity attitudes, before being used for decision, as in Gajdos et al. (2008). As discussed in detail in Section 5 and Appendix B, the heart of the method applies canonically under such weightings or transformations.

Hill (2023) sets out a formal framework that allows for axiomatic foundations for (2), as well as a range of generalizations.[8] In particular, he shows that introducing ILs allows one to overcome the well-known problem of separating the α factor from the set of priors under α-maxmin EU (Ghirardato et al., 2004; Eichberger et al., 2011), and obtain complete identification of the model. As discussed in more detail there (see notably Hill, 2023, Section 3.3), to the extent that the mixture coefficient reflects a taste (for ambiguity), the use of a single α (or W under (1)) in the evaluation of bets and ILs is consistent with the common practice of using a single utility function for the evaluation of both risky and uncertain prospects, or with the insistence in some parts of the ambiguity literature on the 'portability' of the parameters representing ambiguity attitudes across decision situations (see e.g. Marinacci, 2015, p 1051).

2.4. Matching Probability Intervals

To illustrate our approach, take, as in our EXP A, a bag containing 100 green and yellow chips, where the only information available about its composition comes in the form of four prior draws with replacement, one of which was green. Consider the event E: "the next randomly drawn chip will be yellow". Concerning this event, a multiple-prior decision maker will form a probability interval $[\underline{p}(E),\overline{p}(E)]$ on the basis of the information provided and her beliefs about the proportion of yellow chips in the bag; for instance, it might be $[0.5,0.9]$. The corresponding interval for the complementary event—which tracks the proportion of green chips in the bag—is $[1-\overline{p}(E), 1-\underline{p}(E)]$, i.e. $[0.1,0.5]$ in this example. Our aim is to elicit the interval $[\underline{p}(E),\overline{p}(E)]$ for event E.

Standard matching probabilities do not suffice to reveal this subjective probability interval. Indeed, under α-maxmin EU, eliciting matching probabilities for the bets on yellow (E) and green (E^C), $MP(E)$ and $MP(E^c)$, results in the following two equations

[8] See Grant et al. (2019) for an axiomatisation of a special case of (1).

$$\begin{aligned} MP(E) &= \alpha \underline{p}(E) + (1-\alpha)\overline{p}(E), \\ MP(E^c) &= \alpha(1-\overline{p}(E)) + (1-\alpha)(1-\underline{p}(E)). \end{aligned} \tag{3}$$

in three unknowns. Suppose, for instance, that the elicited matching probabilities for yellow and green were 0.66 and 0.26 respectively. Simple calculation reveals that this is consistent with the DM's actual interval for yellow (E) being $[0.5, 0.9]$ if $\alpha = 0.6$, but it would yield the interval $[0.65, 0.75]$ under $\alpha = 0.9$. Since α is unknown, the probability interval for the target event is not uniquely determined by matching probabilities. To solve this identification problem, Hill (2023) supplements the formal setup of the original α-maxmin EU model with the possibility to calibrate subjective probability intervals against objective interval lotteries. As noted previously, in the current paper, the latter are operationalized through red-and-blue-balled urns with partially known composition.

The decision maker's probability interval for E can be mapped to a unique (objective) interval by finding the interval-valued urn $[r, 1-b]$ such that she is indifferent between betting on the yellow chip from the bag (E) and the red ball from the urn, and between betting on the green chip from the bag and the blue ball from the urn. Formally, this yields:

$$(z, [r, 1-b], 0) \sim (z, E, 0), \tag{4}$$

$$(0, [r, 1-b], z) \sim (0, E, z). \tag{5}$$

We call $[r, 1-b] \in \mathscr{I}$ such that indifferences (4) and (5) hold the *matching probability interval* (MPI) of the event E.

Plugging these indifferences into (1) yields a pair of equations that are clearly satisfied by $r = \underline{p}(E)$, $1-b = \overline{p}(E)$. Under the α-maxmin EU model (2) with $\alpha \neq \frac{1}{2}$, this is the unique solution (Proposition A.2, Appendix A): hence there is a unique MPI, which identifies the subjective probability interval $[\underline{p}(E), \overline{p}(E)]$.[9] As noted in Appendix B, under generic cases of the weighting or probability-interval transformation generalizations of α-maxmin EU discussed in Section 2.3, the MPI is also unique. So to elicit a subject's probability interval for the event

[9]More precisely, under (2) with $\alpha \neq \frac{1}{2}$, the indifferences yield the equations:

$$\begin{aligned} \alpha r + (1-\alpha)(1-b) &= \alpha \underline{p}(E) + (1-\alpha)\overline{p}(E), \\ \alpha(1-(1-b)) + (1-\alpha)(1-r) &= \alpha(1-\overline{p}(E)) + (1-\alpha)(1-\underline{p}(E)). \end{aligned} \tag{6}$$

from which α drops out, yielding a unique solution for $\underline{p}, \overline{p}$.

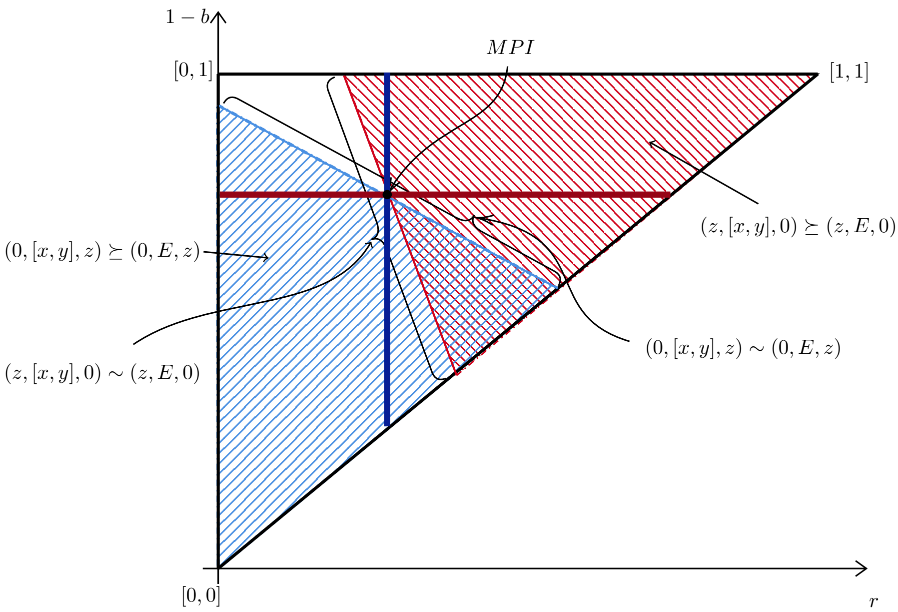


Figure 1: Matching Probability Interval in the space $\mathscr{I}$ of interval-valued urns, for an event E.

E under the main cases of (1), it suffices to find the MPI of E.

The MPI can be conceptually illustrated on Figure 1. A point in the black-edged triangle, (x,y), represents the urn $[x,y]$—i.e. with at least proportion x of red balls and at least proportion $1-y$ of blue ones. As such, it represents two interval lotteries: $(z,[x,y],0)$, the bet on red from the urn, and $(0,[x,y],z)$, the bet on blue. The red hatched area represents the upper contour set (under (2)) of the bet $(z,E,0)$ in the space of interval lotteries corresponding to bets on red: i.e., the set of (x,y) such that $(z,[x,y],0) \succeq (z,E,0)$. The blue hatched area is the upper contour set of the complementary bet $(0,E,z)$ in the space of complementary ILs (corresponding to bets on blue): it is the set of (x,y) such that $(0,[x,y],z) \succeq (0,E,z)$. The boundaries of these sets (the diagonal red and blue lines respectively) represent the indifference curves of $(z,E,0)$ (resp. $(0,E,z)$), in the space of 'red' (resp. 'blue') ILs. The matching probability interval corresponds to the black point at the intersection of these two lines.

Note that standard lotteries and urns with fully known composition correspond to the points on the diagonal ($x=y$) in Figure 1. So the MP of the bet on E is given by the point where the red indifference curve meets the diagonal; and similarly for the MP of E^c and the blue curve. It clear from the Figure that one cannot derive the subject's probability interval from these MPs without knowing the slope of the indifference curves, and this is determined by the ambiguity attitude coefficient α in (2). This is a graphical representation of the previously discussed identification difficulty with MPs. By contrast, the MPI will coincide with the subjective probability interval, independently of the coefficient α.

Name	Preferences	Colour (in Figure 1)
R-B	$(z,[x,y],0) \succeq (z,E,0)$ & $(0,[x,y],z) \succeq (0,E,z)$	Red & Blue
Wh	$(z,[x,y],0) \preceq (z,E,0)$ & $(0,[x,y],z) \preceq (0,E,z)$	White (neither Red nor Blue)
R	$(z,[x,y],0) \succeq (z,E,0)$ & $(0,[x,y],z) \preceq (0,E,z)$	Red
B	$(z,[x,y],0) \preceq (z,E,0)$ & $(0,[x,y],z) \succeq (0,E,z)$	Blue

Table 1: Preference-based division of $\mathscr{I}$

2.5. Elicitation of Matching Probability Intervals

Our strategy for eliciting MPIs is based on an extension of the two-step MP method adopted by Abdellaoui et al. (2021, 2023), where a subject undergoes a 'bisection' binary-choice procedure followed by a 'confirm-or-correct' choice list. Whilst subjects' payments depend solely on the choice list, the binary choice part serves as an aid to filling it in. Here, we develop an analogous two-step procedure that consists in a sequence of binary-choice questions followed by a 'two-dimensional' choice list, tailored for eliciting (two-dimensional) probability intervals instead of (one-dimensional) probability values.

Binary-Choice Procedure. For each event, our subjects first undertake a chained sequence of binary-choice tasks (Section 3.2). Here we set out the general principles of this procedure, leaving full details for Appendix D.1. The logic can be illustrated on Figure 1, notably by dividing the space of interval-valued urns into four preference-defined areas, summarised in Table 1. The procedure is based on the following observation.

Proposition 1. *Suppose preferences are represented according to (2) with $\alpha > \frac{1}{2}$, and let E be an event.*

a. For any $[x,y]$ in the R-B region (i.e. such that the corresponding preferences in Table 1 hold, for E), $\underline{p}(E) \leq x$ and $\overline{p}(E) \geq y$. Moreover, for any $[x,y]$ in the Wh region, $\underline{p}(E) \geq x$ and $\overline{p}(E) \leq y$.

b. For any $[x,y]$ in the R region (i.e. such that the corresponding preferences in Table 1 hold, for E), every $[x',y']$ with $x' \geq x$ and $y' \geq y$ is also in R. Moreover, for any $[x,y]$ in the B region, every $[x',y']$ with $x' \leq x$ and $y' \leq y$ is also in B.

It follows from part a. that if the experimenter has found an interval-valued urn $[x_{RB}, y_{RB}]$ in the R-B region (i.e. with the preference pattern in Table 1, row 1), and a $[x_{Wh}, y_{Wh}]$ in the Wh region, then the MPI is contained in the 'box generated' by these points, i.e. it is in the set $\{[x,y] : x_{Wh} \leq x \leq x_{RB}, y_{RB} \leq y \leq y_{Wh}\}$. The procedure works by searching the smallest such generated box for further points in R-B or Wh, in order to 'reduce' the size of the boxes and hence 'home into' the MPI. In this sense, it is analogous to the bisection procedure for MPs, where preferences indicate that the MP is in a particular interval, and the procedure searches to reduce the width of that interval.

Note that a similar result to Proposition 1 a. does not hold for the R and B regions. However, by part b. it can be concluded, for any interval-valued urn $[x,y]$ in R, that every point North-East of $[x,y]$ is also in R; and similarly for an interval-valued urn in B. So, if the experimenter has just discovered an urn in R (i.e. the elicited preferences for that urn are as specified in Table 1, row 3), then, to seek a point in R-B or Wh, she need not look North-East of this point; and analogously for urns in B. The procedure works, after eliciting preferences for an urn in R and B, by performing a bisection along one-dimensional cuts of the space $\mathscr{I}$ guided by this observation, until an urn in R-B or Wh is found, whence the procedure in the previous paragraph applies again. Details are provided in Appendix D.1. In particular, as shown there (Proposition D.1), the procedure canonically converges to the subject's probability interval for the event, not only under the α-maxmin EU model (2) but also under the generalizations discussed above (see also Section 5 and Appendices A and B.2).

Note finally that the procedure used has an in-built 'precision bias'. Whenever no urn in the R-B or Wh regions has been found, the procedure deliberately moves closer to the space of precise urns (see Appendix D.1). In this way, if there is any misclassification of subjects due to no urns being found in the R-B or Wh regions, the tendency would be for the procedure to represent them as more precise than they actually are.

Two-dimensional Choice List Procedure. For each event, after the binary-choice questions, subjects face a two-dimensional choice list, already filled in according to their responses on the previous procedure. They may modify the preferences encoded on this choice list before confirming. Only the confirmed preferences qualify for payment: so it is essential that the mechanism realized by such choice lists is incentive compatible. We now set out the theory underlying the two-dimensional choice lists. It relies on the following Proposition.

Proposition 2. *Suppose preferences are represented according to* (1). *For any event E, and any* $[r,1-b] \in \mathscr{I}$, $[r,1-b]$ *is a matching probability interval of E if and only if*

$$
\begin{aligned}
(z,[q,1-b],0) &\succ (z,E,0) \quad \textit{for all } q > r, \\
(z,[q,1-b],0) &\prec (z,E,0) \quad \textit{for all } q < r,
\end{aligned}
\tag{7}
$$

and

$$
\begin{aligned}
(0,[r,q],z) &\prec (0,E,z) \quad \textit{for all } q > 1-b, \\
(0,[r,q],z) &\succ (0,E,z) \quad \textit{for all } q < 1-b.
\end{aligned}
\tag{8}
$$

To illustrate, consider any MPI $[r,1-b]$ of an event E, so that the indifference (4) is satisfied. Clearly, under (1), it follows that (7) holds. On Figure 1, this determines the preferences involving the 'red' ILs corresponding to the bold red (horizontal) line. To the left of the MPI, the bet on E is preferred to the IL corresponding to the bet on red from the urn $[q,1-b]$ (i.e. with probability $[q,1-b]$ of winning); to the right of the MPI, the IL is preferred to the bet; and at the MPI, the two are indifferent. Likewise, the indifference (5) concerning the complementary bet determines preferences involving the 'blue' ILs corresponding to the bold blue (vertical) line, with the MPI being the point on that line where preferences 'switch' from the bet concerning E to the IL. The red (horizontal) and blue (vertical) bold lines in Figure 1 are thus analogous to a pair of choice lists, and the MPI is the switching point on each of them. We henceforth refer to the combination of the two as a *2D choice list*. By Proposition 2, we know that any urn that is a switching point on both branches of a 2D choice list is an MPI.

Inspired by this observation, consider an incentivization mechanism in which a subject who confirms the interval-valued urn $[r,1-b]$ for event E is remunerated as follows. Each urn $[x,y]$ in the 2D choice list—i.e. each $[x,y] \in \{[x,y] \in \mathscr{I} : y = 1-b\} \cup \{[x,y] \in \mathscr{I} : x = r\}$—determines a binary choice $\Phi_{[r,1-b],E}([x,y])$, defined as follows:

$$
\Phi_{[r,1-b],E}([x,y]) = \begin{cases} \{(z,E,0),(z,[x,y],0)\} & \text{if } y = 1-b \\ \{(0,E,z),(0,[x,y],z)\} & \text{if } x = r, y \neq 1-b \end{cases}
\tag{9}
$$

In terms of Figure 1, if the urn is on the horizontal line going through $[r,1-b]$ (e.g. the bold red horizontal line in the Figure, if $[r,1-b]$ is the MPI), the choice is between the bet on the event and the bet on red from the urn; if it is on the vertical line going through $[r,1-b]$, the choice is between the complementary bet and the complementary IL. The incentive scheme selects an option for each of these possible choices: for the choice corresponding to urn $[x,y]$ it selects $\phi_{[r,1-b],E}([x,y])$, defined by:

$$\phi_{[r,1-b],E}([x,y]) = \begin{cases} (z,E,0) & \text{if } y = 1-b,\ x < r \\ (z,[x,y],0) & \text{if } y = 1-b,\ x \geq r \\ (0,[x,y],z) & \text{if } x = r, y < 1-b \\ (0,E,z) & \text{if } x = r, y > 1-b \end{cases} \tag{10}$$

I.e. if the urn $[x,y]$ has $y = 1-b$, $x < r$, then this selects the option $(z,E,0)$—the subject 'plays' the bet on E—and similarly for the other cases. The incentive mechanism first draws an urn $[x,y]$ at random from $\{[x,y] \in \mathscr{I} : y = 1-b\} \cup \{[x,y] \in \mathscr{I} : x = r\}$, and hence the choice $\Phi_{[r,1-b],E}([x,y])$; it then pays the subject according to the outcome of the selected bet or IL, $\phi_{[r,1-b],E}([x,y])$. It follows immediately from Proposition 2 that this mechanism is incentive compatible in the sense of weak dominance.

Corollary 1. *Suppose preferences are represented according to (1), and let $[r,1-b]$ be such that, for every urn $[x,y] \in \{[x,y] \in \mathscr{I} : y = 1-b\} \cup \{[x,y] \in \mathscr{I} : x = r\}$, $\phi_{[r,1-b],E}([x,y])$ is a weakly dominant option in $\Phi_{[r,1-b],E}([x,y])$. Then $[r,1-b]$ is a matching probability interval of E.*

In other words, among all probability intervals that the subject could report, only matching probability intervals are such that the option selected by the mechanism is (weakly) preferred, no matter the choice in the 2D choice list that is played 'for real'. Hence implementing this incentive scheme on a subject's confirmed 2D choice list incentivizes reporting her MPI for the event. Since precise probabilities (and SEU) are a special case of multiple priors (respectively, Eq. (1)), this mechanism functions equally for Bayesian decision makers, who are incentivized to report their precise probabilities. We set out the experimental implementation of 2D choice lists in Section 3.2.

Note finally the depth of the analogy with choice lists for MPs. There, MPs are determined by the switching point, i.e. the maximum probability for which the subject prefers the bet on the target event over the lottery with that probability of winning. Similarly, the proposed probability-interval incentive mechanism elicits a single point, which is the switching point on each branch of the 2D choice list. Moreover, in standard MP choice lists, the switching point determines the preferences in the rest of the choice list by stochastic dominance. Similarly,

	(Frequency, Sample size)					
	Group A			Group B		
Choice-based MPI	(0.50, 4)	(0.25, 20)	(0.50, 100)	(0.25, 4)	(0.50, 20)	(0.25, 100)
Stated probability interval	(0.25, 4)	(0.50, 20)	(0.25, 100)	(0.50, 4)	(0.25, 20)	(0.50, 100)

Table 2: Prior information faced by subjects in EXP A: frequency of green chips in the previous sample, and sample size.

Proposition 2 guarantees that the elicited point determines the other preferences in the 2D choice list according to a probability-interval analogue of stochastic dominance, which states that, between ILs $(z,[r,1-b],0)$ and $(z,[r',1-b],0)$, decision makers prefer the prospect with higher lower probability.[10] Finally, in standard MP choice lists, this property underlies the incentive compatibility: it ensures that only the MP is such that, no matter the choice played 'for real' from the choice list, the selected option is preferred by the subject. Corollary 1 establishes an analogous result for MPIs and the proposed incentive mechanism.

3. Experimental Methods

We applied our probability-interval elicitation method in three experiments. One, EXP A, evaluated the effectiveness of our method in a simple setup using artificial sources of uncertainty. The target events were the color of the next chip randomly drawn from bags filled with 100 yellow and green chips, where the only information about each bag's content came from earlier draws conducted with replacement. The other two experiments, EXP N1 and EXP N2, involved uncertain events stemming from natural sources: the minimum winter temperatures in Paris and Sydney in EXP N1, and the test scores for two admission pathways at a French Business School in EXP N2.

3.1. Subjects

233 students completed the experiment: 101 from the INSEAD-Sorbonne Behavioral Lab (Paris, France) for experiment EXP A, 80 from university of Paris 1 for EXP N1 and 52 from HEC Paris Business School for EXP N2. Subjects' choices were collected through computer-based individual interviews that lasted about one hour in each study. Each individual interview started with a video presentation of the experimental instructions, followed by comprehension

[10]This 'Lower Stochastic Dominance' property, which is equivalent to the assumption (Section 2.3) that W is strictly increasing in the lower bound, is behind the preference patterns in Proposition 2; see Appendix A.

questions and one training MPI elicitation task (on an event not involved in the ensuing experiment). Appendix D.2 contains screenshots and a link to the video instructions for EXP A.[11] In all experiments, subjects were told that there were no right or wrong answers, and that they could ask any question regarding the experiment. Differences in experimental instructions between the experiments are explained in the sequel.

3.2. Artificial sources of uncertainty

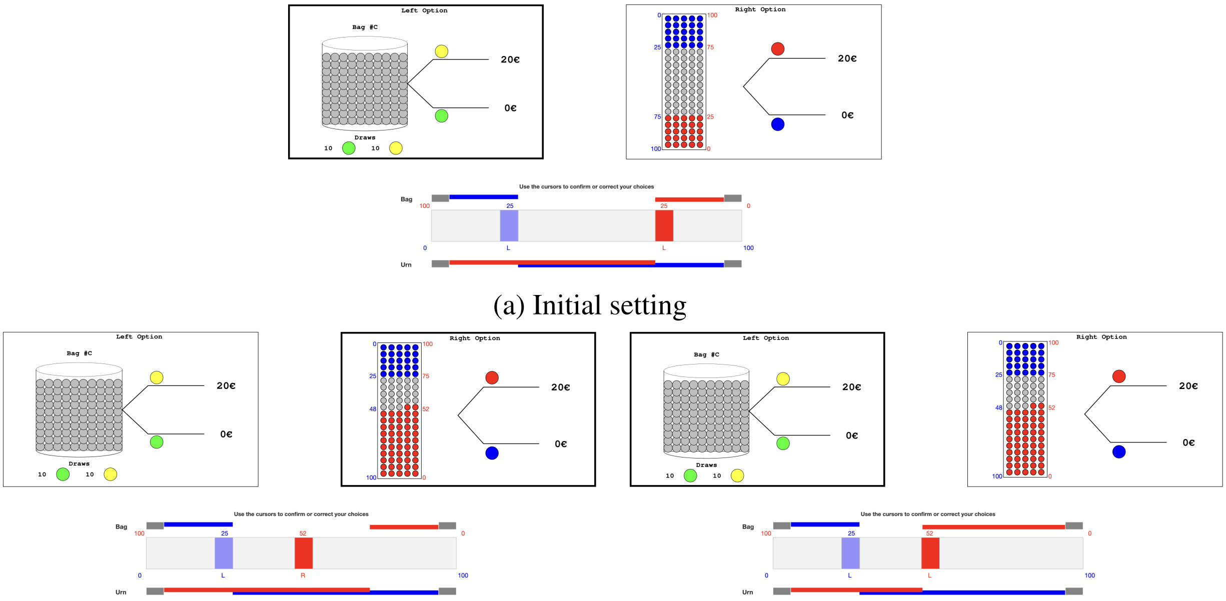
Sources and choice tasks. The sources of uncertainty in EXP A were physical, opaque, labeled bags containing 100 green or yellow chips, with prior information about the composition coming in the form of prior draws with replacement (Appendix D.2). Different bags corresponded to different prior information, i.e., different sample size and frequency of green in the preceding draws.

Subjects were randomly allocated to one of two groups. The tasks for each group are specified in Table 2. Each group carried out two blocks of tasks. Each block involved three different bags. For a given group of subjects and a given bag in the *choice-based elicitation* block, subjects' probability intervals for the event that the next chip drawn from the bag was green were elicited using the proposed method. Then the subjects were asked to state their (precise) probability that the next chip was green (on a one-cursor slider). In the *stated* block, for each bag, subject were asked to state their probability interval for the next chip being green (on a standard, two-cursor slider), and then, as for the choice-based task, give their (precise) probability. The order of the bags within blocks was randomized, as was the order of the blocks in each group. As is clear from Table 2, all subjects provided probability intervals for each bag: for one group, these were elicited using our method; for the other they were stated directly.

Elicitation procedure. As noted in Section 2.5, our choice-based method for eliciting probability intervals follows the general two-step structure adopted by Abdellaoui et al. (2021, 2023) for MP elicitation: a binary-choice procedure is first used to aid subjects to fill in responses on a choice list, which they then confirm or modify.

More specifically, for each event E (e.g. drawing green from a specific bag), we first applied the binary-choice procedure set out in Section 2.5 and Appendix D.1. Each stage of the procedure consisted of two binary choices involving bets concerning E and bets on the color

[11] Beyond those who completed the experiment, 19 subjects in EXP A, 12 in EXP N1 and 0 in EXP N2 did not pass the comprehension check. They received a flat payment but were not given the possibility to continue the experiment.



(a) Initial setting

(b) After having moved red cursor

(c) After having clicked on left option

Figure 2: 2D confirmation choice list: displays.

of a ball drawn from a partially known 100-ball urn with a specified minimum proportions b and $r \leq 1-b$ of blue and red balls, respectively. All bets involved the same winning and losing outcomes. For each event and urn used, we collected the subject's choice in the decision between the bet on the event E and the bet on the next ball drawn from the urn being blue; in the subsequent choice question, we elicited their choice in the decision between the bet on E^c (or against E) and the bet on the next ball drawn from the same urn being red. (See Figure D.6, Appendix D.2 for illustrative examples of binary choices in our experiments.) These elicitations situated the urn in one of the areas in Table 1. The urn proposed in the next stage depended on the preferences elicited in the previous choices according to the binary-choice procedure (Section 2.5 and Appendix D.1). The subjective probability interval for E elicited at the end of the procedure is deduced from the preferences over such bets, as specified in the cited sections. The procedure continued until the interval was estimated to a precision of 0.15 if it was not degenerate, 0.05 if it was degenerate (i.e. corresponded to a precise probability), or up to 12 stages, whichever came first. The probability interval produced was fed into the next, two-dimensional choice-list 'confirmation' step of the elicitation procedure.

To illustrate the 'confirmation' procedure, Figure 2a shows the screen that a subject would see after the binary-choice procedure returns an interval $[0.25, 0.75]$ for the draw of a green ball from the specified bag. The corresponding 2D choice list is materialized by means of a two-cursor scrollbar. The red and blue cursors in the scrollbar determine minimum number of red

	Sources	Events $E_{t_i} = \{t' \in T : t' \leq t_i\}$ for t_i:
EXP N1	Paris	$-2, 2, 5, 8$
	Sydney	$15, 18, 20, 22$
EXP N2	Maths	$7, 10, 12, 15, 17$
	Contraction	$7, 10, 12, 15, 17$

Table 3: Natural sources of uncertainty and events in EXP N1 and EXP N2

and blue balls respectively, hence specifying the urn on the right. The chosen option between the bet on the bag (on the left) and the bet on the specified urn is highlighted. By moving the cursors, the subject can scan the choices associated to different urns. In particular, when moving the red cursor, the blue cursor remains fixed at the pre-specified value: so the subject scans all the urns with the same minimum number of blue balls but differing minimum numbers of red balls. In terms of Figure 1, this corresponds to the choices represented by the horizontal line through $[0.25, 0.75]$. When the red cursor is set far to the left, the minimum number of winning red balls in the urn is low, and the bet on the urn is less attractive (this corresponds to urns on the left of the horizontal line). As the red cursor is shifted further to the right, the minimum number of winning red balls increases, and the bet on the urn becomes more attractive: as indicated on Figure 1, the preference switches in favour of the bet on the urn at some point. Figure 2b illustrates the state of the scrollbar when the red cursor is moved to 52. The subject can modify her choices, for any setting of the red cursor, by clicking on the preferred bet. As illustrated in Figure 2c, which shows the result of clicking on the bag for the previous cursor setting, this updates the slider. Note in particular that the horizontal lines above and below the slider, which indicate the preferred options for each choice, are updated. Similarly, the subject can scan and modify choices involving urns with different minimum number of blue balls (for a fixed minimum number of red balls) by moving the blue cursor. See Appendix D.2 for further details, as well as a link to an online version of EXP A.

After any modifications, subjects had to reconfirm all of the associated choices, by moving one cursor then the other, before continuing on to the next phase of the experiment. The precision of the scrollbar, and hence subject responses, was to the nearest 0.01 (to the precise minimum number of red and blue balls out of 100 respectively).

3.3. Natural sources of uncertainty

Sources. Each of EXP N1 and EXP N2 involved two comparable natural sources of uncertainty. The type of source in EXP N1 was the minimum daily temperature over the previous

November–March period; the sources differed in the city whose temperature was of interest—Paris, where the experiment was carried out, and Sydney. The typical winning event E_{t_i} in this case was of the form: "the minimum temperature on day D in Paris (or Sydney) was less than or equal to t_i", where D was a randomly chosen day in the specified period (see Section 3.4). For each source in EXP N1, we chose temperature value t_i's close to the 10%, 33%, 66% and 90% percentiles of the true distribution (Table 3).

EXP N2 involved marks in two of the previous year's entrance exams for admission at undergraduate level to a prominent French business school, HEC Paris.[12] The subjects in the experiment had sat these exams either in the previous Spring or in the one before. The sources differed in the exam considered: a Maths exam, which is generally considered to be 'objectively marked', and the 'Contraction' exam—a summary of a philosophical or literary text—whose marking is considered more 'unpredictable' by candidates and students. Indeed, the marks in the latter exam have higher variance.[13] The typical winning event E_{t_i} here corresponds to: "candidate C obtained a mark less than or equal to t_i in the Maths (or Contraction) exam" for a randomly drawn candidate C. We used the same values for both sources (Maths and Contraction), picked so they would seem to reasonably scan the range and correspond to comparable points in the true distribution over Contraction scores, where they were at the 3%, 15%, 33%, 68% and 86% percentiles (Table 3).[14]

Choice tasks in EXP N1. Each subject undertook three blocks of tasks. Each of the first two blocks concerned a single source (Paris or Sydney), and involved the elicitation of the probability intervals for each of the events in the source (Table 3). The order of these two blocks was randomized. In each block, the subject first declared, in an non-incentivized manner and using a scrollbar, her estimated maximum and minimum values for the minimum temperature on the unidentified day D (see Section 3.4). This is standard procedure in expert elicitation for unbounded sources, aimed at combating anchoring bias (Morgan, 2014), and played no role in our elicitation. Then the elicitation procedure set out in Section 2.5 and implemented as in EXP A (Section 3.2) was applied for each event in the source. Within each block, the two extreme events (i.e. lowest and highest temperature points) were asked first, in a random order, followed by the other two events, in a random order.

[12] All candidates to this school at undergraduate level must apply in an entrance stream, each of which involves a different set of exams. The exams whose marks were involved in this experiment were sat by all candidates in both the 'ECS' (scientific) and 'ECE' (economics) streams. All subjects in this experiment were students admitted to the school through one of these streams.

[13] The variance of marks for Maths is 3.77, where it is 9.92 for Contraction.

[14] They were at the 0%, 0%, 2%, 21% and 60% of the true distribution of Maths scores.

The final block involved the elicitation of MPs for the events in Paris treatment, using the two-step bisection-then-choice-list procedure from Abdellaoui et al. (2021) (see Appendix D.2 for details). MPs were elicited for each event E_{t_i} in this source and its complement $E^c_{t_i}$ (Table 3). The order of elicitations was randomized in this block.

Choice tasks in EXP N2. Each subject undertook two blocks of tasks. Each of the blocks concerned a single source (Maths or Contraction), and involved the elicitation of the probability intervals for each of the events in the source (Table 3). The order of the blocks was randomized, as was the order of the events in each block. In each block, the elicitation procedure set out in Section 2.5 and implemented as in EXPs A and N1 was applied for each event in the source. Each block ended with an omnibus confirmation screen, in which the interval-valued urns elicited for each of the events in the source were displayed in graphical form (see Appendix D.2 for details). The subject was given the opportunity to select and modify any of her responses for the events in the source. This screen, the sources and the larger number of events elicited per source (see Table 3) were the central differences with respect to EXP N1.

3.4. Incentivizing Subjects

Participants in all studies received a flat payment of €10. Additionally, a random incentive system was implemented, which was entirely analogous to those standardly used to implement elicitation of MPs. In EXPs N1 and N2, after the presentation of the instructions and before the beginning of the experiment, the subject chose a number from a given range, which identified an individual case of the variable of interest (the day D, if the source was minimum temperature; the candidate C, if the source was the mark). The exact case identified was specified according to a spreadsheet that would only be revealed to the subject at the end of the experiment. At the end of each of the three experiments, a choice list (a 2D choice list in EXPs A and N2; a 2D choice list or MP-choice list in EXP N1) and choice on it were selected at random by the computer.[15] The subject was then paid according to the decision she had made on that choice. If she had chosen, say, the bet on the event that the minimum temperature in Paris is less than or equal to 2°C, then the day which she chose was revealed, as well as daily temperature data for the November–March period, and she won if the minimum temperature on that day was indeed 2°C or less; if not, she lost. Or, in EXP A, if she had chosen the bet that the next

[15]More precisely, for the selected choice list, a color—red or blue—was selected at random, and then an urn on the branch of the 2D choice list corresponding to that color was selected at random.

chip in a certain bag was green, then a chip was drawn randomly from that bag and she won according to its color. If she had chosen the urn, then she composed the appropriate urn—she counted the specified minimum numbers of red and blue balls, with the remaining balls coming from pre-constructed Ellsberg urns (of unknown composition). Then a ball was drawn from the constructed urn, and she was paid according to whether she bet on the color of that ball or not. All bets yielded €20 if won, and nothing otherwise.

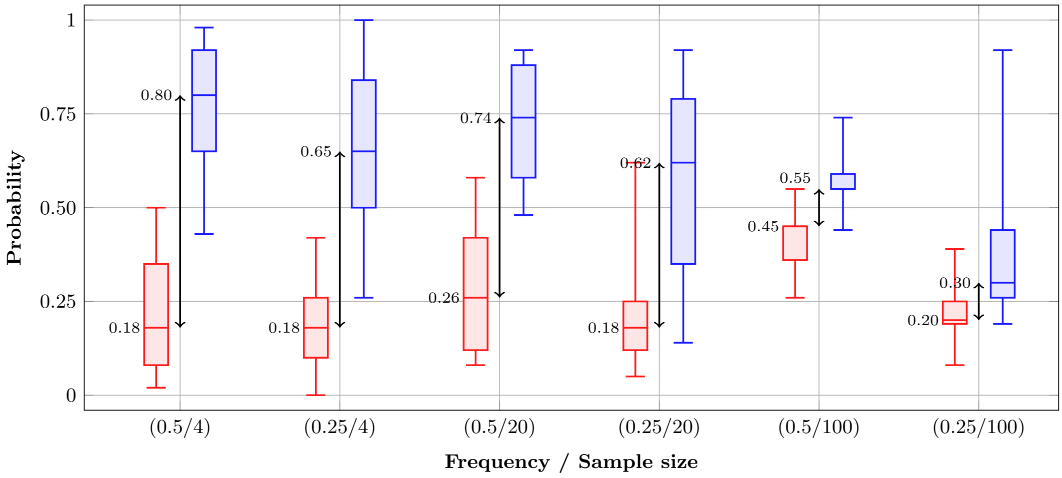
4. Results

4.1. Performance and Validation

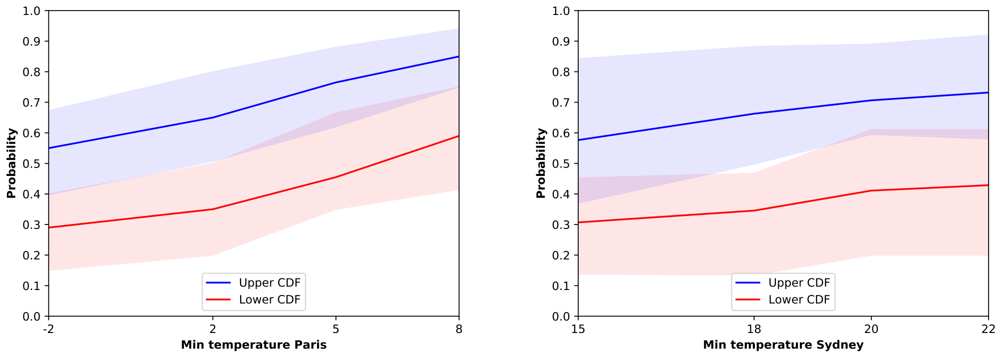
Figure 3 plots the 25%, 50% and 75% quantiles (Interquartile Ranges, i.e. IQRs) of the upper and lower probabilities and CDFs for all events elicited used our method and all experiments (see Table C.1 in Appendix C.1 for basic descriptive statistics). This Figure already gives some early indications about our results, and the performance of our elicitation method.

EXP A. First of all, Panel (a) of Figure 3 shows how the 'balance' of evidence, as represented by the observed frequency of green chips in EXP A, affects the position of the elicited probability intervals: they are higher when the observed frequency is larger. For each sample size, unpaired t-tests and Mann-Whitney tests reject the null hypothesis of equal midpoints of the elicited interval for different observed frequencies in the previous draws ($p < 0.001$ in all cases), with the means being higher for higher frequencies. Since one would expect such sensitivity of posterior beliefs, the method passes this first 'validation' check.

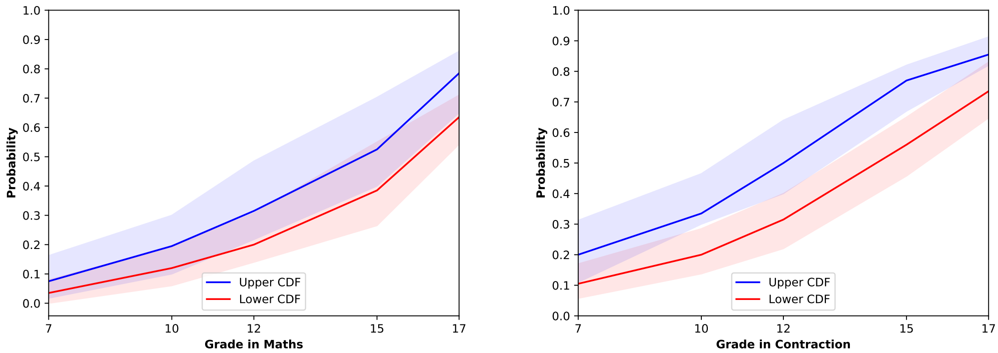
Another possible test for the method, that can be applied under the artificial source of uncertainty, is the comparison with the posterior probabilities of a Bayesian who updates a uniform prior with the same information by Bayes rule. For 80% of events, across all subjects, this 'objective Bayesian' probability was contained in the elicited probability interval, suggesting that in the vast majority of cases, subjects did not rule out the Bayesian probability in forming their posterior probability intervals. Moreover, even in the cases where the Bayesian probability was not in the interval, it was not far, with the average minimal distance to the interval among instances where the Bayesian probability was not contained in it being less than 0.06. Unsurprisingly, the midpoints of the elicited intervals were substantially correlated with the Bayesian probability: the Spearman correlation was 0.65.



(a) EXP A



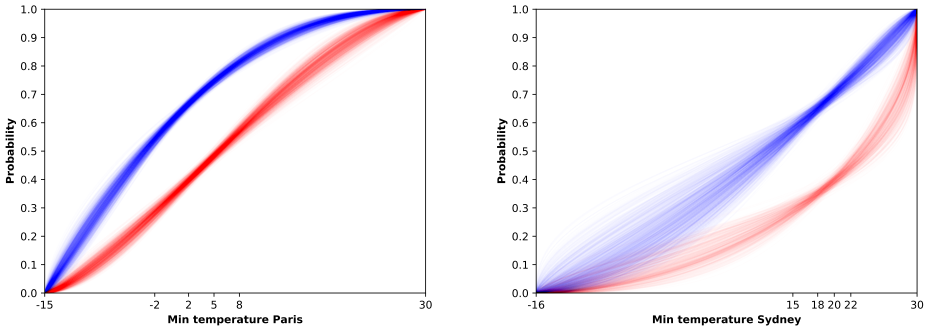
(b) EXP N1



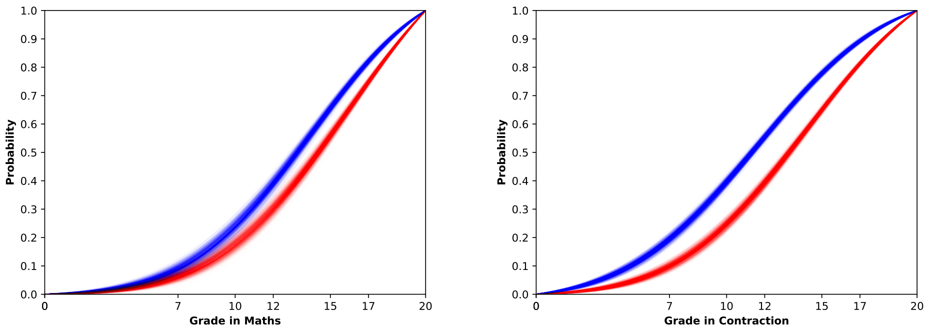
(c) EXP N2

Figure 3: IQRs of upper and lower probabilities and CDFs

EXPs N1 & N2. The experiments eliciting interval-valued CDFs for natural sources of uncertainty suggest that the general message of validity extends to such contexts. Echoing the sensitivity to frequencies found in EXP A, the upper and lower CDFs in the other experiments differ across subjects and events—thus suggesting the consistency of the method. A validity



(a) EXP N1 (Beta distribution: Min temperature)



(b) EXP N2 (Truncated Normal distribution: Grade)

Figure 4: Bayesian estimation of lower and upper CDFs: plots of 1000 samples from MCMC.

test in this context would examine whether the upper and lower CDFs are increasing—an issue that can be investigated using the Kendall rank.[16] As reported in Table C.5, Appendix C.1, the median Kendall τ_b is far greater than 0 for all sources, pointing to increasing upper and lower CDFs. In EXP N2, where subjects were given the opportunity to confirm their replies on all the 2D choice lists for a source (Section 3.3), CDFs were strictly increasing (Kendall rank of 1) for the vast majority of subjects. In EXP N1, there were more violations of monotonicity; however, the median Kendall ranks for upper and lower CDFs in the Paris treatment were similar those obtained under the more standard MP elicitation (Table C.5). This suggests that violations were not unique to the probability-interval elicitation method proposed here.

Finally, we re-analyse the data from these experiments under a standard Bayesian approach, estimating hyperparameters for upper and lower CDFs using a MCMC procedure. Figure 4 plots 1000 MCMC samples for each of the upper and lower distributions, for each source, under the parametric distributions for upper and lower CDFs that offer the best fit (see Tables

[16]The Kendall τ_b is an indicator of ordinal association: the value 1 indicates that the CDFs or MPs are strictly increasing; 0 suggests that there is no association between the elicited probability and the size of the event; -1 indicates a strictly decreasing relationship between the two.

C.16, C.15 and C.17; Appendix C.2). They suggest that the proposed elicitation technique supports parametric estimation of subjective probability intervals in the population, insofar as they chime with expectations given the nature of the events. For instance, they suggest that the dispersion of subjective upper and lower probabilities is larger for the temperature source (EXP N1) than the grade source (EXP N2), which could be related to the fact that all subjects in EXP N2 had sat both exams, and were very interested in the marking, several months before. Also, within EXP N1, there is more dispersion in the estimated distributions for Sydney than for Paris, as would be expected given the less familiar nature of the former source for Paris subjects.[17]

4.2. Imprecision

Overall Imprecision. Our raw data (Figure 3) suggest that subjects' beliefs are often *imprecise*: i.e. there is a gap between their upper and lower probabilities, as indicated in Figure 3a by the arrows connecting the median upper and lower probabilities for each bag. For further analysis, we define a subject's *Imprecision concerning an event E* to be the width of her elicited probability interval for E, i.e. $\overline{p}(E) - \underline{p}(E)$. A subject's *Average Imprecision* across all elicited events in EXP A, or across all elicited events in a source in EXPs N1 and N2, gives an indication of how imprecise the subject's beliefs are, on average, across the relevant events. Naturally, an SEU decision maker will assign precise probabilities to all events, and hence have imprecision 0 (for all events and sources).

Figure 5 displays the 25%, 50% and 75% quantiles, and max and min subject-level Average Imprecision across all sources in all experiments (see also Table C.7, Appendix C.1). It clearly suggests a prevalence of imprecision, with mean and median Average Imprecision greater than 0.1 for most sources and experiments. Binomial tests reject the hypothesis of equal probability for the Average Imprecision to be equal to vs. greater than 0 for each source ($p < 0.001$ in all cases), with a clear majority of subjects—99 out of 101 in EXP A, 79 out of 80 in EXP N1, and 52 out of 52 in EXP N2—having strictly positive Imprecision on average. The data on the number of precise events—events for which the subject's probability interval has zero width—tells a similar story, with not more than around 5% of subjects giving precise probabilities for all events in a single source (Table C.8, Appendix C.1).[18]

[17] More precisely, it is clear from Tables C.17a and C.17b that the standard deviations of the parameters for the Paris source are lower than for Sydney.

[18] Further analysis, reported in Appendix C.1, confirms that the observed imprecision in elicited probability intervals cannot be explained by imprecision in the elicitation procedure.

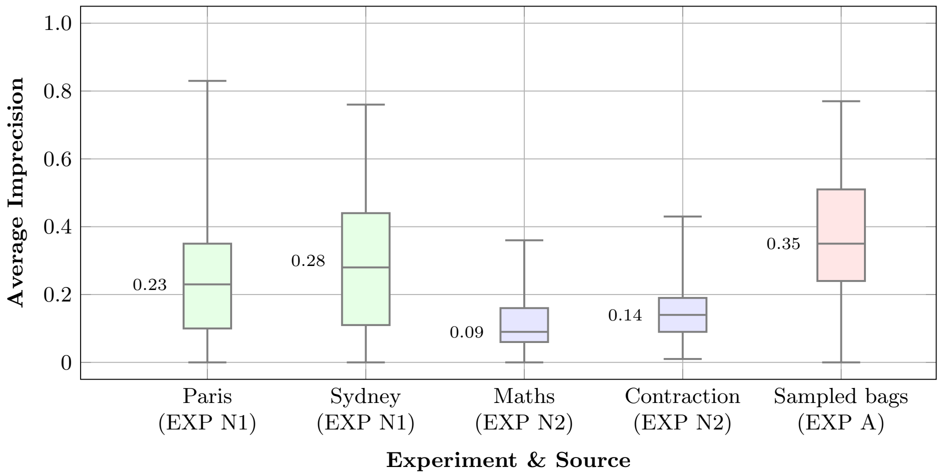


Figure 5: IQRs of Average Imprecision for all events in EXP A and across sources in EXPs N1 and N2

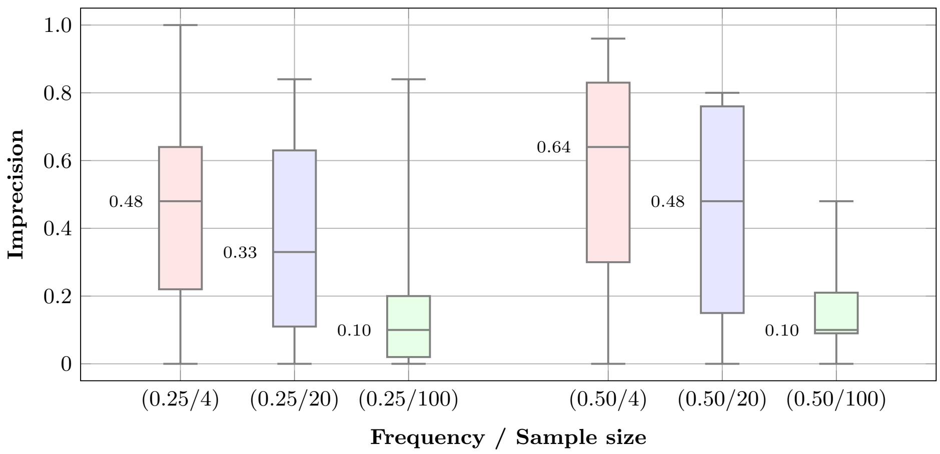
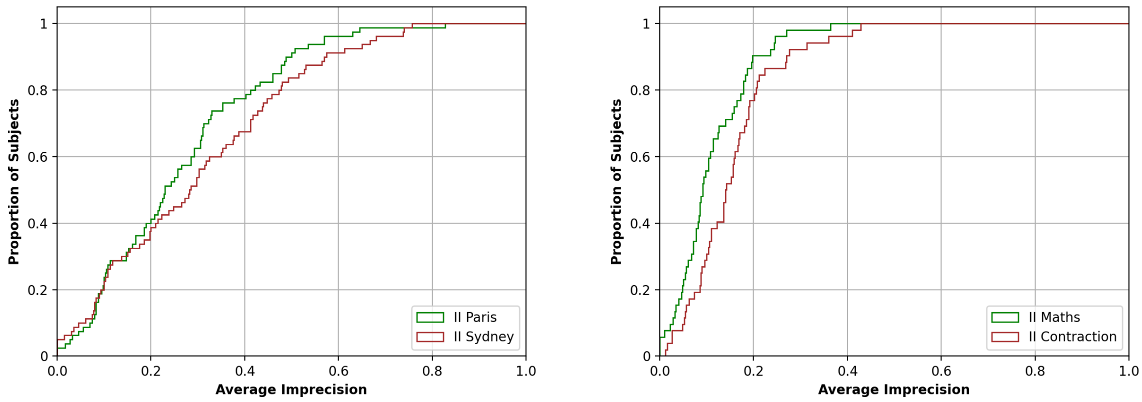


Figure 6: IQRs of Imprecision at varying frequencies and sample sizes, EXP A

Information and familiarity. A reasonable hypothesis is that, *ceteris paribus,* subjects' beliefs are more imprecise concerning events with which they are less familiar, or about which they feel as if they have less knowledge or information. In terms of multiple priors models, this corresponds to wider probability intervals for events for which there is less information. Given the explicit control on the information available via the observed sample size, EXP A allows for a particularly clear examination of the effect of information on imprecision.

As is clear from Figure 6, which displays the 25%, 50% and 75% quantiles, and max and min Imprecision across subjects for each frequency and sample size observed, Imprecision decreases with sample size. Recall (Table 2, Section 3.2) that every subject's interval was elicited, for a given frequency, at sample sizes 4 and 100, so the relationship between Imprecision and



(a) EXP N1 (b) EXP N2

Figure 7: CDFs of Average Imprecision per source across subjects

information can be tested at a within-subject level. Paired t-tests reject the null hypothesis of equal Imprecision across these sample sizes, for both frequencies explored ($p < 0.001$ in both cases). Binomial tests of the null hypothesis of equal chance of one Imprecision being higher come to the same conclusion ($p < 0.001$ in both cases), with 49 out of 51 subjects (respectively 45 out of 50 subjects) having a more imprecise interval for the smaller sample size at observed frequency 0.5 (resp. 0.25). That the proposed elicitation method captures an expected relationship between imprecision and information in the carefully controlled environment of EXP A further bolsters its credentials.

Whilst allowing for less control, the natural sources of uncertainty used in EXPs N1 and N2 also support differences in perceived information, with (Paris-based) subjects likely to be less familiar with the weather in Sydney than that in Paris, and the Contraction exam in EXP N2 generally being considered to be 'less predictable' than the Maths one (Section 3.2). The CDFs of the Average Imprecision per source across subjects, plotted in Figure 7, suggest a relationship between imprecision and familiarity or predictability of the source. The CDF for Contraction—known as the less predictable exam—is entirely to the right of that for Math, indicating a larger Average Imprecision at the subject level. Similarly, the CDFs for Sydney—the less familiar source—is to the right of that for Paris for a large range of values, suggesting that more imprecision for this source. A paired t-test barely fails to reject the null hypothesis of identical Average Imprecision across the sources in EXP N1 ($p = 0.0895$), whilst it rejects it for EXP N2 ($p = 0.0016$). A Binomial test comes to similar conclusions ($p = 0.576$ for EXP N1; $p = 0.017$ for EXP N2), with 45 out of 80 (resp. 35 out of 52) subjects having a more imprecise interval on average for Sydney in EXP N1 (resp. Contraction in EXP N2).

Event-level Imprecision. We also investigate imprecision at the event level within sources in EXPs N1 and N2. One-way ANOVAs of the Imprecision (dependent variable) against the event (factor) reject the null hypothesis of identical imprecision across all events for the sources in EXP N2 ($p < 0.001$ for Maths; $p = 0.003$ for Contraction), whilst failing to reject it for the sources in EXP N1 (Table C.9, Appendix C.1). These conclusions are also illustrated in CDFs of the Imprecision for each elicited event in each source, across subjects (Figure C.1, Appendix C.1). This suggests not only that imprecision is widespread, but that imprecision may be event dependent within sources, as one would expect if some events are intuitively more uncertain than others. For instance, the least imprecise event in EXP N2 involves, for both sources, the lowest grade, where many subjects are presumably more sure of their judgements.

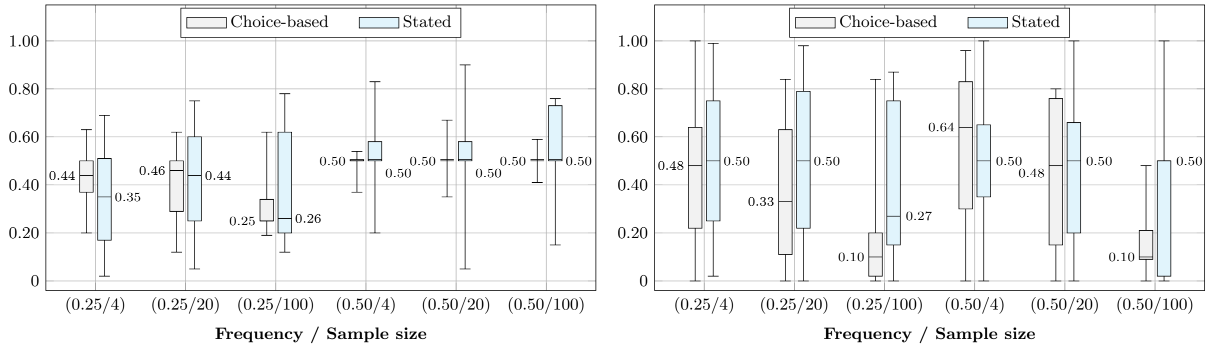
In summary, our method reveals that, when beliefs are elicited with a method allowing for (non-degenerate) probability intervals, imprecision is widespread, at least for the events considered here. Crucially, we recover an expected relationship between imprecision and perceived information in both controlled artificial sources and natural ones. This can be seen as providing further indirect evidence for the solidity of the proposed elicitation method. Finally, at least within some sources, the extent of imprecision may depend on the event.

4.3. Matching versus Stated Probability Intervals

Recall that in EXP A, the same events (concerning bags as characterised by frequencies and sample sizes) were elicited across different subjects using different methods: some subjects underwent the proposed incentive-compatible method, while others were asked for stated probability intervals, as in Giustinelli et al. (2022). This permits between-subject comparison of the results of interval elicitation under the two methods.

Figure 8 displays the 25%, 50% and 75% quantiles, and min and max of the distribution of the interval midpoints and Imprecision across subjects, for each event (concerning a bag characterised by frequency and sample size) and elicitation method. In the aggregate, the stated intervals are roughly comparable to those elicited under the proposed incentive-compatible method for most events, though the dispersion across subjects may differ at some points. Given the theoretical well-foundedness of the method developed here, this could be understood as providing validation for the use of stated intervals in large-scale field studies aimed at eliciting aggregate characteristics.

The Figure does however suggest some interesting differences between the intervals elicited by our method and stated intervals. For one, there is a greater dispersion in the *position* of inter-



(a) Midpoints

(b) Imprecision

Figure 8: IQRs of probability interval midpoints and Imprecision across frequencies and sample sizes in EXP A, for the proposed choice-based elicitation method and stated intervals.

vals, as is clearly visible on the plots of the midpoints in Figure 8a. This is particularly notable for the 'symmetric' case of frequency 0.5, which are more tightly centred on the midpoint of 0.5 under the choice-based elicitation method. More importantly, the imprecision tends to be *lower* under incentive-compatible elicitation as compared to stated intervals, especially when there is lots of information (Figure 8b). For instance, both unpaired t-tests and Mann-Whitney tests reject the null hypothesis of equal imprecision for both frequencies under sample size 100 ($p < 0.001$ in all cases), though this is not the case for smaller sample sizes. This suggests that the expected link between information and imprecision is tighter for intervals elicited under our method. Moreover, it hints that the use of stated interval elicitation may tend to overestimate imprecision, especially in information-rich environments.

Moreover, we can undertake analysis at the individual subject level, using the Average Imprecision, defined in Section 4.2, for each subject and each elicitation method. The Spearman correlation between subjects' Average Imprecision under our elicitation method vs stated probabilities intervals is 0.23 ($p < 0.05$), suggesting that the stated method does fairly well at identifying subjects whose intervals are wider on average.

4.4. Stated probabilities, matching probabilities and the α-maxmin EU mixture coefficient

Recall that, in EXP A, subjects provided their stated (precise) probabilities for the events; comparison with the elicited probability intervals may provide another 'sanity check' for the method. For 77% of events, across all subjects, the stated probabilities were contained in the intervals elicited by our method; this figure rose to 81% if one removes stated probabilities suggesting limited effort on the part of subjects.[19] Across such responses, stated probabilities were strongly correlated with the midpoints of the elicited intervals (Spearman correlation of 0.49).[20]

Further insight can be gleaned from EXP N1, which contained a choice-based elicitation of MPs for the events in the Paris source (Section 3.3). As noted in the Introduction and Section 2.4, under the α-maxmin EU model (2), MPs for an event and its complement generate a pair of equations (Eq. (3)) that cannot be solved for α and the upper and lower probabilities of the event in general. However, drawing on the elicited MPs and our elicitations of upper and lower probabilities, they can be used to elicit the mixture coefficient α. Under analysis using the raw data, the median α across subjects is 0.80 (Table C.19, Appendix C.3); a Bayesian estimation of the α in tandem with the lower and upper CDFs (see Appendix C.2, Table C.17a) yields mean value 0.81. As discussed at more length in Section 5, this is, to our knowledge, the first direct choice-based elicitation of the α in the α-maxmin EU model that fully controls for the set of priors by eliciting the relevant information about them without invoking supplementary assumptions. Moreover, it is consistent with the findings for stated probabilities in EXP A: for $\alpha > 0.5$, the MP of an event is below the midpoint of the corresponding probability interval.

5. Discussion

Several general conclusions emerge from our implementation of the proposed multiple-prior elicitation method over a range of different sources of uncertainty. The first concerns its feasibility and validity. The method produces probability intervals that are reasonable: they are sensitive to the aspects of available information (in experimental contexts where that is controlled), and in general consistent with 'objective' (Bayesian update) probabilities where available. The second concerns the extent and determinants of imprecision. Although precise probabilities are permitted by our procedure, many subjects' elicited probability intervals are imprecise – they do not reduce to a single probability value – for the events considered here. Crucially, we recover an expected relationship between imprecision and perceived information in both controlled artificial sources and natural ones. This can be seen as providing further indirect

[19]More precisely, recalling that the stated probability task was not incentivised, and that it was completed on a one-cursor slider (Section 3.2), in just under 10% of cases, subjects provided stated probabilities that were very close to the default slider setting of 0. The reported proportion removes all responses below 0.1 (noting that the lowest observed frequency is 0.25).

[20]Moreover, stated probabilities were typically close to, yet below, the midpoints of the corresponding elicited intervals: in the mean, stated probabilities were 0.05 less than the midpoints.

evidence for the solidity of the proposed elicitation method.

Our third general conclusion concerns the comparison of our method with probability intervals stated directly by subjects. Several studies have compared different methods for precise probability elicitation in laboratory settings (e.g. Trautmann and Kuilen, 2015; Hollard et al., 2016). These could be used to ground and improve belief elicitation beyond the lab. After all, if a simpler method yields similar results to a more complex one with better properties – in terms of incentive compatibility, for instance – then such comparisons can bolster confidence in the use of the former method in the field. Our between-subject comparison in EXP A between the proposed elicitation method and stated probability intervals is the first such exercise for multiple prior or imprecise probability beliefs, to our knowledge (see also the related literature discussion below). It suggests that, in aggregate, stated probability-interval methods give a fairly good approximation to the intervals provided by our choice-based method, though they may overestimate the extent of imprecision in information-rich environments. Whilst this is a first study, and others are doubtless required, this bodes well for the use of stated probability-interval methods in the field (e.g. Giustinelli et al., 2022), as well as for the solidity of their conclusions.

Finally, as concerns the well-known identification problem for the α-maxmin EU model, we draw on our probability-interval elicitation to perform the first elicitation of the model's mixture coefficient that is fully general and controls for beliefs (see discussion below).

We now discuss the robustness of our procedure, related literature, and some directions for future development.

Robustness. Although Hurwicz α-maxmin EU is one of most general decision models in the literature taking as belief component a set of priors, the core of our method applies beyond this model. As set out in Sections 2.3–2.5, it operates on a general model (Eq. (1)) that can incorporate probability weighting in the style of Prospect Theory (Wakker, 2010) or transformations of probability intervals in the style of Gajdos et al. (2008). More precisely, the notion of MPI remains well-defined for all such extensions, and the decision maker's subjective probability interval is always a MPI (Section 2.4). Moreover, MPIs are essentially unique for generic cases of such extensions (Appendix B). The incentivization mechanism, which is implemented in the 2D choice list part of the method, is incentive compatible under the most general form of model (1), as evidenced by Proposition 2 and the accompanying discussion in Section 2.5. So it applies not only under α-maxmin EU, but also under the probability weighting or probability-interval transformation generalizations just mentioned. As concerns the first, binary-choice part of our method, although Proposition 1 underlying it is stated under α-maxmin EU with $\alpha > \frac{1}{2}$, it is actually a corollary of a stronger result (Proposition A.1, Appendix A). This result covers the aforementioned generalizations under conditions analogous to $\alpha > \frac{1}{2}$ (Appendix B). Moreover, as discussed in Appendix B, there is independent evidence that such conditions hold for most of our subjects. Note that the generality of the second, 2D choice list part of the procedure is more important, for this is the part that counts for incentivizing subjects' responses (Sections 2.5 and 3.4). Finally, the method also applies under decision models that do not belong to the family (1), such as the multiple-prior minimax expected regret model (Appendix B, footnote 1).

The 2D choice list mechanism is incentive compatible in the sense of weak dominance (Corollary 1, Section 2.5). This demands, for the reported interval-valued urn $[r, 1-b]$, that, for every choice between the bet on red from an urn $[x, 1-b]$ and the bet on the event E, the subject prefers the option selected by the incentive mechanism; and similarly for choices between bets on blue from urns $[r, y]$, for varying y, and the bet against E. The set of choices between the bet on the event and bets on urns with varying minimal numbers of red balls forms a branch of the 2D choice list, and can itself be thought of as a (standard one-dimensional) choice list. The weak-dominance notion of incentive compatibility focuses on the choices in this list in isolation from the way the number of blue balls b is set; and similarly for the other branch. Our implementation was designed to favor such isolation, notably via the realization of 2D choice lists by a single scrollbar with two cursors (Figure 2, Section 3.2 and Appendix D.2). Visually very different from Figure 1, this presentation is less suggestive of opportunities for strategically reporting the interval to influence the set of choices used for remuneration. Notwithstanding this, the extent to which such strategic reasoning has been employed by the subjects in our experiments is ultimately an empirical question, and we treat it as such. On this front, our elicitation method has the advantage that such reasoning leads to easily recognizable choice patterns. As discussed in Appendix B.2, for a subject represented by (2) with $\alpha \in (0,1)$ and any set of priors, her optimal response to the 2D choice list task when reasoning strategically[21] is one of the intervals $[0,0], [0,1], [1,1]$. However, no subjects gave such responses for all events elicited, with only one subject across all three experiments giving such an interval for more than half of the elicited events (Table C.6, Appendix C.1; see Appendix B.2 for further details). This suggests that strategic reasoning is extremely infrequent among our subjects.

[21] As set out in the cited Appendix, under strategic reasoning, the subject considers the choice of MPI as a choice of a (second-order) lottery over ILs and particular bets for or against E.

Related literature. Our elicitation method relates to existing experimental and theoretical literature on multiple prior models, and the α-maxmin EU model in particular. Part of this literature is concerned with testing or comparing such models (e.g. Hey et al., 2010; Baillon and Bleichrodt, 2015); by contrast, the aim here is to elicit probability intervals in the context of a general multiple prior model. Likewise, there is a literature using matching probabilities or certainty equivalents to study willingness to bet on objectively-given probability intervals (e.g. Baillon et al., 2012; Chew et al., 2017) . The present paper, by contrast, uses such interval-valued urns with the distinct aim of eliciting subjective probability intervals.

Multiple prior beliefs are most relevant in situations where agents typically do not hold precise probability distributions determining preferences, so to be generally applicable, a multiple-prior elicitation method should avoid assuming underlying precise probabilities. The assumption that subjects have precise probabilistic beliefs which completely determine the contributions of events to their (potentially non-expected utility) preferences is called *probabilistic sophistication* (Machina and Schmeidler, 1992; Chew and Sagi, 2006). As emphasized in the Introduction, our method avoids all assumptions of this sort. This arguably sets it apart from much of the theoretical literature and virtually all of the experimental literature on multiple prior models.

On the theory side, the challenge of incentive-compatible elicitation under α-maxmin EU (2) is compounded by identification issues, arising from the fact that different pairs of mixture coefficient α and sets of priors can represent the same preferences (see Introduction and Section 2.4). Proposed approaches include pinning down the set of priors using 'unambiguous preferences' (Ghirardato et al., 2004), though this has problems in finite state spaces (Eichberger et al., 2011), or enrichening the state space to include an infinite product structure and invoking symmetry axioms (Klibanoff et al., 2021). Another line of attack concentrates on special cases where the set of priors is generated by a precise probability distribution. For instance, Gul and Pesendorfer (2014, 2015) and Chateauneuf et al. (2007) obtain a unique identification of α and the set of priors: the former when the set is generated as extensions of a precise probability measure on a subalgebra of events; the latter when it is generated from a precise probability measure via ε-contamination, i.e. the mixture with the set of all probability measures.[22] Since in both cases, preferences and sets of priors are generated from precise probabilities, they assume some form of probabilistic sophistication. Our approach, by contrast, deliberately eschews such assumptions as inadmissible in many situations of interest. Rather, it follows

[22]Formally, the assumption is that the set of priors $\mathscr{C} = \{(1-\varepsilon)p + \varepsilon q : q \in \Delta\}$, where Δ is the space of all probability measures, p is an element of Δ and $\varepsilon \in [0,1]$.

the theoretical approach developed by Hill (2023), who resolves the identification issue for α-maxmin EU and a range of extensions by using interval lotteries, with no need for specific richness assumptions on the state space, probabilistic sophistication, or any other non-standard assumptions on the set of priors.

On the experimental front, there is a small literature dedicated to incentive-compatible elicitation of multiple priors. One family of approaches purports to elicit them as the support of second-order beliefs, represented as a probability measure over the space of probability measures. Beyond the assumption of second-order probabilities, which is foreign to the original multiple prior models (Gilboa and Schmeidler, 1989; Bewley, 2002; Ghirardato et al., 2004), and the fact that they import an assumption of probabilistic sophistication, albeit at the second-order level, these often make further assumptions about the role of these second-order beliefs in choice. For instance, Qiu and Weitzel (2016) propose a method that relies on the assumption that a subject's opinions about others subjects' matching probabilities coincides with the uncertainty surrounding her own assessment.

Another family of approaches draws on the probabilistically-sophisticated special case of α-maxmin EU studied by Chateauneuf et al. (2007), where the subject's set of priors is generated as the ε-contamination of a single probability measure. Dimmock et al. (2015); Baillon et al. (2018b,a) use elicitation of MPs or certainty equivalents to estimate 'ambiguity indices', which they claim can be used to back out the mixture coefficient α and the set of priors. However, as shown by Baillon et al. (2021, Theorem 16 & Section 7.3, Eq. (20)), these indices are only guaranteed to yield the subject's set of priors if they are generated from a precise probability measure by ε-contamination, in which case preferences are represented by the Chateauneuf et al. (2007) model. So, though unwarranted in situations where multiple prior decision models come to the fore, this elicitation technique assumes probabilistic sophistication. In fact, our data provides empirical insight into the relevance of their assumption. Whilst Chateauneuf et al. (2007) implies that the imprecision (in the sense of Section 4.2) is the same for all events,[23] our observations reject this equality for the sources in EXP N2 (Section 4.2; see also Table C.9 and Figure C.1, Appendix C.1): these are thus sources for which their method's underlying assumption does not hold. Of course, this does not bode well for the general applicability of their method. It does not follow that it is never viable; indeed, our data indicate that the Paris source in EXP N1 may satisfy their assumptions. Moreover, we can estimate the ambiguity indices used in the aforementioned papers on the basis of the data from our study (EXP N1,

[23] If the set of priors is as defined in footnote 22, then, for any E, $(1-\varepsilon)p(E) \in [0, 1-\varepsilon]$, so the probability interval for event E is $[(1-\varepsilon)p(E), (1-\varepsilon)p(E)+\varepsilon]$, and hence the event has imprecision ε.

Paris treatment) and under their assumption about the set of priors;[24] doing so, we find, for instance, that they yield the value 0.82 for the mixture coefficient α—which, reassuringly, is close to the Bayesian and raw estimates reported in Section 4.4. So our elicitation method is not only more general and robust, insofar as it applies in situations where the assumptions underlying their approach do not hold; moreover, it can evaluate precisely in which cases they do hold. In those cases, their approach, implemented on our data, gives the same result as our 'direct' elicitation.

Another related branch of literature focuses on scoring rules. Hossain and Okui (2013) provides a scoring rule in the absence of expected utility preferences: since it elicits precise probabilities under probabilistic sophistication, it does not tackle the issue of multiple-prior elicitation. Scoring rules have also been proposed for most-likely intervals for the value of an unknown parameter (Winkler and Murphy, 1979; Schlag, 2015). Typically, they are incentive compatible under the assumption that the subject is a Subjective Expected Utility maximiser with a precise probability distribution (Schlag, 2015, Section 5), and hence under an assumption stronger than probabilistic sophistication. By applying them where the unknown parameter at issue is itself a probability, such scoring rules could conceivably be repurposed to elicit probability intervals. However, given that they rely on the assumption of expected utility—here at the second-order level—they would need to suppose precise probabilities in order to elicit imprecise ones; as noted, this seems inappropriate for situations where multiple priors are relevant. As mentioned in the Introduction, the difficulty of developing scoring rules that avoid such probabilistic assumptions is further underlined by an impossibility result showing that there are no real-valued continuous strictly proper scoring rules for multiple-prior probability intervals (Seidenfeld et al., 2012, Prop 5).

Going beyond the lab, there is a large and growing literature on elicitation of multiple priors or imprecise probabilities in a range of disciplines, from economics to climate science. All such elicitation exercises of which we are aware use stated probability intervals, and as such are not incentive compatible. For instance, Giustinelli et al. (2022) elicit beliefs on dementia and long-term care decisions in a large-scale representative survey (over 1000 subjects), allowing stated probabilities to be interval-valued. Consistently with our results (Section 4.2), they find

[24] Specifically, Baillon et al. (2018b) propose the average of $1 - MP(E) - MP(E^c)$ over a selection of events as their measure of the 'ambiguity aversion index' b. The average for the events elicited here can be deduced directly from Table C.20 (Appendix C.4), as around 0.16. On the other hand, under (2) with the specified form for the set of priors (see footnote 22), their 'a-insensitivity index' $a = \varepsilon$. Under such sets of priors, as noted in footnote 23, every E has imprecision ε. The Average Imprecision measured by our method (Section 4.2 and Table C.7) thus gives an estimate of their a: it is around 0.25. The mixture coefficient α is related to these indices by $\alpha = \frac{1}{2}\left(\frac{b}{a}+1\right)$ (Baillon et al., 2021), yielding the value in the text.

widespread imprecision. They argue forcefully for the importance of probability-interval elicitation for reducing survey bias and understanding attitudes to and behavior in the face of high-uncertainty events, such as whether one will develop dementia and whether to insure against it. In another approach, in different domain, Kriegler et al. (2009) elicit beliefs of selected scientists (around 50 subjects) concerning climate tipping points, allowing participants to state probability intervals for these (notoriously uncertain) events. Such expert elicitations, which involve often time-consuming and individualised sessions with selected experts, have emerged as a central tool for managing complex uncertainties (Morgan, 2014). Though they have traditionally aimed at eliciting precise probabilities, Kriegler et al. (2009) shows that imprecision is widespread for some events, which once again argues for the relevance of probability-interval elicitation.

Future Directions Our method can shed some much-needed light on the criticisms of stated approaches centred on their lack of incentive compatibility and theoretical grounding. The preliminary comparison from EXP A shows that, in the aggregate, the stated approach yields similar results to our incentive-compatible decision-theoretically-well-founded method. As reported in Section 4.3, beyond this general match, there is a significant correlation in the Average Imprecision between the two methods across subjects. This suggests that, roughly, subjects with larger intervals as elicited by our method will tend to provide larger intervals in the stated task. Our comparison thus arguably provides justification for certain uses of stated elicitation: results found using stated methods that bear on mean imprecision or on tendencies across subjects promise to hold up under our more theoretically rigorous method. Other results concerning the comparison—for instance, the fact that stated intervals are considerably wider than those elicited by our method in information-rich situations (Section 4.3)—flag potential limits. If the aim is to study absolute amounts of imprecision in contexts where there is plenty of information, perhaps stated probability intervals are not a sufficiently robust tool.

This suggests one direction for future research. As noted, stated probability intervals are typically used in large-scale surveys (such as Giustinelli et al. 2022). The sorts of comparisons conducted in lab settings in EXP A shed light on their performance, and in particular the performance loss with respect to incentive-compatible methods for specific research questions. As such, they provide indications of expected performance in the field. Further research can expand our comparison, by identifying more precisely the sorts of characteristics of intervals where stated methods fair well, by extending the comparison to natural (as opposed to artificial) sources of uncertainty, or by mapping the performance of different refinements of the stated approach. A properly grounded probability-interval elicitation method, of the sort developed in this paper, can serve as a tool for designing and evaluating simpler methods for use in large-scale studies.

Moreover, although our method was developed with the aim of demonstrating the possibility and feasibility of choice-based incentive-compatible probability-interval elicitation and investigating some basic characteristics of subjective probability intervals, future research could operationalise simpler, parametrised versions, with fewer choice questions. Such versions could be more implementable, for instance in field studies. Some large-scale surveys use choice tasks without necessarily incentivising them (e.g. Falk et al., 2018), and questions formulated in terms of bets may trigger different cognitive mechanisms to those formulated in terms of stated judgements. Our method could thus lay the foundations of a bet-based approach to add to the arsenal of probability-interval elicitation procedures used in practice.

Finally, analogous possibilities exist for expert elicitation exercises (Kriegler et al., 2009; Morgan, 2014). Compared to survey studies, these typically involve fewer subjects (experts), with each spending more time; accordingly, more precision is desired of the elicitation at the individual level. Aggregate-level performance of an elicitation method—of the sort suggested for stated methods by the results in Section 4.3—is less relevant for such exercises. Our experiments suggest the promise of our method to provide individual-level probability-interval elicitation with theoretically well-founded incentive-compatibility properties. Probability elicitation exercises in decision analysis often use bet-based choice tasks without necessarily incentivising them (e.g. Clemen and Reilly, 2013); again, our method, applied in this context, complements existing stated approaches to eliciting probability intervals.

6. Conclusion

This paper proposes and implements a solution to the open problem of choice-based incentive-compatible elicitation of multiple prior beliefs. It comprises a new preference-based notion—Matching Probability Intervals—and a probability-interval analogue of a state-of-the-art elicitation procedure for matching probabilities. Our elicitation operates under the Hurwicz α-maxmin EU model as well as a range of generalizations, and in the absence of strong assumptions about subjects' sets of priors, most notably any form of probabilistic sophistication.

Our implementation of the elicitation method, in three experiments to elicit subjective probability intervals and upper and lower CDFs over artificial and natural sources of uncertainty, testifies to its validity and feasibility. It finds a predominance of imprecision—intervals of non-zero width—across our subjects, for all explored sources, showing it to be related to information, familiarity or predictability. It also compares our choice-based elicitation with stated probability-interval methods, showing that they yield similar results in aggregate. Our method also allows us to perform what, to our knowledge, is the first elicitation of the mixture coefficient in the α-maxmin EU model that fully controls for beliefs.

Acknowledgements

We would like to thank Itzhak Gilboa, Simon Grant, Charles Manski and Peter Wakker, as well as audiences at D-TEA, RUD and SAET, for stimulating discussion and excellent feedback on early versions of this paper. We thank the Université Paris 1 Experimental Economics lab, the INSEAD-Sorbonne Behavioral Lab, Raffael Grasberger, Pierre-Antoine Amiand-Leroy, Maxime Frolov, Nawel Ayadi and Esso-Wazam Tchalanga for research assistance.

Funding

This research is supported by grants of the French National Research Agency (ANR), "Investissements d'Avenir" (LabEx Ecodec/ANR-11-LABX-0047) (ANR) and the ANR project DUSUCA (ANR-14-CE29-0003-01). We gratefully acknowledge the support of Hi! Paris.

References

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

Abdellaoui, M., H. Bleichrodt, and C. Gutierrez (2023): "Unpacking Overconfident Behavior When Betting on Oneself," *Management Science*.

Abdellaoui, M., H. Bleichrodt, and O. LâHaridon (2008): "A tractable method to measure utility and loss aversion under prospect theory," *Journal of Risk and Uncertainty*, 36, 245.

Abdellaoui, M., E. Diecidue, E. Kemel, and A. Onculer (2021): "Temporal Risk Resolution: Utility versus Probability Weighting Approaches," *Management Science*.

Anscombe, F. J. and R. J. Aumann (1963): "A Definition of Subjective Probability," *The Annals of Mathematical Statistics*, 34, 199–205.

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

BAILLON, A., H. BLEICHRODT, U. KESKIN, O. LÂHARIDON, AND C. LI (2018a): "The Effect of Learning on Ambiguity Attitudes," *Management Science*, 64, 2181–2198.

BAILLON, A., H. BLEICHRODT, C. LI, AND P. P. WAKKER (2021): "Belief hedges: Measuring ambiguity for all events and all models," *Journal of Economic Theory*, 198, 105353.

BAILLON, A., L. CABANTOUS, AND P. P. WAKKER (2012): "Aggregating imprecise or conflicting beliefs: An experimental investigation using modern ambiguity theories," *Journal of Risk and Uncertainty*, 44, 115–147.

BAILLON, A., Z. HUANG, A. SELIM, AND P. P. WAKKER (2018b): "Measuring ambiguity attitudes for all (natural) events," *Econometrica*, 86, 1839–1858.

BERGER, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis*, New York: Springer.

BERGER, J. O., D. R. INSUA, AND F. RUGGERI (2000): "Bayesian robustness," *Robust Bayesian Analysis*, 1–32.

BEWLEY, T. F. (2002): "Knightian decision theory. Part I," *Decisions in Economics and Finance*, 25, 79–110.

BOREL, E. (1939): "Valeur pratique et philosophie des probabilitÃ©s.[Practical and philosophical value of probability]," *TraitÃ© du calcul des probabilitÃ©s et de ses applications [Treatise of probability calculus and its applications]*, 4.

BOSE, S. AND L. RENOU (2014): "Mechanism design with ambiguous communication devices," *Econometrica*, 82, 1853–1872.

CARNEY, M., JON CUNLIFFE, DAVE RAMSDEN, ANDREW HALDANE, JONATHAN HASKEL, MICHAEL SAUNDERS, SILVANA TENREYRO, AND GERTJAN VLIEGHE (2019): "Inflation Report," *Bank of England*.

CERRONI, S. (2020): "Eliciting farmersâ subjective probabilities, risk, and uncertainty preferences using contextualized field experiments," *Agricultural Economics*, 51, 707–724, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/agec.12587.

CHATEAUNEUF, A., J. EICHBERGER, AND S. GRANT (2007): "Choice under uncertainty with the best and worst in mind: Neo-additive capacities," *Journal of Economic Theory*, 137, 538–567.

CHEW AND J. S. SAGI (2006): "Event exchangeability: Probabilistic sophistication without continuity or monotonicity," *Econometrica*, 74, 771–786.

CHEW, S. H., B. MIAO, AND S. ZHONG (2017): "Partial ambiguity," *Econometrica*, 85, 1239–1260, publisher: Wiley Online Library.

Clemen, R. T. and T. Reilly (2013): *Making Hard Decisions with Decision Tools*, Cengage Learning.

Delavande, A., X. GinÃ©, and D. McKenzie (2011): “Measuring subjective expectations in developing countries: A critical review and new evidence,” *Journal of development economics*, 94, 151–163, publisher: Elsevier.

Dimmock, S. G., R. Kouwenberg, and P. P. Wakker (2015): “Ambiguity attitudes in a large representative sample,” *Management Science*.

Dow, J. and S. R. da Costa Werlang (1992): “Uncertainty aversion, risk aversion, and the optimal choice of portfolio,” *Econometrica: Journal of the Econometric Society*, 197–204, publisher: JSTOR.

Eichberger, J., S. Grant, D. Kelsey, and G. A. Koshevoy (2011): “The alpha-MEU model: A comment,” *Journal of Economic Theory*, 146, 1684–1698.

Ellsberg, D. (1961): “Risk, Ambiguity, and the Savage Axioms,” *Quart. J. Econ.*, 75, 643–669.

Engelberg, J., C. F. Manski, and J. Williams (2011): “Assessing the temporal variation of macroeconomic forecasts by a panel of changing composition,” *Journal of Applied Econometrics*, 26, 1059–1078, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.1206.

Epstein, L. G. and M. Schneider (2010): “Ambiguity and Asset Markets,” *Annual Review of Financial Economics*, 2, 315–346.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018): “Global evidence on economic preferences,” *The Quarterly Journal of Economics*, 133, 1645–1692.

Gajdos, T., T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud (2008): “Attitude toward imprecise information,” *J. Econ. Theory*, 140, 27 – 65.

Garlappi, L., R. Uppal, and T. Wang (2007): “Portfolio selection with parameter and model uncertainty: A multi-prior approach,” *Review of Financial Studies*, 20, 41–81.

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004): “Differentiating ambiguity and ambiguity attitude,” *J. Econ. Theory*, 118, 133–173.

Gilboa, I. and D. Schmeidler (1989): “Maxmin expected utility with non-unique prior,” *J. Math. Econ.*, 18, 141–153.

Giustinelli, P., C. F. Manski, and F. Molinari (2022): “Precise or Imprecise Probabilities? Evidence from Survey Response on Late-onset Dementia,” *Journal of the European Economic Association*, 20.

Grant, S., P. Rich, and J. D. Stecher (2019): “An Ordinal Theory of Worst- and Best-Case Expected Utility,” SSRN Scholarly Paper ID 3369078, Social Science Research Network, Rochester, NY.

Guiso, L. and G. Parigi (1999): “Investment and demand uncertainty,” *The Quarterly Journal of Economics*, 114, 185–227, publisher: MIT Press.

Gul, F. and W. Pesendorfer (2014): “Expected uncertain utility theory,” *Econometrica*, 82, 1–39.

——— (2015): “Hurwicz expected utility and subjective sources,” *Journal of Economic Theory*, 159, 465–488.

Hey, J., G. Lotito, and A. Maffioletti (2010): “The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity,” *Journal of Risk and Uncertainty*, 41, 81–111.

Hill, B. (2023): “Beyond uncertainty aversion,” *Games and Economic Behavior*, 141, 196–222.

——— (2024): “Prices vs Quantities under Severe Uncertainty,” SSRN 4826405, HEC Paris.

Hoffman, M. D. and A. Gelman (2014): “The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.” *J. Mach. Learn. Res.*, 15, 1593–1623.

Hollard, G., S. Massoni, and J.-C. Vergnaud (2016): “In search of good probability assessors: an experimental comparison of elicitation rules for confidence judgments,” *Theory and Decision*, 80, 363–387, publisher: Springer.

Hossain, T. and R. Okui (2013): “The binarized scoring rule,” *Review of Economic Studies*, 80, 984–1001.

Ilut, C. L. and M. Schneider (2014): “Ambiguous business cycles,” *American Economic Review*, 104, 2368–99.

Karanki, D. R., H. S. Kushwaha, A. K. Verma, and S. Ajit (2009): “Uncertainty Analysis Based on Probability Bounds (P-Box) Approach in Probabilistic Safety Assessment,” *Risk Analysis*, 29, 662–675.

Klibanoff, P., S. Mukerji, K. Seo, and L. Stanca (2021): “Foundations of ambiguity models under symmetry: alpha-MEU and smooth ambiguity,” *Journal of Economic Theory*, 105202.

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

Kriegler, E., J. W. Hall, H. Held, R. Dawson, and H. J. Schellnhuber (2009): “Imprecise probability assessment of tipping points in the climate system,” *Proceedings of the national Academy of Sciences*, 106, 5041–5046.

Machina, M. J. and D. Schmeidler (1992): “A More Robust Definition of Subjective Probability,” *Econometrica*, 60, 745–780.

Manski, C. F. (2004): “Measuring expectations,” *Econometrica*, 72, 1329–1376, publisher: Wiley Online Library.

MARINACCI, M. (2015): "Model uncertainty," *Journal of the European Economic Association*, 13, 1022–1100.

MASTRANDREA, M. D., C. B. FIELD, T. F. STOCKER, O. EDENHOFER, K. L. EBI, D. J. FRAME, H. HELD, E. KRIEGLER, K. J. MACH, AND P. R. MATSCHOSS (2010): "Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties," *Intergovernmental Panel on Climate Change (IPCC)*.

MORGAN, M. G. (2014): "Use (and abuse) of expert elicitation in support of decision making for public policy," *Proceedings of the National Academy of Sciences*, 111, 7176–7184.

QIU, J. AND U. WEITZEL (2016): "Experimental evidence on valuation with multiple priors," *Journal of risk and uncertainty*, 53, 55–74.

SAVAGE, L. J. (1954): *The Foundations of Statistics*, New York: Dover, 2nd edn 1971.

SCHLAG, K. H. (2015): "A method to elicit beliefs as most likely intervals," *Judgment and Decision Making*, 10, 456.

SEIDENFELD, T., M. J. SCHERVISH, AND J. B. KADANE (2012): "Forecasting with imprecise probabilities," *International Journal of Approximate Reasoning*, 53, 1248–1261.

STOYE, J. (2011): "Axioms for minimax regret choice correspondences," *Journal of Economic Theory*, 146, 2226–2251.

TRAUTMANN, S. T. AND G. KUILEN (2015): "Belief elicitation: A horse race among truth serums," *The Economic Journal*, 125, 2116–2135.

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

WALLEY, P. (1991): *Statistical reasoning with imprecise probabilities*, London: Chapman and Hall.

WINKLER, R. L. AND A. H. MURPHY (1979): "Use of Probabilities in Forecasts of Maximum and Minimum Temperatures," *Meteorological Magazine*, 108, 317–329.

A. Proofs

For differentiable W as in representation (1), let $\partial_1 W([x,y])$ denote the partial derivative of W with respect to the first coordinate, x, at $[x,y]$, and similarly for $\partial_2 W([x,y])$ and the second coordinate.

Proposition 1 is a corollary of the following Proposition, the uniqueness of the MPI (Proposition A.2 below), and the fact that (2) corresponds to a case of (1) where W is differentiable, $\partial_1 W([x,y]) = \alpha$ and $\partial_2 W([x,y]) = 1 - \alpha$.

Proposition A.1. *Let E be an event, and suppose preferences are represented according to (1) with a unique MPI for E and W differentiable with $\partial_1 W([\underline{p}(E), \overline{p}(E)]) > \partial_2 W([\underline{p}(E), \overline{p}(E)])$.*

a. For any $[x,y]$ in the R-B region (i.e. such that the corresponding preferences in Table 1 hold, for E), $\underline{p}(E) \leq x$ and $\overline{p}(E) \geq y$. Moreover, for any $[x,y]$ in the Wh region, $\underline{p}(E) \geq x$ and $\overline{p}(E) \leq y$.

b. For any $[x,y]$ in the R region (i.e. such that the corresponding preferences in Table 1 hold, for E), every $[x',y']$ with $x' \geq x$ and $y' \geq y$ is also in R. Moreover, for any $[x,y]$ in the B region, every $[x',y']$ with $x' \leq x$ and $y' \leq y$ is also in B.

Proof. Part b. follows directly from the fact that, under (1), given that W is increasing in both bounds, whenever $x \leq x'$ and $y \leq y'$, then $(z,[x,y],0) \preceq (z,[x',y'],0)$ and $(0,[x,y],z) \succeq (0,[x',y'],z)$.

As concerns part a., since W is increasing in both bounds, if $[x,y]$ is such that $x < \underline{p}(E)$ and $y < \overline{p}(E)$, then $(z,[x,y],0) \preceq (z,E,0)$ and $(0,[x,y],z) \succeq (0,E,z)$, so $[x,y]$ is in the B region. Similarly, for $[x,y]$ such that $x > \underline{p}(E)$ and $y > \overline{p}(E)$, $[x,y]$ is in the R region.

For each x such that there exists y with $(0,[x,y],z) \sim (0,E,z)$, let $J(x)$ be y such that this indifference holds. By Lower Stochastic Dominance, i.e. the fact that W is strictly increasing in the first coordinate (footnote 10), there is a unique $J(x)$ for all such x. By construction $\{[x,y] : (0,[x,y],z) \sim (0,E,z)\} = \{[x,J(x)]\}$. This set, which we call Ind_B, is the indifference curve for $[\underline{p}(E), \overline{p}(E)]$ in the space of bets on blue. Note that, since W is differentiable, so is J with $\frac{dJ}{dx}(x) = -\frac{\partial_2 W([x,J(x)])}{\partial_1 W([x,J(x)])}$ on its domain.

For each y such that there exists x with $(z,[x,y],0) \sim (z,E,0)$, let $I(y)$ be x such that this indifference holds. By Lower Stochastic Dominance, i.e. the fact that W is strictly increasing in the first coordinate (footnote 10), there is a unique $I(y)$ for all such y. By construction $\{[x,y] : (z,[x,y],0) \sim (z,E,0)\} = \{[I(y),y]\}$. This set, which we call Ind_R, is the indifference curve for $[\underline{p}(E), \overline{p}(E)]$ in the space of bets on red. Note that, since W is differentiable, so is I with $\frac{dI}{dy}(y) = -\frac{\partial_2 W([I(y),y])}{\partial_1 W([I(y),y])}$ on its domain.

Since there is a unique MPI, there exists a unique $[x,y]$ at the intersection of the two indifference curves; i.e. a unique $[x,y]$ with $y = J(x)$ and $x = I(J(x))$. For any sufficiently small $dx > 0$, $[x+dx, J(x)+J'(x)dx]$ belongs to the blue indifference curve Ind_B. Similarly, $[x+I'(J(x))J'(x)dx, J(x)+J'(x)dx]$ belongs to the red indifference curve Ind_R. Hence, for $x+dx$ with small $dx > 0$, the blue indifference curve Ind_B is 'above' the red indifference curve Ind_R (as in Figure 1) if and only if $I'(J(x))J'(x) < 1$. Substituting in the derivatives of I and J, this holds if and only if $\partial_1 W([x,y]) > \partial_2 W([x,y])$. Since the MPI is unique, it follows that the blue indifference curve is 'above' the red indifference curve for all $x' > x$. By similar reasoning, the blue indifference curve is 'below' the red one for all $x' < x$. The result follows from the fact that $[\underline{p}(E), \overline{p}(E)]$ is the MPI, the previously noted fact about points to the South-West and

North-East of $[\underline{p}(E), \overline{p}(E)]$, and the definition of the R-B region (respectively Wh region) in Table 1 as those urns 'below' the blue indifference curve and 'above' the red (resp. 'above' the blue one and 'below' the red one).

□

Proof of Proposition 2. Under (1), it follows from the first preference pattern in Proposition 2 that $W([q, 1-b]) > W([\underline{p}(E), \overline{p}(E)])$ for all $q > r$ and $W([q, 1-b]) < W([\underline{p}(E), \overline{p}(E)])$ for all $q < r$, and similarly for the others. By the continuity of W, it thus follows from the first two preferences that $W([r, 1-b) = W([\underline{p}(E), \overline{p}(E)])$, and from the second pair of preferences that $W([b, 1-r]) = W([1-\overline{p}(E), 1-\underline{p}(E)])$. It thus follows that $(z, [r, 1-b], 0) \sim (z, E, 0)$ and $(0, [r, 1-b], z) \sim (0, E, z)$, so $[r, 1-b]$ is a MPI for E, as required. The converse direction is an immediate consequence of the fact that W is strictly increasing in the lower bound. □

Finally, we state for completeness the result on the uniqueness of the MPI.

Proposition A.2. *For any decision maker represented according to (2) with $\alpha \neq \frac{1}{2}$, and for any event E, there is a unique MPI for E.*

Proof. Existence is immediate from Eqs. (4) and (5). Uniqueness is immediate from the linearity of the indifference curves in $\mathscr{I}$-space (see Figure 1). □

Online Appendix

B. Theoretical Appendix: Robustness of the method

In this Appendix, we discuss the robustness of the central elements of our proposal. Whilst we concentrate below on models of the general form (1), note that the method also applies under other multiple-prior decision models, most notably multiple-prior minimax (expected) regret.[1] We begin by discussing the robustness of the notion of MPI, before turning to the method for eliciting them.

B.1. Matching Probability Intervals

As noted in Section 2.4, under general preferences of the form (1), the notion of MPI is well defined, and the subjective probability interval is a MPI. However, uniqueness of the MPI is guaranteed only if there is a unique solution to the equations corresponding to the preferences (4) and (5), and this only occurs if W satisfies the following 'single-crossing property': every pair of red-and-blue indifference curves in Figure 1 cross at most once.[2] Whether this is the case, and how often it is not, will depend on the functional form of W. We thus consider what form of uniqueness holds for reasonable W.

For instance, the MPI is clearly unique when W is linear and non-symmetric[3]—and hence for α-maxmin EU whenever $\alpha \neq \frac{1}{2}$ (Proposition A.2). In Section 2.3, we mentioned two other more general interesting cases. One is when W incorporates probability weighting, e.g. is of the form $W([x,y]) = \alpha w(x) + (1-\alpha)w(y)$ for a (probability) weighting function w. As noted previously, this form can incorporate findings on probability weighting for (two-outcome) lotteries, via w. For such W, if w takes the neo-additive form often used in literature (Chateauneuf et al., 2007; Wakker, 2010; Baillon et al., 2021), then w is linear except at 0 and 1, so the previous observation implies that MPIs are unique. Moreover, even for non-linear weighting functions, calculation of relevant cases suggests that MPIs are typically unique. As an example, Figure B.1 plots red and blue indifference curves for the specified form of W with w being the popular Prelec weighting function with the parameters found by Abdellaoui et al. (2011) for a Paris temperature source (i.e. one that is similar to the source we used in EXP N1) and an α

[1] This model evaluates the choice of act f from a menu M according to $-\max_{p \in \mathscr{C}} \mathbb{E}_p (\max_{g \in M} u(g(s)) - u(f(s)))$, where $\mathbb{E}_p$ is the expectation with respect to probability measure p and $\mathscr{C}$ is the set of priors (e.g. Berger, 1985; Stoye, 2011). It is straightforward to show that for the choices used by our method—namely binary choices between bets on independent events, in the sense that the joint (multi-prior) distribution over the pair of relevant events is a 'type-1 product' (Walley, 1991, Sect. 9.3.5) of the multiple priors beliefs about each—preferences under this rule correspond to preferences under maxmin-EU (i.e. (2) with $\alpha = 1$) with the same set of priors.

[2] Technically, for every $A, B \in \mathbb{R}$, $|\{[x,y] \in \mathscr{I} : W([x,y]) = A, W([1-y, 1-x]) = B\}| \leq 1$.

[3] I.e. it is not the case that $W([x,y]) = W([y,x])$ for all $[x,y]$.

of 0.8 (i.e. close to the value we found for α; Section 4.4). Clearly, red and blue indifference curves typically only cross (at most) once, as required for uniqueness of MPI. Even in the cases where there are multiple MPIs, there will be at most two, with one close to the boundary.

Another interesting case is when W incorporates a transformation of probability intervals, i.e. $W([\underline{p},\overline{p}]) = \alpha\underline{\varphi([\underline{p},\overline{p}])} + (1-\alpha)\overline{\varphi([\underline{p},\overline{p}])}$ where α is as in (2) and $\varphi : \mathscr{I} \to \mathscr{I}$, with $\mathscr{I}$ the space of probability intervals, is a probability-interval transformation function. A canonical transformation φ would take a 'central point' of the interval and contract the interval around it, i.e. it would be of the form $\varphi([\underline{p},\overline{p}]) = \left[\varepsilon(\beta\underline{p}+(1-\beta)\overline{p})+(1-\varepsilon)\underline{p}, \varepsilon(\beta\underline{p}+(1-\beta)\overline{p})+(1-\varepsilon)\overline{p}\right]$, where β determines the 'central point' and ε encodes the extent of the contraction.[4] This is a generalisation of the contraction representation in Gajdos et al. (2008), in which 'contractions' of objectively provided sets of priors feature in a maxmin-EU decision rule. Such W is clearly linear and, for canonical α, ε and β, non-symmetric,[5] so the previous observation implies that MPIs are unique under such probability-interval transformation preferences.

In summary, for all generalisations of α-maxmin EU belonging to the general class defined in Section 2.3, MPIs are well-defined, and the subject's probability interval is always a MPI. Moreover, for reasonable extensions of various sorts, MPIs continue to be unique.

B.2. MPI Elicitation

Recall (Sections 2.5 and 3.2) that, following techniques developed for eliciting matching probabilities (Abdellaoui et al., 2021, 2023), we develop a two-step elicitation method for MPI elicitation. We discuss the robustness of the two steps in turn.

Binary-choice procedure The binary-choice step (Section 2.5 and Appendix D.1) is based on the division of space of probability intervals $\mathscr{I}$ into regions (Table 1) and Proposition 1 dictating 'where' the MPI is relative to points in the various regions. For decision makers represented according to the α-maxmin EU model (2), Proposition 1 a. only holds if $\alpha > \frac{1}{2}$. Proposition A.1 applies for the general decision model (1); in this sense, the binary-choice step of our elicitation method is robust to the decision model, but requires the equivalent of $\alpha > \frac{1}{2}$, as specified in the Proposition. Under the notable generalisations of α-maxmin EU discussed in Sections 2.3 and B.1—i.e. involving neo-additive probability weighting or a probability-interval transformation contracting the interval around the midpoint—it is straightforward to check that this condition reduces to $\alpha > \frac{1}{2}$. Hence, we can focus on this assumption underlying

[4] For instance, $\beta = \frac{1}{2}$ yields a contraction around the midpoint of the interval.

[5] More precisely, by basic algebra, W is non-symmetric whenever $\alpha \neq \frac{1-2\varepsilon\beta}{2(1-\varepsilon)}$, so for the 'contraction around the midpoint' case discussed in the previous footnote ($\beta = \frac{1}{2}$), W is unique whenever $\alpha \neq \frac{1}{2}$. The condition for uniqueness of MPI under this generalisation is thus the same as that under α-maxmin EU (Proposition A.2).

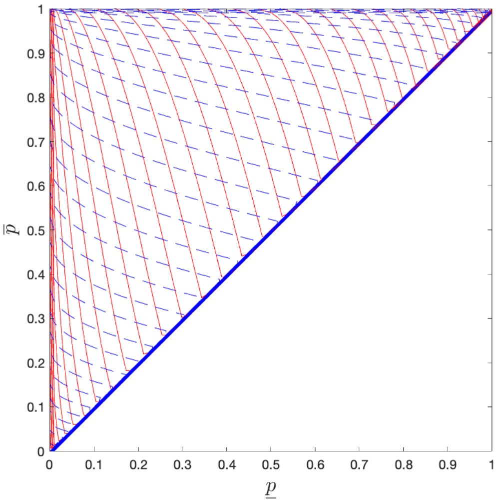


Figure B.1: Indifference curves in probability interval space $\mathscr{I}$ under (1) with $W(x,y) = \alpha w(x) + (1-\alpha)w(y)$.
Red lines: indifference curves for IL $(z,[p,q],0)$: i.e. curves of the form $\alpha w(x) + (1-\alpha)w(y) = C$. Blue lines: indifference curves for IL $(0,[p,q],z)$: i.e. curves of the form $\alpha w(1-y) + (1-\alpha)w(1-x) = D$. Parametrisation: Prelec weighting function $w(x) = \left(e^{-(-ln(x))^{\alpha}}\right)^{\beta}$ with $\alpha = 0.54$ and $\beta = 0.85$ (Abdellaoui et al., 2011); $\alpha = 0.8$.

the binary-choice procedure, in the knowledge that it is common to most decision models of interest.

Proposition 1 a. guarantees that if the elicited point (on Figure 1) is in the R-B region (respectively, Wh region), then the MPI is North-West (resp., South-East) of it. When $\alpha < \frac{1}{2}$, the opposite holds: e.g. the MPI is North-West of the elicited point not when it is in R-B, but when it is in Wh. So the procedure applied to such decision makers would 'move' in the wrong direction: e.g. instead of looking South-East for the MPI after finding a point in R-B, it would look North-West. When $\alpha = \frac{1}{2}$, the red and blue indifference curves in Figure 1 are parallel, so there will canonically be no points in the R-B and Wh regions. We now review evidence on the value of α for our subjects, as well as on the functioning of the procedure.

We find little evidence for widespread $\alpha \leq \frac{1}{2}$ among our subjects. First of all, the elicitation of α reported in Section 4.4 finds median and 25 percentile values significantly above $\frac{1}{2}$ (Table C.19), indicating that less than 25% of subjects have $\alpha \leq \frac{1}{2}$. Moreover, under the α-maxmin EU model, the sum of the MP of an event and that of its complement is less than (respectively, greater than) one precisely when $\alpha > \frac{1}{2}$ (resp. $\alpha < \frac{1}{2}$; see Appendix C.4); so we can use our matching probability data to check for the sign of $\alpha - \frac{1}{2}$. Table C.20 (Appendix C.4) displays the descriptive statistics on this sum for the Paris treatment where MPs were elicited, confirming again that $\alpha > \frac{1}{2}$ for over 75% of subjects.

As concerns the functioning of the procedure, since it 'moves' in the wrong direction for subjects with $\alpha < \frac{1}{2}$, no such subjects will pass through both points in Wh and in R-B. However, 383 applications of the procedure out of 704 in EXP N1 (300 out of 606 in EXP N2; 155 out of 299 in EXP A) passed through points in Wh and R-B. Whilst there were nevertheless applications which passed through points in R-B but not Wh (152 in EXP N1, 77 in EXP N2, 69 in EXP A) and in Wh but not R-B (114 in EXP N1, 105 in EXP N2, 72 in EXP A), these would be expected if the procedure functioned correctly and the probability intervals were large (respectively small). Moreover, for all subjects in all experiments, there was at least one event with a point in R-B or Wh, which is inconsistent with widespread $\alpha = \frac{1}{2}$ among subjects. The evidence thus does not support misfunctioning of the procedure.

The binary-choice procedure is only the first step of the elicitation method. Even if it does not work properly for some decision makers, they have the opportunity to correct their responses in the second, 2D choice list step. So the ultimate performance of the whole method depends more centrally on the validity of this step—an issue to which we now turn.

2D Choice List Procedure As described in Sections 2 and 3, the incentive compatibility of our elicitation method depends on the incentive compatibility of its second, 2D choice-list confirmation step. As set out in Section 2.5, this is guaranteed by Proposition 2, which applies under the fully general decision model (1), and not just under the α-maxmin EU special case. Hence the incentive compatibility of our method is robust across a range of multiple-prior decision models, including all those mentioned previously.

As discussed in Sections 2.5 and 5, this stage of the method is incentive compatible whenever subjects treat the two branches of the 2D choice list in isolation from each other. If, by contrast, a subject reasons strategically across the two branches of the 2D choice list, then the choice of MPI is conceptualised as the choice of a (second-order) lottery assigning a probability to playing a bet for or against E or to playing specific ILs according to the mechanism. Assuming the α-maximin EU model (2) at both levels, the subject evaluates each such second-order lottery using the expectation over the values of the bets and ILs. For any reported interval $[\underline{q}, \overline{q}]$ in this task, the incentive mechanism defined in Section 2.5 determines the probability of the bet or IL 'received',[6] which determines in turn the utility of reporting $[\underline{q}, \overline{q}]$ when the true beliefs are $[\underline{p}(E), \overline{p}(E)] = [\underline{p}, \overline{p}]$. Finding the optimum numerically for a grid of values of $\underline{p}, \overline{p}, \alpha \in [0,1]$ using Matlab, we find that, for every $(\underline{p}, \overline{p}, \alpha)$ (with $\overline{p} \geq \underline{p}$) except for $\underline{p} = 0,\ \overline{p} = 1,\ \alpha = 0$, and those with $\underline{p} = 0.5,\ \alpha = 1$ or $\overline{p} = 0.5,\ \alpha = 1$, the optimal response under this strategic reasoning is situated at one or several of the 'vertices' of the space of probability intervals in Figure 1, i.e. $[0,0], [0,1], [1,1]$. For $\underline{p} = 0,\ \overline{p} = 1,\ \alpha = 0$ and $\underline{p} = 0.5$, $\alpha = 1$ or $\overline{p} = 0.5$, $\alpha = 1$ with $\underline{p} \neq \overline{p}$, the optimum is situated at all points on one of the boundaries of the probability-interval space, i.e.

[6] Specifically, the probabilities of receiving the bet on E, the IL on red, the bet on E^c, the IL on blue are $\frac{\underline{q}}{\overline{q}+1-\underline{q}}$, $\frac{\overline{q}-\underline{q}}{\overline{q}+1-\underline{q}}$, $\frac{1-\overline{q}}{\overline{q}+1-\underline{q}}$, $\frac{\overline{q}-\underline{q}}{\overline{q}+1-\underline{q}}$, respectively.

$\{[0,y] : y \in [0,1]\}, \{[x,1] : x \in [0,1]\}, \{[x,y] : x \in [0,1], y = x\}$. When $\underline{p} = 0.5, \overline{p} = 0.5, \alpha = 1$, the utility above is constant, so all points maximise it.

It follows that, for any subject with $\alpha \in (0,1)$ who responds strategically in this manner, all her responses will be at a vertex of the space $\mathscr{I}$. Our elicitation of α suggests however that the vast majority of subjects have α in this range. Even for subjects with $\alpha = 0$ or 1 reasoning strategically, they will have more than one response in the interior of $\mathscr{I}$ only if $\alpha = 1$ and they assign precise probability of 0.5 to several elicited events. In EXPs N1 and N2, involving nested events, this would correspond to a peculiar (e.g. bimodal) distribution on the variable of interest (temperature, marks); in EXP A, involving events related to varying types of information (frequencies and samples sizes), it would indicate complete insensitivity to prior information, which contradicts the findings reported in Sections 4.1 and 4.2. As is clear from Table C.6 (Appendix C.1), no subjects give vertex responses for all elicited events, with only one subject (across all experiments) giving vertex responses for over half of the elicited events. Moreover, the vast majority of subjects (73 out of 80 in EXP N1; 51 out of 52 in EXP N2; 97 out of 101 in EXP A) gave more than one response in the interior of $\mathscr{I}$. The data thus clearly suggests that strategic reasoning is extremely infrequent in our sample.

C. Supplementary Statistics

Data and code for its analysis are available online here.[7]

C.1. Descriptive Statistics

Table C.1 reports the basic descriptive statistics on the upper (U) and lower (L) bounds of the elicited probability intervals, over the three experiments. Table C.2 reports the descriptive statistics of the upper and lower bounds of the stated probability intervals, for EXP A.

Tests for EXP A Tables C.3 and C.4 report tests that the midpoints (respectively Imprecision) is the same across bags with the same sample size and different frequencies (respectively the same frequency and different sample sizes) in EXP A. Throughout this Appendix, unless specified, all tests are two-sided.

Monotonicity in EXPs N1 and N2 Table C.5 reports the descriptive statistics for the individual level Kendall τ_b correlation coefficients between the size of events (i.e. the t for events E_t, as specified in Table 3) and the upper (resp. lower) probabilities or MPs elicited for each

[7] If the link does not work, the address is: `https://osf.io/yvax4/`.

(Sample size, frequency)	Mean		Std		Min		25%		50%		75%		Max	
	L	U	L	U	L	U	L	U	L	U	L	U	L	U
(4, 0.5)	0.22	0.77	0.14	0.15	0.02	0.43	0.08	0.65	0.18	0.80	0.35	0.92	0.50	0.98
(4, 0.25)	0.19	0.65	0.11	0.21	0.00	0.26	0.10	0.50	0.18	0.65	0.26	0.84	0.42	1.00
(20, 0.5)	0.28	0.73	0.16	0.14	0.08	0.48	0.12	0.58	0.26	0.74	0.42	0.88	0.58	0.92
(20, 0.25)	0.21	0.59	0.13	0.23	0.05	0.14	0.12	0.35	0.18	0.62	0.25	0.79	0.62	0.92
(100, 0.5)	0.42	0.57	0.07	0.06	0.26	0.44	0.36	0.55	0.45	0.55	0.45	0.59	0.55	0.74
(100, 0.25)	0.22	0.39	0.06	0.19	0.08	0.19	0.19	0.26	0.20	0.30	0.25	0.44	0.39	0.92

(a) Lower and upper probabilities, EXP A (Overall sample $n = 101$)

E_t	Mean		Std		Min		25%		50%		75%		Max	
	L	U	L	U	L	U	L	U	L	U	L	U	L	U
$t = -2$	0.29	0.55	0.17	0.21	-0.01	0.09	0.15	0.40	0.29	0.55	0.40	0.67	0.7	1.0
$t = 2$	0.38	0.65	0.22	0.19	0.00	0.23	0.20	0.51	0.35	0.65	0.50	0.80	1.0	1.0
$t = 5$	0.48	0.74	0.23	0.17	0.00	0.25	0.35	0.62	0.46	0.76	0.66	0.88	1.0	1.0
$t = 8$	0.57	0.82	0.24	0.14	0.05	0.50	0.42	0.75	0.59	0.85	0.75	0.94	1.0	1.0

(b) Lower and upper probabilities Paris, EXP N1 (Overall sample $n = 80$)

E_t	Mean		Std		Min		25%		50%		75%		Max	
	L	U	L	U	L	U	L	U	L	U	L	U	L	U
$t = 15$	0.31	0.58	0.22	0.27	0.00	0.01	0.14	0.37	0.26	0.56	0.45	0.45	0.95	0.99
$t = 18$	0.35	0.66	0.26	0.24	0.00	0.03	0.14	0.50	0.32	0.69	0.47	0.47	1.00	1.00
$t = 20$	0.41	0.71	0.27	0.23	-0.01	0.01	0.20	0.60	0.40	0.76	0.61	0.61	1.00	1.00
$t = 22$	0.43	0.73	0.26	0.23	-0.01	0.00	0.20	0.58	0.39	0.80	0.61	0.61	1.00	1.00

(c) Lower and upper probabilities Sydney, EXP N1 (Overall sample $n = 80$)

E_t	Mean		Std		Min		25%		50%		75%		Max	
	L	U	L	U	L	U	L	U	L	U	L	U	L	U
$t = 7$	0.05	0.12	0.07	0.14	-0.01	0.00	0.00	0.02	0.04	0.08	0.08	0.16	0.35	0.16
$t = 10$	0.15	0.23	0.12	0.17	0.00	0.00	0.06	0.10	0.12	0.20	0.19	0.30	0.50	0.30
$t = 12$	0.24	0.35	0.15	0.18	0.00	0.04	0.14	0.22	0.20	0.32	0.31	0.48	0.63	0.48
$t = 15$	0.40	0.54	0.19	0.20	0.08	0.08	0.26	0.40	0.38	0.52	0.55	0.70	0.73	0.70
$t = 17$	0.60	0.75	0.16	0.15	0.18	0.22	0.54	0.65	0.64	0.78	0.71	0.86	0.86	0.86

(d) Lower and upper probabilities Maths, EXP N2 (Overall sample $n = 52$)

E_t	Mean		Std		Min		25%		50%		75%		Max	
	L	U	L	U	L	U	L	U	L	U	L	U	L	U
$t = 7$	0.12	0.22	0.08	0.14	0.01	0.02	0.06	0.11	0.11	0.20	0.17	0.31	0.36	0.56
$t = 10$	0.22	0.37	0.12	0.14	0.02	0.06	0.14	0.30	0.20	0.34	0.29	0.46	0.50	0.65
$t = 12$	0.33	0.51	0.13	0.14	0.14	0.20	0.22	0.40	0.32	0.50	0.40	0.64	0.60	0.74
$t = 15$	0.54	0.74	0.14	0.11	0.19	0.40	0.46	0.67	0.56	0.77	0.65	0.82	0.83	0.90
$t = 17$	0.71	0.86	0.13	0.07	0.25	0.60	0.65	0.82	0.74	0.86	0.83	0.91	0.90	1.00

(e) Lower and upper probabilities Contraction, EXP N2 (Overall sample $n = 52$)

Table C.1: Descriptive Statistics: lower and upper bounds of the elicited probability intervals in the three experiments

(Sample size, frequency)	Mean		Std		Min		25%		50%		75%		Max	
	L	U	L	U	L	U	L	U	L	U	L	U	L	U
(4, 0.5)	0.26	0.78	0.21	0.22	0.00	0.40	0.02	0.54	0.25	0.80	0.50	1.00	0.66	1.00
(4, 0.25)	0.11	0.63	0.13	0.35	0.00	0.03	0.01	0.30	0.05	0.61	0.22	1.00	0.38	1.00
(20, 0.5)	0.26	0.75	0.21	0.24	0.00	0.10	0.10	0.60	0.30	0.73	0.40	1.00	0.80	1.00
(20, 0.25)	0.16	0.65	0.15	0.32	0.00	0.10	0.00	0.30	0.15	0.72	0.25	1.00	0.50	1.00
(100, 0.5)	0.36	0.70	0.21	0.24	0.00	0.30	0.36	0.50	0.49	0.55	0.50	1.00	0.52	1.00
(100, 0.25)	0.17	0.56	0.14	0.34	0.00	0.20	0.00	0.25	0.24	0.31	0.25	1.00	0.70	1.00

Table C.2: Lower and upper bounds of stated probability intervals, EXP A

	t-tests		Mann-Whitney tests	
sample size	t-statistic	p-value	U-statistic	p-value
20	-5.69	< 0.001	503	< 0.001
100	-13.09	< 0.001	193.5	< 0.001
4	-5.83	< 0.001	555.5	< 0.001

Table C.3: Unpaired t-tests and Mann-Whitney tests of the hypothesis that the midpoint is the same for each pair of bags with the same sample size and frequencies 0.25 and 0.5, EXP A.

	t-tests			Binomial tests		
frequency	Deg. free.	Statistic	p-value	Deg. free.	Statistic	p-value
0.5	50	9.87	< 0.001	51	2	< 0.001
0.25	49	6.71	< 0.001	50	5	< 0.001

Table C.4: Paired t-tests and binomial tests of the hypothesis that the Imprecision is the same for each pair of bags with the same frequency of green and sample sizes 4 and 100, EXP A.

	EXP N1: MP		EXP N1: Paris		EXP N1: Sydney		EXP N2: Contraction		EXP N2: Maths	
	L	U	L	U	L	U	L	U	L	U
Count	74	78	79	78	78	78	52	52	52	52
Mean	0.62	0.66	0.56	0.56	0.27	0.41	0.99	0.99	0.98	1.00
Std	0.46	0.38	0.45	0.47	0.59	0.50	0.02	0.03	0.07	0.01
Min	-0.91	-0.91	-0.91	-0.91	-1.00	-1.00	0.95	0.80	0.53	0.95
25%	0.55	0.55	0.33	0.33	-0.14	0.00	1.00	1.00	1.00	1.00
50%	0.71	0.69	0.67	0.67	0.33	0.55	1.00	1.00	1.00	1.00
75%	0.91	0.91	1.00	1.00	0.67	0.91	1.00	1.00	1.00	1.00
Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table C.5: Individual-level Kendall τ_b descriptive statistics for all sources and tasks in EXP N1 and N2

Note τ_b is not defined for some subjects in EXP N1 (because of too many ties), and they were dropped.

source in EXPs N1 and N2. As could have been expected, in EXP N1, the frequency of monotonicity violations appears to increase with the difficulty of the choice task, with the MP task being arguably easier than that for probability-interval elicitation, and the task for Paris, the more familiar source for our subjects, being easier than that for Sydney.

Elicited points on a vertex Table C.6 reports counts of the number of subjects with a given number of elicited points on a vertex of the space $\mathscr{I}$ of interval-valued urns in Figure 1.

Point	EXP N1										EXP N2												EXP A			
	Paris					Sydney					Maths						Contraction									
	# subjects					# subjects					# subjects						# subjects						# subjects			
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	5	0	1	2	3	4	5	0	1	2	3
$[0,0]$, $[0,1]$ or $[1,1]$	79	0	0	1	0	78	1	0	1	0	46	5	1	0	0	0	52	0	0	0	0	0	95	6	0	0

Table C.6: Number of subjects with the specified number of elicited points being on a vertex of $\mathscr{I}$.

	EXP 1: Paris	EXP 1: Sydney	EXP 2: Maths	EXP 2: Contraction	EXP A
Count	80	80	52	52	101
Mean	0.26	0.30	0.11	0.15	0.36
Std	0.17	0.20	0.08	0.09	0.18
Min	0.00	0.00	0.00	0.01	0.00
25%	0.10	0.11	0.06	0.09	0.24
50%	0.23	0.28	0.09	0.14	0.35
75%	0.35	0.44	0.16	0.19	0.51
Max	0.83	0.76	0.36	0.43	0.77

Table C.7: Average Imprecision for each source in EXP N1 and N2, and across all events in EXP A (Section 4.2): descriptive statistics.

# subjects	EXP N1: Paris	EXP N1: Sydney	EXP N2: Maths	EXP N2: Contraction	EXP A
0	51	48	20	31	74
1	14	18	14	12	23
2	7	8	12	5	2
3	6	2	3	3	2
4	2	4	0	1	-
5	-	-	3	0	-
Total	80	80	52	52	101

Table C.8: Number of subjects with given number of precise events, per source.

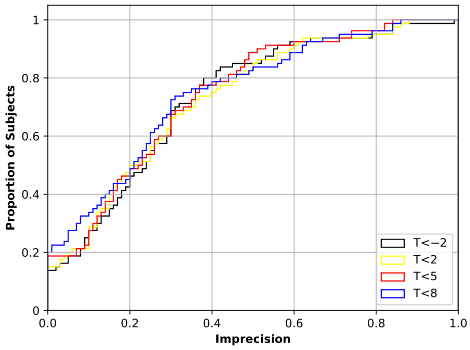
	Source	F	p-value		Source	F	p-value
EXP N1	Paris	0.1048	0.957	EXP N2	Contraction	4.0352	0.003
	Sydney	0.4769	0.698		Maths	5.863	0.00015

Table C.9: ANOVAs of the Imprecision related to an event (dependent variable) on the event (factor), for each source. (H_0: the Imprecision is identical across all events in the source.)

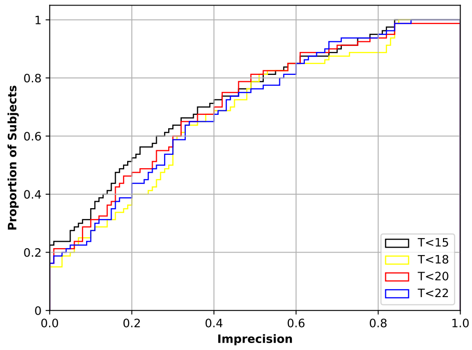
Imprecision Table C.7 presents the descriptive statistics for the Average Imprecision, whereas Table C.8 displays counts of the number of subjects with various numbers of precise elicited points. Table C.9 presents the results of ANOVAs of the Imprecision concerning an event against the event, for each source in EXPs N1 and N2, where the null hypothesis is that Imprecision is invariant across events. Figure C.1 plots CDFs of the Imprecision for each elicited event in each of the sources in EXPs N1 and N2, across subjects.

Matching versus Stated Probability Intervals Tables C.10 and C.11 present the descriptive statistics for the imprecision and midpoints of probability intervals, across subjects, for each bag (characterised by a sample size and frequency) and for the choice-based elicitation method and stated probability intervals respectively.

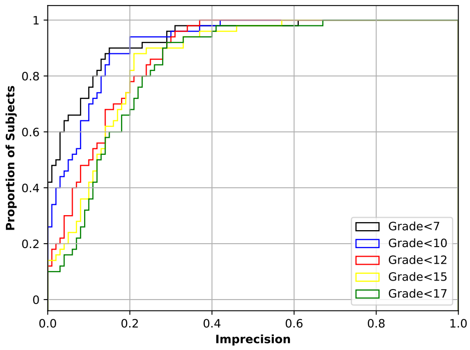
Modifications in the 2D choice list step Tables C.12 and C.13 provide data on the number of subjects that modified their interval in the 2D choice list confirmation step of the elicitation procedure, and the total number of modifications, including those introducing precision or imprecision.



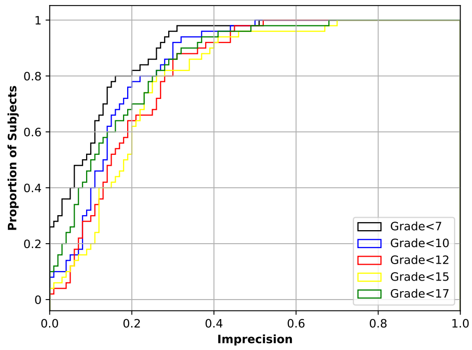
(a) EXP N1: Paris source



(b) EXP N1: Sydney source



(c) EXP N2: Maths source



(d) EXP N2: Contraction source

Figure C.1: CDFs of Imprecision across subjects in EXPs N1 and N2, for each elicited event

	sample size	frequency	mean	std	min	25%	50%	75%	max
Imprecision	4	0.5	0.55	0.29	0.00	0.30	0.64	0.83	0.96
		0.25	0.47	0.28	0.00	0.22	0.48	0.64	1.00
	20	0.5	0.45	0.28	0.00	0.15	0.48	0.76	0.80
		0.25	0.39	0.29	0.00	0.11	0.33	0.63	0.84
	100	0.5	0.15	0.12	0.00	0.09	0.10	0.21	0.48
		0.25	0.16	0.20	0.00	0.02	0.10	0.20	0.84
Midpoint	4	0.5	0.50	0.02	0.37	0.50	0.50	0.50	0.54
		0.25	0.42	0.09	0.20	0.37	0.44	0.50	0.63
	20	0.5	0.50	0.04	0.35	0.50	0.50	0.50	0.67
		0.25	0.40	0.12	0.12	0.29	0.46	0.50	0.62
	100	0.5	0.50	0.03	0.41	0.50	0.50	0.50	0.59
		0.25	0.31	0.10	0.19	0.25	0.25	0.34	0.62

Table C.10: EXP A: Descriptive statistics for Imprecision and midpoints of elicited probability intervals, for each bag (sample size and frequency).

The binary-choice procedure continued until the interval was estimated to a predetermined precision (Section 3.2 and Appendix D.1). To ascertain whether this imprecision in the estimate, combined with a reticence of subjects to modify their intervals in the 2D choice list step of our elicitation procedure, drives our findings about imprecision in beliefs, we repeat our analyses of overall imprecision in Section 4.2 under the assumption that in all such ‘possibly

	sample size	frequency	mean	std	min	25%	50%	75%	max
Imprecision	4	0.5	0.52	0.29	0.00	0.35	0.50	0.65	1.00
		0.25	0.52	0.32	0.02	0.25	0.50	0.75	0.99
	20	0.5	0.48	0.28	0.00	0.20	0.50	0.66	1.00
		0.25	0.49	0.28	0.00	0.22	0.50	0.79	0.98
	100	0.5	0.34	0.27	0.00	0.02	0.50	0.50	1.00
		0.25	0.39	0.29	0.00	0.15	0.27	0.75	0.87
Midpoint	4	0.5	0.52	0.16	0.20	0.50	0.50	0.58	0.83
		0.25	0.37	0.21	0.02	0.17	0.35	0.51	0.69
	20	0.5	0.51	0.18	0.05	0.50	0.50	0.58	0.90
		0.25	0.40	0.21	0.05	0.25	0.44	0.60	0.75
	100	0.5	0.53	0.18	0.15	0.50	0.50	0.73	0.76
		0.25	0.37	0.21	0.12	0.20	0.26	0.62	0.78

Table C.11: EXP A: Descriptive statistics for Imprecision and midpoints of stated probability intervals, for each bag (sample size and frequency).

Number of Modifications	EXP N1: Paris	EXP N1: Sydney	EXP N2: Maths	EXP N2: Contraction	EXP A
0	15	14	4	2	8
1	22	25	4	5	24
2	14	20	9	13	36
3	15	9	12	9	33
4	14	12	15	13	-
5	-	-	8	10	-
Total	80	80	52	52	101

Table C.12: Number of subjects who modified the interval provided in the 2D choice list step of the procedure for the given number events, per source.

For row $n = 0, \ldots, 5$, the entry in each column reports the number of subjects for which the result of the binary-choice procedure differed from interval confirmed in the 2D choice list for precisely n events in the specified experiment and source.

precise' cases, the imprecision is zero. More precisely, for every subject and event, we define the Possible Imprecision for that event and subject to be zero if: a. the binary-choice procedure halts due to the stopping rule but where, were it to continue, it could have arrived to a precise interval (i.e. an interval of zero width); and b. the subject did not modify the reported interval in the 2D choice list step of the elicitation procedure. For all other events, the Possible Imprecision coincides with the Imprecision. Table C.14 provides descriptive statistics for Average Possible Imprecision. As concerns the other results on Average Imprecision reported in Section 4.2, binomial tests reject the hypothesis of equal probability for the Average Possible Imprecision to be equal to vs. greater than 0 for each source ($p < 0.001$ in all cases), with a clear majority of subjects—99 out of 101 in EXP A, 79 out of 80 in EXP N1, and 52 out of 52 in EXP N2—having strictly positive Possible Imprecision on average. The similarity with the results concerning (uncorrected) Imprecision suggest that imprecision in our elicitation procedure is not a main driver of the imprecision in elicited beliefs.

Number of Modifications	EXP N1: Paris	EXP N1: Sydney	EXP N2: Maths	EXP N2: Contraction	EXP A
Total	151	140	158	160	195
Introducing precision	13	16	7	2	33
Introducing imprecision	17	15	37	31	0

Table C.13: Number of modifications in the 2D choice list step of the elicitation procedure across all subjects, per experiment and source.

Each entry reports the number of tasks in the specified experiment and source for which the result of the binary-choice procedure differed from the interval confirmed in the 2D choice list, across all subjects. The second row reports the number of such tasks where the result of the binary-choice procedure is imprecise (i.e. of strictly positive Imprecision), whereas the interval confirmed in the 2D choice list is precise (i.e. of zero Imprecision); the third row reports the number of such tasks where the result of the binary-choice procedure is precise, whereas the interval confirmed in the 2D choice list is imprecise.

	EXP 1: Paris	EXP 1: Sydney	EXP 2: Maths	EXP 2: Contraction	EXP A
Count	80	80	52	52	101
Mean	0.25	0.29	0.11	0.15	0.35
Std	0.18	0.20	0.08	0.09	0.18
Min	0	0	0	0.01	0.00
25%	0.09	0.11	0.05	0.09	0.22
50%	0.23	0.28	0.09	0.14	0.35
75%	0.35	0.44	0.16	0.19	0.51
Max	0.83	0.76	0.36	0.43	0.77

Table C.14: Average Possible Imprecision for each source in EXP N1 and N2, and across all events in EXP A (Section C.1): descriptive statistics.

C.2. Bayesian estimation for EXPs N1 and N2

C.2.1. Statistical approach

Estimation of upper and lower CDFs in EXP N1 and EXP N2 Recall that T denotes the space of possible values of the variables of interest (minimum temperatures in EXP N1, grades in EXP N2). For each source, we estimate general models of the form:

$$\begin{cases} \underline{p}(E) = \underline{f}(E) + \underline{\varepsilon} \\ \overline{p}(E) = \overline{f}(E) + \overline{\varepsilon} \end{cases} \tag{11}$$

where $\underline{p}(E)$ (resp. $\overline{p}(E)$) are the elicited lower (resp. upper) probabilities of events E in Table 3, $\underline{f}$ and $\overline{f}$ are CDFs over T from specified two-parameter families (Table C.15), with parameters $\underline{a}, \underline{b}$ (resp. $\overline{a}, \overline{b}$), and $\underline{\varepsilon}$ and $\overline{\varepsilon}$ are zero-mean normal distributions with variance $\underline{\sigma}^2$ and $\overline{\sigma}^2$ respectively.

For each equation, the parameter space is $\Theta \subseteq \mathbb{R}^3$, with a typical point $(\underline{a}, \underline{b}, \underline{\sigma})$ (resp. $(\overline{a}, \overline{b}, \overline{\sigma})$) specifying an $\underline{f}$ (resp. $\overline{f}$) and the variance of the relevant error term. We specify the following priors over the hyperparameters : a, b, σ are realisations from $A \sim N(\mu_a, \sigma_a^2)$,

	Temperature (EXP N1)	Grade (EXP N2)
Family 1	Truncated Normal $\mathcal{N}(a,b)$	Truncated Normal $\mathcal{N}(a,b)$
Family 2	Beta $B(a,b)$	Beta $B(a,b)$
Support	[min of min stated temperature, max of max stated temperature]	[0,20]

Table C.15: Families of distributions over T (temperature; mark)
Note the minima and maxima in the first column are taken across all subjects' responses (Section 3.3)

$B \sim N(\mu_b, \sigma_b^2)$ and $\Sigma = \sigma_\sigma \mid Y \mid$ with $Y \sim N(0,1)$.

We use a MCMC-like approach to estimate the posterior distributions of these distributions through the use of the Python package PyMC3, and more specifically, the No-U-Turn Sampler algorithm (NUTS) (Hoffman and Gelman, 2014).

The likelihood of observations $x_1, ..., x_n$ pertaining to $t_1, \ldots, t_n$ (e.g. elicited lower probabilities for cumulative events $E_{t_i} = \{t \in T : t \leq t_i\}$) given the point $(a,b,\sigma) \in \Theta$ is:

$$L(a,b,\sigma|x_1,\ldots,x_n) \quad = \quad \prod_{i\in\{1,\ldots,n\}} \varphi\left(\frac{x_i - f_{(a,b)}(\{t \leq t_i\})}{\sigma}\right)$$

where $f_{(a,b)}$ is the CDF with parameters a,b and φ is the density of the normal distribution. Hence the likelihood of hyperparameters $\mu_a, \sigma_a^2, \mu_b, \sigma_b^2, \mu_\sigma, \sigma_\sigma^2$ given observations $x_1 \ldots x_n$ is :

$$\begin{aligned} & L(\mu_a, \sigma_a^2, \mu_b, \sigma_b^2, \mu_\sigma, \sigma_\sigma^2|x_1,\ldots,x_n) \\ = \quad & \int_{(a,b,\sigma)\in\Theta} L(a,b,\sigma|x_1,\ldots,x_n)dp(a,b,\sigma|\mu_a, \sigma_a^2, \mu_b, \sigma_b^2, \mu_\sigma, \sigma_\sigma^2) \end{aligned}$$

$L(\mu_{\underline{a}}, \sigma_{\underline{a}}^2, \mu_{\underline{b}}, \sigma_{\underline{b}}^2, \mu_{\underline{\sigma}}, \sigma_{\underline{\sigma}}^2|\underline{x_1},\ldots,\underline{x_n})$ and $L(\mu_{\overline{a}}, \sigma_{\overline{a}}^2, \mu_{\overline{b}}, \sigma_{\overline{b}}^2, \mu_{\overline{\sigma}}, \sigma_{\overline{\sigma}}^2|\overline{x_1},\ldots,\overline{x_n})$ are used by the NUTS algorithm to estimate the posterior distributions of A, B and Σ, where $\underline{x_1},\ldots,\underline{x_n},\overline{x_1},\ldots,\overline{x_n}$ are the elicited lower and upper probabilities respectively, under the parametric families for f given in Table C.15.

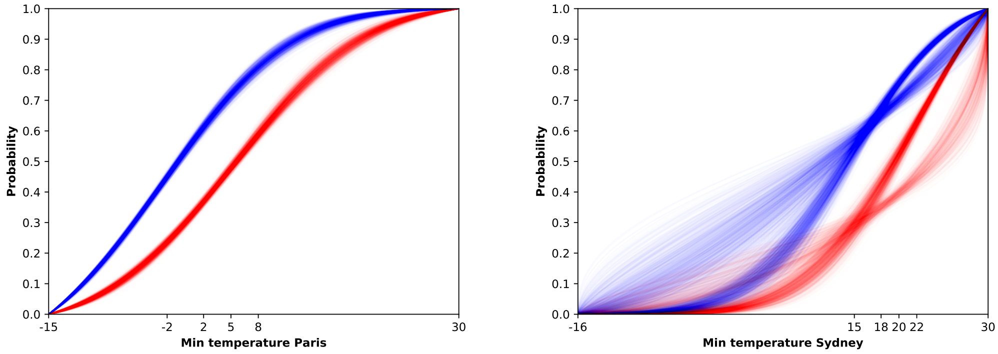
Likelihood estimation of α in EXP N1 (Paris treatment) For the Bayesian estimation of the mixture coefficient α in the α-maxmin EU model, we supplement the general model (11) with the following equations

$$\begin{cases} MP(E) = \alpha\underline{p}(E) + (1-\alpha)\overline{p}(E) + \varepsilon_{\underline{\alpha}} \\ 1 - MP(E^c) = \alpha\overline{p}(E) + (1-\alpha)\underline{p}(E) + \varepsilon_{\overline{\alpha}} \end{cases} \tag{12}$$

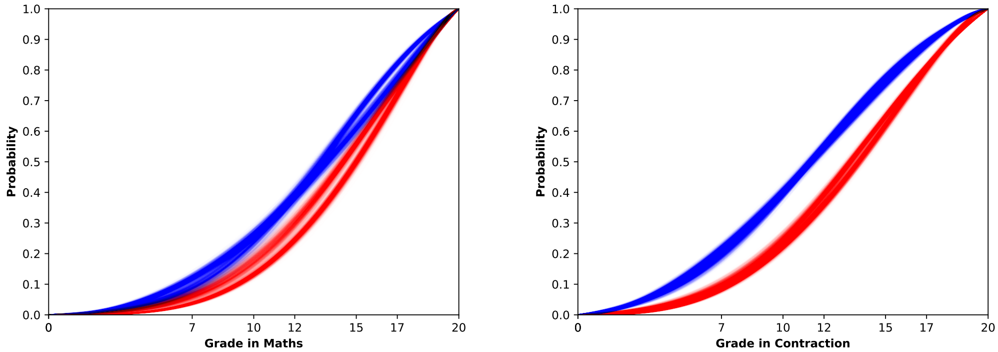
which are discussed in Section 4.4. We assume that α follows a beta distribution $B(a_\alpha, b_\alpha)$, and the $\varepsilon_{\overline{\alpha}}$ and $\varepsilon_{\underline{\alpha}}$ are zero-mean normal distributions, with the hyperparameters independent

	Distribution	EXP N1		EXP N2	
		Paris	Sydney	Mathematics	Contraction
AIC	Normal	706.65	700.79	411.22	385.52
	Beta	648.26	684.36	416.18	390.64
BIC	Normal	711.42	705.56	415.12	389.42
	Beta	653.02	689.12	420.08	394.54

Table C.16: AIC and BIC under (truncated) normal and Beta specifications for CDFs (Table C.15).



(a) EXP N1



(b) EXP N2

Figure C.2: Bayesian estimation of lower and upper CDFs: plots of 1000 samples from MCMC (Truncated Normal distribution for EXP N1; Beta distribution for EXP N2)

and normally distributed, as above, with variances $\sigma^2_{\overline{\alpha}}$ and $\sigma^2_{\underline{\alpha}}$.

The MPs have been elicited for the Paris treatment in EXP N1. The hyperparameters concerning the upper and lower CDFs discussed above and those for α were estimated under the model composed of (11) and (12) using the NUTS algorithm, with the procedure set out above.

C.2.2. Analysis

Table C.16 presents the AIC and BIC criteria of goodness of fit for the parametric forms in Table C.15, which justify the choice of forms to present in Figure 4. Figure C.2 displays the upper and lower distributions under the parametric families not shown in Figure 4. Table C.17 reports statistics on the distribution over parameters under the estimated hyperparameters.

	mean		sd		mc_error		hpd_2.5		hpd_97.5		n_eff		Rhat	
	N	B	N	B	N	B	N	B	N	B	N	B	N	B
$\underline{a}$	5.25	1.43	0.31	0.23	0.01	0.01	4.65	1.04	5.87	1.91	1009.11	569.30	1.0	1.0
$\overline{a}$	-2.57	1.64	0.37	0.29	0.01	0.01	-3.27	1.13	-1.87	2.24	542.99	571.99	1.0	1.0
a_α	3.42	1.07	1.62	0.16	0.06	0.01	0.53	0.73	6.38	1.39	630.90	541.73	1.0	1.0
b_α	1.80	2.46	1.04	0.35	0.05	0.01	0.10	1.76	3.80	3.17	429.61	522.49	1.0	1.0
$\underline{b}$	11.35	4.32	0.75	1.76	0.02	0.06	9.71	1.06	12.66	7.87	1214.33	812.98	1.0	1.0
$\overline{b}$	11.00	1.90	0.64	1.08	0.03	0.04	9.78	0.18	12.17	3.92	614.95	606.36	1.0	1.0
$\underline{\sigma}$	0.22	0.22	0.01	0.01	0.00	0.00	0.20	0.20	0.23	0.23	1043.67	1163.74	1.0	1.0
$\overline{\sigma}$	0.18	0.18	0.01	0.01	0.00	0.00	0.17	0.17	0.19	0.19	1055.83	1239.11	1.0	1.0
$\sigma_{\overline{\alpha}}$	0.21	0.21	0.01	0.01	0.00	0.00	0.20	0.20	0.23	0.23	930.43	1134.15	1.0	1.0
$\sigma_{\underline{\alpha}}$	0.19	0.19	0.01	0.01	0.00	0.00	0.17	0.17	0.20	0.20	909.78	1409.46	1.0	1.0
α	0.81	0.81	0.04	0.04	0.00	0.00	0.74	0.74	0.88	0.88	754.54	1079.99	1.0	1.0

(a) Paris (EXP N1)

	mean		sd		mc_error		hpd_2.5		hpd_97.5		n_eff		Rhat	
	N	B	N	B	N	B	N	B	N	B	N	B	N	B
$\underline{a}$	22.03	1.12	0.45	0.55	0.01	0.04	21.11	0.31	22.80	2.21	1130.92	1.91	1.0	1.57
$\overline{a}$	14.66	0.14	0.46	0.32	0.01	0.03	13.78	-0.27	15.48	0.67	876.71	1.07	1.0	4.19
$\underline{b}$	9.62	1.32	0.95	0.48	0.03	0.02	7.88	0.49	11.67	2.24	1018.57	320.38	1.0	1.00
$\overline{b}$	9.04	0.94	0.85	0.24	0.02	0.01	7.42	0.50	10.78	1.37	882.98	321.66	1.0	1.00
$\underline{\sigma}$	0.26	0.25	0.01	0.01	0.00	0.00	0.23	0.23	0.28	0.27	933.24	522.03	1.0	1.00
$\overline{\sigma}$	0.25	0.24	0.01	0.01	0.00	0.00	0.23	0.22	0.27	0.26	831.75	671.56	1.0	1.00

(b) Sydney (EXP N1)

	mean		sd		mc_error		hpd_2.5		hpd_97.5		n_eff		Rhat	
	N	B	N	B	N	B	N	B	N	B	N	B	N	B
$\underline{a}$	15.88	3.76	0.17	0.09	0.01	0.0	15.57	3.58	16.22	3.93	788.64	978.44	1.0	1.0
$\overline{a}$	5.40	1.46	0.28	0.05	0.01	0.0	4.86	1.36	5.95	1.55	955.04	900.67	1.0	1.0
$\underline{b}$	13.97	2.27	0.17	0.09	0.00	0.0	13.65	2.10	14.33	2.43	1218.24	1049.31	1.0	1.0
$\overline{b}$	5.03	1.22	0.25	0.05	0.01	0.0	4.58	1.12	5.53	1.32	928.02	1031.36	1.0	1.0
$\underline{\sigma}$	0.14	0.16	0.01	0.01	0.00	0.0	0.13	0.14	0.16	0.17	1158.52	1311.22	1.0	1.0
$\overline{\sigma}$	0.17	0.19	0.01	0.01	0.00	0.0	0.16	0.17	0.19	0.20	958.54	1247.27	1.0	1.0

(c) Maths (EXP N2)

	mean		sd		mc_error		hpd_2.5		hpd_97.5		n_eff		Rhat	
	N	B	N	B	N	B	N	B	N	B	N	B	N	B
$\underline{a}$	14.17	3.06	0.13	0.09	0.00	0.0	13.93	2.90	14.42	3.23	1090.28	888.13	1.0	1.00
$\overline{a}$	5.32	1.57	0.22	0.05	0.01	0.0	4.90	1.48	5.78	1.67	1068.01	937.24	1.0	1.00
$\underline{b}$	11.61	1.96	0.13	0.07	0.00	0.0	11.37	1.82	11.87	2.11	1167.95	478.84	1.0	1.01
$\overline{b}$	5.39	1.56	0.21	0.06	0.01	0.0	4.99	1.45	5.78	1.66	1209.38	460.19	1.0	1.00
$\underline{\sigma}$	0.12	0.14	0.01	0.01	0.00	0.0	0.11	0.13	0.13	0.15	1501.55	966.76	1.0	1.00
$\overline{\sigma}$	0.12	0.13	0.01	0.01	0.00	0.0	0.11	0.12	0.13	0.15	1144.85	894.44	1.0	1.00

(d) Contraction (EXP N2)

Table C.17: Statistics for parameters under Bayesian estimation; Normal (N) and Beta (B) parametrisations
Note mc_error: Monte Carlo procedure standard error; hdp_2.5 / hdp_97.5: Highest posterior density 2.5 and 97.5 percentiles; n_eff : count of iteration in the MCMC procedure.

C.3. EXP N1: Matching Probability data and analysis of α

Table C.18 provides descriptive statistics on the elicited MPs in EXP N1. Table C.19 provides descriptive statistics on the α estimated from the raw data (from Eqs. (3)). These equations cannot be applied to estimate α whenever the upper and lower probabilities of an event coincide, i.e. $\underline{p}(E) = \overline{p}(E)$; Table C.19 performs the estimates using all events for which the equations can be applied—and hence only removes the two subjects for which the upper and lower probabilities coincide for all events (Table C.8).

t	$MP(E_t)$								$1-MP(E_t^c)$						
	count	mean	std	min	25%	50%	75%	max	mean	std	min	25%	50%	75%	max
−2	80	0.35	0.21	0.02	0.17	0.37	0.47	1.00	0.50	0.19	0.03	0.38	0.48	0.63	0.98
2	80	0.44	0.20	0.02	0.27	0.47	0.57	0.97	0.59	0.19	0.23	0.48	0.57	0.74	0.98
5	80	0.54	0.23	0.02	0.37	0.55	0.68	0.97	0.71	0.20	0.23	0.53	0.73	0.92	0.98
8	80	0.60	0.21	0.17	0.47	0.57	0.76	0.97	0.77	0.17	0.43	0.63	0.80	0.93	0.98

Table C.18: Descriptive statistics for $MP(E_i)$ and $1-MP(E_i^c)$ in Paris treatment, EXP N1

	count	mean	std	min	25%	50%	75%	max
α	78	0.97	0.66	-0.32	0.62	0.80	1.17	3.84

Table C.19: Descriptive statistics for α, estimated from raw data according to Eqs. (3). Estimation conducted across all subjects such that, for any least one event E, $\underline{p}(E) \neq \overline{p}(E)$.

C.4. Elicitation-free check of $\alpha > \frac{1}{2}$

Under the α-maxmin EU model (2), it follows from Eqs. (3) that

$$MP(E) + MP(E^c) = 1 + (\overline{p}(E) - \underline{p}(E)).(1-2\alpha)$$

Since $\overline{p}(E) - \underline{p}(E) \geq 0$ by definition, it follows that, whenever there is imprecision, $MP(E) + MP(E^c) < 1$ if and only if $\alpha > \frac{1}{2}$.

Table C.20 displays the descriptive statistics for the sum $MP(E) + MP(E^c)$ for the Paris source in EXP1. It is clear that the vast majority of subjects have a sum of MPs less than 1 indicating an α greater than 0.5. Indeed, over 80% of subjects have sum of MPs less than or equal to 1.

t	count	mean	std	min	25%	50%	75%	max
−2	80	0.84	0.20	0.29	0.71	0.89	0.98	1.31
2	80	0.85	0.20	0.29	0.73	0.89	0.99	1.34
5	80	0.83	0.22	0.24	0.69	0.89	0.99	1.29
8	80	0.83	0.18	0.39	0.69	0.89	0.99	1.26

Table C.20: Empirical distribution of average $MP(E_t) + MP(E_t^c)$ across all events for which MPs were elicited (those concerning Paris temperature in EXP N1).

D. Experimental design and displays

D.1. Binary-choice procedure

D.1.1. Introduction and setup

Our binary-choice procedure is fully described in Figures D.2–D.5; the algorithm, coded in Python, is provided in the supplementary materials here. Figure D.2 sets out the general structure (and stopping rules). At each step of the procedure, preferences are elicited for a single probability interval $[\underline{p}_i, \overline{p}_i]$: i.e. preferences between the bet on the event and the IL $(z, [\underline{p}_i, \overline{p}_i], 0)$, and between the bet on the complement event and the complementary IL $(0, [\underline{p}_i, \overline{p}_i], z)$. The heart of the procedure, detailed in Figures D.3–D.5, involves specification of the next probability interval proposed for elicitation on the basis of the preferences concerning the previous intervals. We first set out the notation used in the presentation of these parts of the procedure, before explaining informally its main steps. Throughout, we adopt the Euclidean topology on $\mathscr{I} \subseteq \mathbb{R}^2$, and let $d(\bullet, \bullet)$ be the Euclidean distance. Moreover, recall from Section 2.1 that an interval-valued urn $[p, q]$, i.e. with a minimum proportion p of red balls and a minimum proportion $1 - q$ of blue balls, corresponds to a probability interval; we shall present the procedure in terms of the latter here.

The procedure draws on two formal elements. The first is the assignment of interval-valued urns—or equivalently probability intervals—to one of four preference-defined regions, as set out in Table 1 (Section 2.5). For instance, in Figure D.1, which we shall use to illustrate the procedure, the probability intervals already elicited are the dots coloured white, red, blue and red-blue according to the region they belong to.

The second element is a 'polar'-style coordinate system for the set of probability intervals $\mathscr{I}$, under which, informally, $(m, \alpha) \in [0, 0.5] \times [0, 1]$ is the probability interval that is α along the piecewise-linear line that goes through the probability intervals $[0, 0]$, $[1, 1]$, and $[m, 1 - m]$ (corresponding to the urn with at least proportion m of red balls and at least proportion m of blue balls). The thick grey line in Figure D.1 is one such line. Formally, $\sigma : \mathscr{I} \to [0, 0.5] \times [0, 1]$ is defined by:

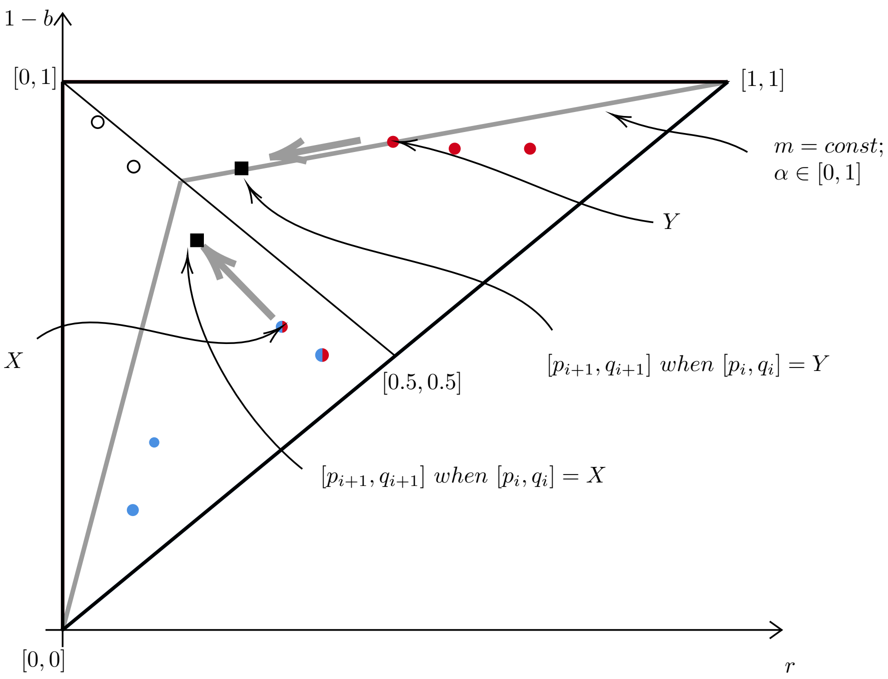


Figure D.1: Binary Choice Procedure.

$$\sigma([p,q]) = \begin{cases} (\frac{p}{p+q}, \frac{p+q}{2}) & p \leq 1-q, p+q \in (0,2) \\ (\frac{1-q}{2-p-q}, \frac{p+q}{2}) & p > 1-q, p+q \in (0,2) \\ (0,0) & p=q=0 \\ (0,1) & p=q=1 \end{cases} \tag{13}$$

It is straightforward to check that σ is a well-defined function on $\mathscr{I}$. Every point except for $[0,0], [1,1]$ corresponds to a unique line (parametrised by m) and 'distance' along that line (parametrised by α). $[0,0]$ (respectively $[1,1]$) corresponds to a single α, namely 0 (resp. 1), though it lies on all such lines; we set the corresponding $m = 0$ by convention. We write $\sigma_1([p,q])$ (respectively $\sigma_2([p,q])$) for the first (resp. second coordinate) of $\sigma([p,q])$. Since this is a simple change of coordinates, we shall write $(m,\alpha) \in B$ as short for $\sigma^{-1}(m,\alpha) \in B$, and similarly for other cases.

D.1.2. Presentation of main steps

As discussed in Section 2.5 (Proposition 1), elicited points in the R-B and Wh regions determine an area in $\mathscr{I}$ 'between the R-B and the Wh points' to which the MPI must belong. The general aim of the procedure is thus to find progressively 'closer' points in R-B and Wh, hence reducing the size of this area. This motivates the two main steps in the determination of the next

Procedure Binary Choice Procedure: structure and stopping rules

Set $[p_1, q_1] = [0.3, 0.7]$, $i = 1$ and $El = \emptyset$; /* $[p_i, q_i]$ is the last interval for which preferences were elicited, and $El \not\ni [p_i, q_i]$ is the set of intervals for which preferences have been elicited previously. So $El \cup \{[p_i, q_i]\}$ is the set of all elicited intervals, including the one under consideration. */

Set $rb_{El} = \arg\max_{[p,q] \in El \cap R-B} q$, $w_{El} = \arg\min_{[p,q] \in El \cap Wh} q$, $r_{El} = \arg\min_{[p,p] \in El \cap R} p$, $b_{El} = \arg\max_{[p,p] \in El \cap B} p$; /* rb_{El} is the highest elicited point in R-B, r_{El} is the lowest elicited precise point in R, etc. */

Set $D(El) = d(rb_{El}, w_{El})$ if $El \cap R - B \neq \emptyset$ and $El \cap Wh \neq \emptyset$; $\min\{d(w_{El}, [p,p]) : p \in [0,1]\}$ if $El \cap R - B = \emptyset$ and $El \cap Wh \neq \emptyset$; $\min(\min\{d(rb_{El}, [p,1]) : p \in [0,1]\}, \min\{d(rb_{El}, [0,p]) : p \in [0,1]\})$ if $El \cap R - B \neq \emptyset$ and $El \cap Wh = \emptyset$; undefined otherwise. ; /* $D(El)$ is the smallest distance between points in $El \cap R - B$ and $El \cap Wh$ when both non-empty, and the smallest distance to the appropriate boundary when only one non-empty. */

while $|El| < 12$ **do**
- **while** $\neg(D(El) < 0.15)$ & $\neg(d(r_{El}, b_{El}) < 0.05)$ **do**
  - Elicit preferences for $[p_i, q_i]$;
  - Execute algorithm 1 ;
  - Add $[p_i, q_i]$ to El ;
  - **if** $D(El) < 0.15$ **then**
    - **return** $\frac{rb_{El} + w_{El}}{2}$;
    - Stop
  - **if** $d(r_{El}, b_{El}) < 0.05$ **then**
    - **return** $\frac{b_{El} + r_{El}}{2}$;
    - Stop

if $El \cap R - B \neq \emptyset$ and $El \cap Wh = \emptyset$ **then**
- **return** rb_{El}

if $El \cap Wh \neq \emptyset$ and $El \cap R - B = \emptyset$ **then**
- **return** w_{El}

if $El \cap R - B \neq \emptyset$ and $El \cap Wh \neq \emptyset$ **then**
- **return** $\frac{rb_{El} + w_{El}}{2}$

Figure D.2: Binary choice procedure: structure

probability interval to be presented for elicitation, $[p_{i+1}, q_{i+1}]$, on the basis of the previously elicited point $[p_i, q_i]$.

On the one hand, if $[p_i, q_i]$ is in the R-B region (respectively, the Wh region), then by Proposition 1 a. (Section 2.5), the MPI will be North-West of $[p_i, q_i]$ (resp. South-East of $[p_i, q_i]$) in Figure 1: i.e. $\underline{p} \leq p_i$ and $\overline{p} \geq q_i$ (resp. $\underline{p} \geq p_i$ and $\overline{p} \leq q_i$), where the MPI is $[\underline{p}, \overline{p}]$. In such cases, the procedure proposes a $[p_{i+1}, q_{i+1}]$ North-West (resp. South-East) of $[p_i, p_i]$. This exemplified by the $[p_{i+1}, q_{i+1}]$ proposed for point X in Figure D.1. The precise proposal for $[p_{i+1}, q_{i+1}]$ depends on whether there is a point in Wh (resp. R-B); technicalities aside, this

Algorithm 1: Determination of Next Binary Choice

1 **Case 1**

2 **if** $(El \cup \{[p_i, q_i]\}) \cap R - B = (El \cup \{[p_i, q_i]\}) \cap Wh = \emptyset$ **then**

3 **if** *there is no* $[p, q] \in El$ *with* $\sigma_1([p, q]) = m$, *where* $m = \sigma_1([p_i, q_i])$ **then**

4 **if** $[p_i, q_i] \in R$ **then**

5 $[p_{i+1}, q_{i+1}] = \sigma^{-1}(m, \max\{\alpha_0 | (m_0, \alpha_0) \in El \cap B\})$;

6 **if** $[p_i, q_i] \in B$ **then**

7 $[p_{i+1}, q_{i+1}] = \sigma^{-1}(m, \min\{\alpha_0 | (m_0, \alpha_0) \in El \cap R\})$;

8 **if** *there exists* $[p, q] \in El$ *with* $\sigma_1([p, q]) = m$, *where* $m = \sigma_1([p_i, q_i])$, *but there exists no* $[p, q] \in El$ *with* $[p, q] \in B$ *and* $\sigma_1([p, q]) = m$ **then**

9 $[p_{i+1}, q_{i+1}] = \sigma^{-1}(m, \frac{\sigma_2([p_i, q_i])}{2})$;

10 **if** *there exists* $[p, q] \in El$ *with* $\sigma_1([p, q]) = m$, *where* $m = \sigma_1([p_i, q_i])$, *but there exists no* $[p, q] \in El$ *with* $[p, q] \in R$ *and* $\sigma_1([p, q]) = m$ **then**

11 $[p_{i+1}, q_{i+1}] = \sigma^{-1}(m, \frac{1+\sigma_2([p_i, q_i])}{2})$;

12 **if** *there exists* $[p, q], [p', q'] \in El$ *with* $[p, q] \in B$, $[p', q'] \in R$ *and* $\sigma_1([p, q]) = \sigma_1([p', q']) = \sigma_1([p_i, q_i])$=*m* **then**

13 **if** $[p_i, q_i] \in R$ **then**

14

$$[p_{i+1}, q_{i+1}] = \begin{cases} ((m + 0.5)/2, (\sigma_2([p_i, q_i] + \alpha')/2) & m \leq 0.44 \\ (0.5, (\sigma_2([p_i, q_i] + \alpha')/2) & m > 0.44 \end{cases}$$

where $\alpha' = \max\{\alpha_0 | (m, \alpha_0) \in El \cap B\}$;

15 **if** $[p_i, q_i] \in B$ **then**

16

$$[p_{i+1}, q_{i+1}] = \begin{cases} ((m + 0.5)/2, (\sigma_2([p_i, q_i] + \alpha'')/2) & m \leq 0.44 \\ (0.5, (\sigma_2([p_i, q_i] + \alpha'')/2) & m > 0.44 \end{cases}$$

where $\alpha'' = \min\{\alpha_0 | (m, \alpha_0) \in R\}$;

17 **end**

Figure D.3: Determination of Next Binary Choice: Part 1
Notation: σ defined in (13).

is the general strategy of the cases in lines 20-23 and 36-39 of the procedure (Figures D.4-D.5). If the point $[p_{i+1}, q_{i+1}]$ turns out to be in R-B or Wh, this will further restrict the area where the MPI can lie.

On the other hand, if $[p_i, q_i]$ is in the R or B regions, then Proposition 1 a. does not apply; as discussed in Section 2.5, the aim in such cases is to find a point in the R-B or Wh regions, to continue reducing the area containing the MPI. The procedure draws on two observations. First, as mentioned above, any point $[p_i, q_i]$ can be equivalently written in another coordinate system, specifying the line it sits on—parametrised by $m = \sigma_1([p_i, q_i])$—and how 'far' along the line it is—parametrised by $\alpha = \sigma_2([p_i, q_i])$. Second, for $[p_i, q_i]$ in R (respectively B), by Proposition 1 b., all points North-East (resp. South-West) of $[p_i, q_i]$ are also in R (resp. B). So the only points in R-B and W on the line $m = \sigma_1([p_i, q_i])$ corresponding to the point $[p_i, q_i]$ must be South-West of $[p_i, q_i]$, i.e. with lower α (resp. North-East, i.e. with higher α). Accordingly, the procedure proposes a point $[p_{i+1}, q_{i+1}]$ on the line $m = \sigma_1([p_i, q_i])$ but shifted in the appropriate direction, as illustrated by the $[p_{i+1}, q_{i+1}]$ proposed for point Y (lying in the R region) in Figure D.1.

Technicalities aside, this is general strategy for Case 1 (lines 1-17) and the cases in lines 24-34 and lines 40-44 of the procedure (Figures D.3–D.5). Among these cases, all retain the same m (grey line in Figure D.1) except those considered in lines 12-17. These treat cases where no point in R-B or Wh has yet been found; the procedure in these cases increases m during the search, hence looking closer to the 45° line (ie. the line of $[p,q]$ with $p=q$). We use a procedure with this in-built precision bias to favour Bayesian replies (i.e. precise probabilities); in the light of it, our finding of widespread imprecision (Section 4.2) is all the more remarkable.

D.1.3. Convergence

Except for extreme cases, the procedure tends to the MPI.

Proposition D.1. *Let E be an event, and suppose preferences are represented according to (1) with a unique MPI for E and W differentiable with $\partial_1 W([\underline{p}(E),\overline{p}(E)]) > \partial_2 W([\underline{p}(E),\overline{p}(E)]) > 0$. Let $[\underline{p}_n,\overline{p}_n]$ be the result of the procedure in Figures D.3–D.5 (with initial values set as in Figure D.2) applied for n steps. Then $[\underline{p}_n,\overline{p}_n] \to [\underline{p}(E),\overline{p}(E)]$ as $n \to \infty$. Moreover, the procedure also converges in this sense when preferences are represented according to (1) with $\partial_1 W([\underline{p}(E),\overline{p}(E)]) > \partial_2 W([\underline{p}(E),\overline{p}(E)]) = 0$, $\underline{p}(E) \neq 0$ and $\overline{p}(E) \neq 1$.*

Proof. We provide the main steps of the proof here; they rely on technical Lemmas E.1–E.4, which are detailed in Appendix E. We adopt the notation and initial values from Figure D.2; in particular, let El_n be the set of elicited points after n steps. To simplify notation, we set $El_n \cap Wh = El_n^{Wh}$ and $El_n \cap R-B = El_n^D$. As discussed in Section 2.4, the MPI is $[\underline{p}(E),\overline{p}(E)]$. Moreover, by Proposition A.1, at stage n, the MPI is contained in

$$\Phi_n = \left\{ [p,q] \in \mathscr{I} : \begin{array}{c} \max\left\{p' : [p',q'] \in El_n^{Wh}\right\} \leq p \leq \min\left\{p'' : [p'',q''] \in El_n^{D}\right\}, \\ \max\left\{q'' : [p'',q''] \in El_n^{D}\right\} \leq q \leq \min\left\{q' : [p',q'] \in El_n^{Wh}\right\} \end{array} \right\} \tag{14}$$

where the maximum of an empty set is taken to be 0 and the minimum 1.

We reason referring to the cases in the procedure (Figures D.3–D.5). At the beginning of the procedure, it is in Case 1 ($El_0^{Wh} = El_0^D = \emptyset$). By lines 13-16, if no point in Wh or R-B is found, the points elicited by the procedure will reach the space of precise probabilities (i.e. points $[p,q]$ with $p=q$), where it will follow a standard bisection procedure. All such points have σ_1-value of 0.5. It follows from Lemma E.1 that if the MPI is not precise, then a point will be found in R-B, so the procedure moves to Case 2. On the other hand, if the MPI is precise, then, by Lemma E.1 and the bisection character of the procedure on the space of precise probabilities, the points elicited in the procedure will converge to it as required.

Now consider cases where the procedure arrives to Case 2 or 3, i.e. it finds a point in R-B or Wh. By Lemma E.3, $\sigma_1([p_n,q_n]) \to \sigma_1([\underline{p}(E),\overline{p}(E)])$ as $n \to \infty$. We distinguish three cases.

- $\sigma_1([\underline{p}(E),\overline{p}(E)]) > 0$ and $\sigma_1([p_n,q_n]) \neq \sigma_1([\underline{p}(E),\overline{p}(E)])$ for all n. By Proposition A.1

Case 2

if *either* $(El \cup \{[p_i, q_i]\}) \cap R - B \neq \emptyset$ or $El \cup \{[p_i, q_i]\} \cap Wh \neq \emptyset$, but not both **then**

  if $[p_i, q_i] \in R - B$ **then**

$$[p_{i+1}, q_{i+1}] = \begin{cases} \left[\frac{p_i + (p_i + q_i - 1)}{2}, \frac{q_i + 1}{2}\right] & p_i + q_i > 1 \\ \left[\frac{p_i}{2}, \frac{q_i + (p_i + q_i)}{2}\right] & p_i + q_i \leq 1 \end{cases}$$

  if $[p_i, q_i] \in Wh$ **then**

$$[p_{i+1}, q_{i+1}] = \left[\frac{2p_i + q_i}{2}, \frac{p_i + 2q_i}{2}\right]$$

  if $[p_i, q_i] \in R$ **then**

    if $El \cap R - B \neq \emptyset$ **then**

$$[p_{i+1}, q_{i+1}] = \sigma^{-1}\left(\sigma_1([p_i, q_i])), \frac{\sigma_2([p_i, q_i]) + \alpha'}{2}\right)$$

    where

$$\alpha' = \begin{cases} \frac{q'}{2(1 - \sigma_1([p_i, q_i])} & q' \leq 1 - \sigma_1([p_i, q_i]) \\ \frac{q' - 1}{2\sigma_1([p_i, q_i])} + 1 & q' > 1 - \sigma_1([p_i, q_i]) \end{cases}$$

    with $[p', q'] = \arg\max_{[p,q] \in El \cap R - B} q$.

    if $El \cap Wh \neq \emptyset$ **then**

$$[p_{i+1}, q_{i+1}] = \sigma^{-1}\left(\sigma_1([p_i, q_i])), \frac{\sigma_2([p_i, q_i]) + \alpha''}{2}\right)$$

    where

$$\alpha'' = \begin{cases} \frac{p''}{2\sigma_1([p_i, q_i])} & p'' \leq \sigma_1([p_i, q_i]) \\ \frac{p'' + 1 - 2\sigma_1([p_i, q_i])}{2(1 - \sigma_1([p_i, q_i]))} & p'' > \sigma_1([p_i, q_i]) \end{cases}$$

    with $[p'', q''] = \arg\max_{[p,q] \in El \cap Wh} p$.

  if $[p_i, q_i] \in B$ **then**

    if $El \cap R - B \neq \emptyset$ **then**

$$[p_{i+1}, q_{i+1}] = \sigma^{-1}\left(\sigma_1([p_i, q_i])), \frac{\sigma_2([p_i, q_i]) + \alpha''}{2}\right)$$

    where

$$\alpha'' = \begin{cases} \frac{p''}{2\sigma_1([p_i, q_i])} & p'' \leq \sigma_1([p_i, q_i]) \\ \frac{p'' + 1 - 2\sigma_1([p_i, q_i])}{2(1 - \sigma_1([p_i, q_i]))} & p'' > \sigma_1([p_i, q_i]) \end{cases}$$

    with $[p'', q''] = \arg\min_{[p,q] \in El \cap R - B} p$.

    if $El \cap Wh \neq \emptyset$ **then**

$$[p_{i+1}, q_{i+1}] = \sigma^{-1}\left(\sigma_1([p_i, q_i])), \frac{\sigma_2([p_i, q_i]) + \alpha'}{2}\right)$$

    where

$$\alpha' = \begin{cases} \frac{q'}{2(1 - \sigma_1([p_i, q_i])} & q' \leq 1 - \sigma_1([p_i, q_i]) \\ \frac{q' - 1}{2\sigma_1([p_i, q_i])} + 1 & q' > 1 - \sigma_1([p_i, q_i]) \end{cases}$$

    with $[p', q'] = \arg\min_{[p,q] \in El \cap Wh} q$.

end

Figure D.4: Determination of Next Binary Choice: Part 2

and the definition of σ (and in particular the slopes of the lines $\sigma_1([p,q]) = m$ for $m > 0$), it follows that $\min_{[p,q] \in El_n} d([\underline{p}(E), \overline{p}(E)], [p,q])$ tends to 0 as $n \to 0$, whence $[\underline{p}_n, \overline{p}_n] \to$

34 **Case 3**

35 **if** $(El \cup \{[p_i, q_i]\}) \cap R - B \neq \emptyset$ and $(El \cup \{[p_i, q_i]\}) \cap Wh \neq \emptyset$ **then**

36 **if** $[p_i, q_i] \in R - B$ **then**

37
$$[p_{i+1}, q_{i+1}] = \left[\frac{p_i + p^{''}}{2}, \frac{q_i + q^{''}}{2}\right]$$
with $[p^{''}, q^{''}] = \arg\max_{[p,q]\in El\cap Wh} p$.

38 **if** $[p_i, q_i] \in Wh$ **then**

39
$$[p_{i+1}, q_{i+1}] = \left[\frac{p_i + p^{'}}{2}, \frac{q_i + q^{'}}{2}\right]$$
with $[p^{'}, q^{'}] = \arg\min_{[p,q]\in El\cap RB} p$.

40 **if** $[p_i, q_i] \in R$ **then**

41
$$[p_{i+1}, q_{i+1}] = \sigma^{-1}\left(\sigma_1([p_i, q_i]), \frac{\sigma_2([p_i, q_i]) + \min(\alpha', \alpha^{''})}{2}\right)$$
where
$$\alpha' = \begin{cases} \frac{q'}{2(1-\sigma_1([p_i,q_i])} & q' \leq 1 - \sigma_1([p_i, q_i]) \\ \frac{q'-1}{2\sigma_1([p_i,q_i])} + 1 & q' > 1 - \sigma_1([p_i, q_i]) \end{cases}$$
with $[p', q'] = \arg\max_{[p,q]\in El\cap R-B} q$ and
$$\alpha^{''} = \begin{cases} \frac{p^{''}}{2\sigma_1([p_i,q_i])} & p^{''} \leq \sigma_1([p_i, q_i]) \\ \frac{p^{''}+1-2\sigma_1([p_i,q_i])}{2(1-\sigma_1([p_i,q_i]))} & p^{''} > \sigma_1([p_i, q_i]) \end{cases}$$
with $[p^{''}, q^{''}] = \arg\max_{[p,q]\in El\cap Wh} p$.

42 **if** $[p_i, q_i] \in B$ **then**

43
$$[p_{i+1}, q_{i+1}] = \sigma^{-1}\left(\sigma_1([p_i, q_i]), \frac{\sigma_2([p_i, q_i]) + \max(\alpha', \alpha^{''})}{2}\right)$$
where
$$\alpha' = \begin{cases} \frac{q^{''}}{2(1-\sigma_1([p_i,q_i])} & q' \leq 1 - \sigma_1([p_i, q_i]) \\ \frac{q^{''}-1}{2\sigma_1([p_i,q_i])} + 1 & q' > 1 - \sigma_1([p_i, q_i]) \end{cases}$$
with $[p^{''}, q^{''}] = \arg\min_{[p,q]\in El\cap Wh} q$ and
$$\alpha^{''} = \begin{cases} \frac{p'}{2\sigma_1([p_i,q_i])} & p' \leq \sigma_1([p_i, q_i]) \\ \frac{p'+1-2\sigma_1([p_i,q_i])}{2(1-\sigma_1([p_i,q_i]))} & p^{''} > \sigma_1([p_i, q_i]) \end{cases}$$
with $[p', q'] = \arg\min_{[p,q]\in El\cap R-B} p$.

44 **end**

Figure D.5: Determination of Next Binary Choice: Part 3

$[\underline{p}(E), \overline{p}(E)]$ as required.

- $\sigma_1([\underline{p}(E),\overline{p}(E)]) > 0$ and $\sigma_1([p_i,q_i]) = \sigma_1([\underline{p}(E),\overline{p}(E)])$ for some i. By Lemma E.1 and Case 2 (lines 24-33) and Case 3 (lines 40-43), the procedure will, from i onwards, only pass through points with same σ_1-value $\sigma_1([\underline{p}(E),\overline{p}(E)])$, where it will only find points in R and B. Moreover, it follows a bisection-style procedure on the line $\sigma_1([p,q]) = \sigma_1([\underline{p}(E),\overline{p}(E)])$. It follows from standard arguments, Lemma E.1 and representation (1) that this procedure converges to $[\underline{p}(E), \overline{p}(E)]$ as required.

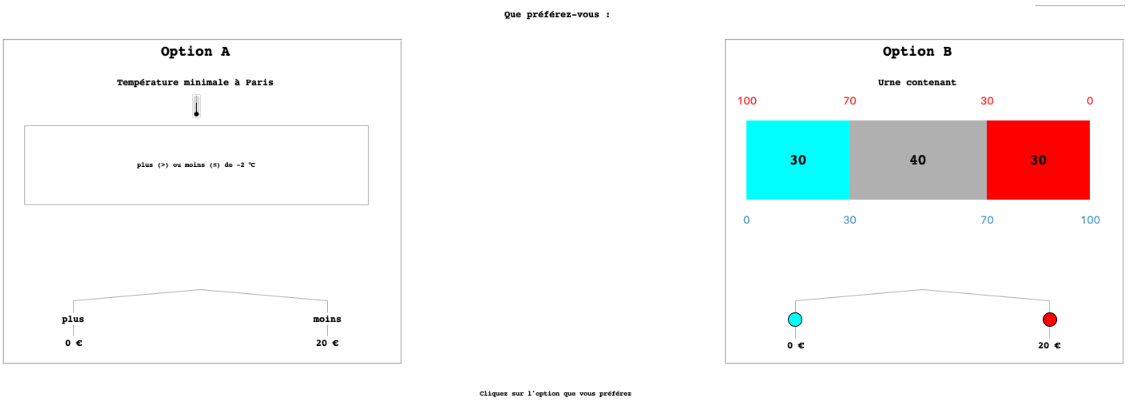


Figure D.6: Display for binary choices

- $\sigma_1([\underline{p}(E),\overline{p}(E)]) = 0$ and $\partial_1 W([\underline{p}(E),\overline{p}(E)]) > \partial_2 W([\underline{p}(E),\overline{p}(E)]) > 0$. Suppose $\underline{p}(E) = 0$; the other case ($\underline{p}(E) \neq 0$ and so $\overline{p}(E) = 1$) is treated similarly. By Lemma E.1, $[p_n, q_n]$ contains a subsequence of points in R-B, with σ_1-value tending to 0. Since $\partial_1 W([\underline{p}(E),\overline{p}(E)]) > \partial_2 W([\underline{p}(E),\overline{p}(E)]) > 0$, by representation (1), for every $q < \overline{p}(E)$, there exists $p > 0$ such that $(z,[p,q],0) \prec (z,E,0)$, and hence such that $[p,q]$ is not in R-B. Moreover, by the representation (and notably the fact that W is strictly increasing in the lower bound), for every $q > \overline{p}(E)$ and $p \geq 0$, $(0,[p,q],z) \prec (0,\left[\underline{p}(E),\overline{p}(E)\right],z) \sim (0,E,z)$, so such $[p,q]$ are not in R-B. It follows that the subsequence of $[p_n,q_n]$ consisting of points in R-B converges to $\left[\underline{p}(E),\overline{p}(E)\right]$, so $[p_n,q_n] \to \left[\underline{p}(E),\overline{p}(E)\right]$ as required.

□

D.2. Elicitation method: displays and details

Experimental material for the experiments is available here.[8] The entire EXP A, which was conducted in English, is online at the following address: behavioralexpe.shinyapps.io/expe2023/.[9] The instruction video for EXP A is available at the following address: youtu.be/dGlfpII8uBQ.

MPI Elicitation Figure D.6 shows the display for a typical choice in the binary-choice step of our elicitation method, in EXP N1. The presentation of the options in EXP N2 is similar. In EXP A, choices are presented as in Figure D.6, with the options drawn as in Figure 2. Note that, in that experiment, both the bag's label and the previous draws from it were presented on

[8]In case of issues with the link, the address is: `https://osf.io/yvax4/`.

[9]Please be aware that, whilst an identical program, the version used for the actual experiment was hosted on another server; there may thus be server issues with this version. Moreover, subject numbers were strictly controlled during the actual experiment: conflicts in subject numbers entered by users in this open-access version may cause you to skip parts of the study. We advise restarting with another subject number if this occurs.

screen throughout all tasks involving each bag. The physical bags (labeled as on the screen) were also present throughout the experiment.[10]

The display for the 2D choice list in EXPs N1 and N2 was as presented in Figure 2, with the bets displayed as in Figure D.6. Note that the red and blue lines above and below the scrollbar encode the preferences between bets on the target event and bets on urns. For instance, the red lines below then above the bar in Figure 2a indicate that, for an urn with at least 25 blue balls and a minimum number of red balls greater than 25, the bet on red from the urn is preferred over the bet on yellow from bag #C, whereas when there are at least 25 blue urns and the minimum number of red balls is less than 25, the bet on the bag is preferred. Similarly for the blue lines and bets on blue from the urn and green from the bag. Readers wishing to enter their preferences using the proposed method can undertake EXP A at the following address: behavioralexpe.shinyapps.io/expe2023/.[11]

In EXP N2, after having undertaken the elicitation tasks for all events in a source, subjects were presented with a final confirmation screen, shown in Figure D.7. All interval-valued urns corresponding to the choices made and confirmed by the subject for the source are presented on the left. They are graphically depicted on the right: the red line shows the minimum number of red balls for each event (mark, in the case of this source), whereas the blue line plots 100 minus the minimum number of blue balls. To change a choice, a subject can either click on the choice on the right hand plot or on the corresponding urn in the sidebar on the left. By doing so, she returns to the corresponding two-cursor scrollbar confirmation screen, as described above and shown in Figure 2. She may modify her choices on this screen as described previously, and must reconfirm before proceeding.

MP elicitation For the Paris treatment in EXP N1, the MP of the bet on a given event was elicited through a two-step procedure, as in Abdellaoui et al. (2021, 2023). First, a candidate MP was determined through a bisection process (Abdellaoui et al., 2008), consisting in a chained sequence of binary choices between the bet on the event and an urn whose composition was fully known. Starting with a binary choice between $(z,[\frac{1}{2},\frac{1}{2}],0)$ and $(z,E,0)$, it then asks a binary choice with the midpoint of the lower (respectively upper) interval $[0,\frac{1}{2}]$ (resp. $[\frac{1}{2},1]$) whenever the subject chooses the former (resp. latter) option, and so on. The displays used for these binary choices were similar to those used in the MPI elicitation (Figure D.6). In the second stage, the complete confirmation (one-dimensional, single cursor) scrollbar-based choice list, filled in according to the prior bisection choices, was displayed for verification. Figure D.8 presents the display for this part of the method. As for the MPI confirmation screen

[10] In EXP A, before undertaking the tasks concerning each bag, subjects were informed of prior draws from it on the screen. The (unknown) compositions of the physical bags were consistent with the reported draws.

[11] Following the instructions, which the reader can skip, the experiment begins with a practice task, involving both steps of our elicitation method.

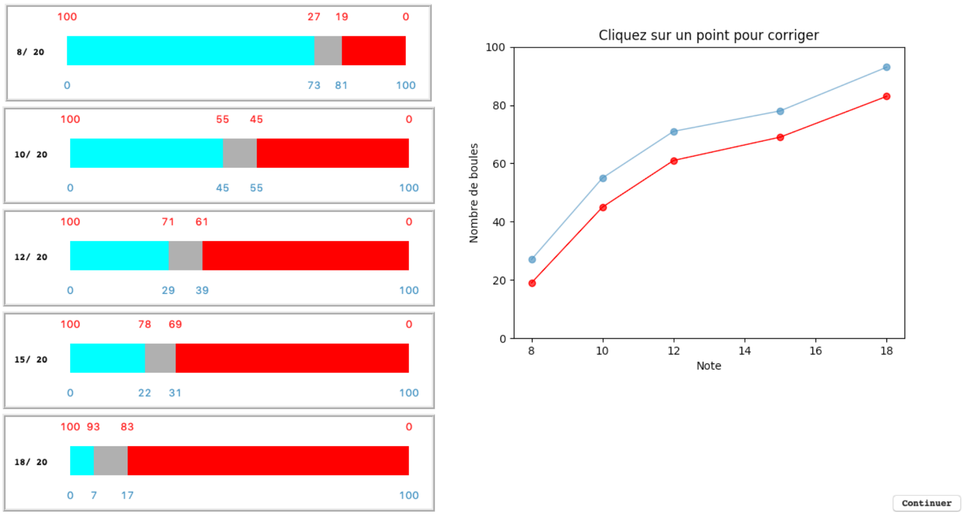


Figure D.7: Omnibus confirmation screen in EXP N2

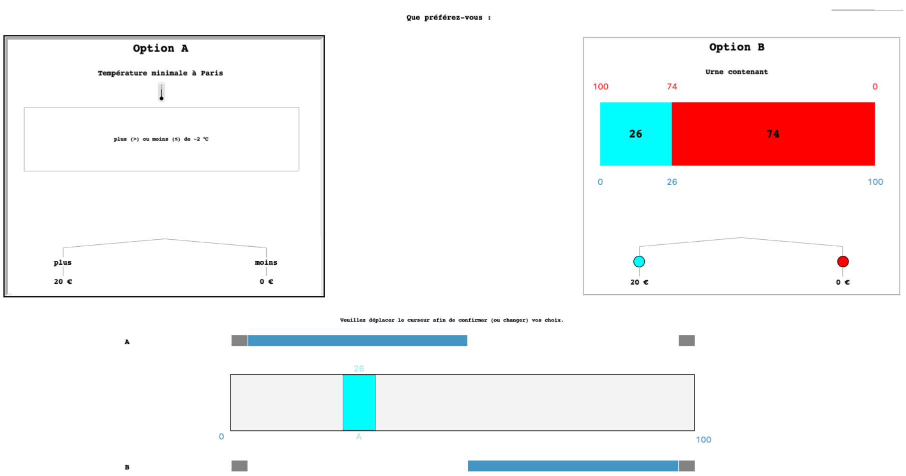


Figure D.8: MP confirmation choice list display in EXP N1

(Section 3.2), the subject may use the cursor to inspect and modify her choices, and must scan all choices before confirming. The precision of the elicited MP was to the nearest 0.05.

E. Technical Appendix: Lemmas for the proof of Proposition D.1

In the following Lemmas, we suppose that preferences are represented according to (1) with a unique MPI for the event of interest E and W differentiable with $\partial_1 W([\underline{p}(E),\overline{p}(E)]) > \partial_2 W([\underline{p}(E),\overline{p}(E)])$, where $[\underline{p}(E),\overline{p}(E)]$ is the subjective probability interval for E. Throughout this section, we adopt the notation set out in Section D.1.3.

Lemma E.1. *For every $m \in [0, 0.5]$:*

- *If $\sigma_1\left(\left[\underline{p}(E), \overline{p}(E)\right]\right) < m$, there exists $[p,q] \in R-B$ with $\sigma_1([p,q]) = m$, but no $[p,q] \in Wh$ with $\sigma_1([p,q]) = m$;*
- *If $\sigma_1\left(\left[\underline{p}(E), \overline{p}(E)\right]\right) > m$, there exists $[p,q] \in Wh$ with $\sigma_1([p,q]) = m$, but no $[p,q] \in R-B$ with $\sigma_1([p,q]) = m$;*
- *If $\sigma_1\left(\left[\underline{p}(E), \overline{p}(E)\right]\right) = m$, each $[p,q]$ with $\sigma_1([p,q]) = m$ and $[p,q] \neq \left[\underline{p}(E), \overline{p}(E)\right]$ belongs to either R or B.*

Proof. Straightforward to check from the representation (1) and the definition of σ (13). (See also Figures 1 and D.1.) □

Lemma E.2. *If $\underline{p}(E) = \overline{p}(E)$ and Case 1 arrives at a point $[p_i, q_i]$ with $p_i = q_i$, then the procedure remains in Case 1, and $[\underline{p}_n, \overline{p}_n] \to [\underline{p}(E), \overline{p}(E)]$ as $n \to \infty$.*

Proof. Once the procedure reaches the subspace of precise probabilities, it executes a standard bisection procedure (lines 13–16, Figure D.3). □

Lemma E.3. *Suppose that the procedure reaches a point $[p_i, q_i]$ in R-B or Wh. Then the sequence $\sigma_1([p_n, q_n]) \to \sigma_1\left(\left[\underline{p}(E), \overline{p}(E)\right]\right)$ as $n \to \infty$.*

Proof. Consider a stage i in the procedure where a point has just been found in R-B or Wh. So the area containing the MPI is Φ_i (Eq. (14)). For the sake of readability, we set $I_{p,q} = [p,q]$. Let

$$
\begin{aligned}
m_i^{Wh} &= \min \sigma_1(\Phi_i) \\
&= \sigma_1\left([\max\left\{p' : I_{p',q'} \in El_n^{Wh}\right\}, \min\left\{q' : I_{p',q'} \in El_n^{Wh}\right\}]\right) \\
m_i^{RB} &= \max \sigma_1(\Phi_i) \\
&= \begin{cases} \sigma_1\left([\min\left\{p'' : I_{p'',q''} \in El_n^{D}\right\}, \max\left\{q'' : I_{p'',q''} \in El_n^{D}\right\}]\right) & \text{if } El_n^D \neq \emptyset \\ 0.5 & \text{otherwise} \end{cases}
\end{aligned}
$$

and $|\Phi_i| = m_i^{RB} - m_i^{Wh}$. The latter is the maximum difference in σ_1 values across all pairs of points in Φ_i. In the first two subcases of Case 3 (lines 35-39), the next probability interval elicited is

$$
\begin{aligned}
I_{p_{i+1}q_{i+1}} = &\frac{1}{2}[\max\left\{p' : I_{p',q'} \in El_n^{Wh}\right\}, \min\left\{q' : I_{p',q'} \in El_n^{Wh}\right\}] \\
&+ \frac{1}{2}[\min\left\{p'' : I_{p'',q''} \in El_n^{D}\right\}, \max\left\{q'' : I_{p'',q''} \in El_n^{D}\right\}]
\end{aligned}
$$

In the second subcase of Case 2 (lines 22-23), where a point in Wh has been found, but no point in R-B, the next probability interval elicited is

$$I_{p_{i+1}q_{i+1}} = \frac{1}{2}\left[\frac{1}{2}\left(\begin{array}{l}\min\left\{p'': I_{p'',q''} \in El_n^{Wh}\right\} + \\ \max\left\{q'': I_{p'',q''} \in El_n^{Wh}\right\}\end{array}\right), \frac{1}{2}\left(\begin{array}{l}\min\left\{p'': I_{p'',q''} \in El_n^{Wh}\right\} + \\ \max\left\{q'': I_{p'',q''} \in El_n^{Wh}\right\}\end{array}\right)\right]$$
$$+\frac{1}{2}\left[\min\left\{p'': I_{p'',q''} \in El_n^{Wh}\right\}, \max\left\{q'': I_{p'',q''} \in El_n^{Wh}\right\}\right]$$

where $[\frac{1}{2}\left(\min\left\{p'': I_{p'',q''} \in El_n^{Wh}\right\} + \max\left\{q'': I_{p'',q''} \in El_n^{Wh}\right\}\right), \frac{1}{2}\left(\min\left\{p'': I_{p'',q''} \in El_n^{Wh}\right\}\right) + \max\left\{q'': I_{p'',q''} \in El_n^{Wh}\right\}]$ is the point on the diagonal of precise probabilities (i.e. degenerate probability intervals) that is closest to $[\min\left\{p'': I_{p'',q''} \in El_n^{Wh}\right\}, \max\left\{q'': I_{p'',q''} \in El_n^{Wh}\right\}]$ (it is on the downwards sloping 45° line from $[\min\left\{p'': I_{p'',q''} \in El_n^{Wh}\right\}, \max\left\{q'': I_{p'',q''} \in El_n^{Wh}\right\}]$). So this point has σ_1-value 0.5.

In first subcase of Case 2 (lines 20-21), where a point in R-B has been found, but no point in Wh, the next probability interval elicited, $I_{p_{i+1}q_{i+1}}$, is a $\frac{1}{2} - \frac{1}{2}$ mix of $[\min\left\{p'': I_{p'',q''} \in El_n^{D}\right\}, \max\left\{q'': I_{p'',q''} \in El_n^{D}\right\}]$ with

$$\begin{cases} \left[\begin{array}{l}\min\left\{p'': I_{p'',q''} \in El_n^{D}\right\} + \\ \max\left\{q'': I_{p'',q''} \in El_n^{D}\right\} - 1\end{array}, 1\right] & \text{if } \begin{array}{l}\min\left\{p'': I_{p'',q''} \in El_n^{D}\right\} \\ +\max\left\{q'': I_{p'',q''} \in El_n^{D}\right\}\end{array} > 1 \\ \left[0, \begin{array}{l}\min\left\{p'': I_{p'',q''} \in El_n^{D}\right\} + \\ \max\left\{q'': I_{p'',q''} \in El_n^{D}\right\}\end{array}\right] & \text{if } \begin{array}{l}\min\left\{p'': I_{p'',q''} \in El_n^{D}\right\} \\ +\max\left\{q'': I_{p'',q''} \in El_n^{D}\right\}\end{array} \leq 1 \end{cases}$$

which is the point on the upper boundary (with either lower bound for the probability interval 0 or upper bound 1) that is on the downwards sloping 45° line from $[\min\left\{p'': I_{p'',q''} \in El_n^{D}\right\}, \max\left\{q'': I_{p'',q''} \in El_n^{D}\right\}]$. This point has σ_1-value 0.

Clearly, in all cases, $m_i^{Wh} < \sigma_1(I_{p_{i+1}q_{i+1}}) < m_i^{RB}$. Moreover, by the rest of the subcases in Cases 2 & 3, if this point is not in R-B or Wh, all the subsequent points elicited will have the same σ_1-value as $I_{p_{i+1}q_{i+1}}$. And whenever a point in R-B is found, the next area containing the MPI, Φ_{i+1}, will have the same minimum σ_1-value m_i^{Wh}, but its maximum value will be replaced by $\sigma_1(I_{p_{i+1}q_{i+1}})$. By Lemma E.4, it follows that

$$|\Phi_i|.\frac{m_i^{Wh}}{m_i^{RB} + m_i^{Wh}} \leq |\Phi_{i+1}|$$
$$\leq |\Phi_i|.\frac{1 - m_i^{Wh}}{(1 - m_i^{RB}) + (1 - m_i^{Wh})}$$

Similarly, whenever a point in Wh is found, the next area containing the MPI, Φ_{i+1}, will have the same maximum σ_1 value m_i^{RB}, but its minimum value will be replaced by $\sigma_1(I_{p_{i+1}q_{i+1}})$, whence

$$|\Phi_i|.\frac{1-m_i^{RB}}{(1-m_i^{RB})+(1-m_i^{Wh})} \leq |\Phi_{i+1}| \leq |\Phi_i|.\frac{m_i^{RB}}{m_i^{RB}+m_i^{Wh}}$$

Since, for any $j > i$, $m_j^{RB} \leq m_i^{RB}$ and $m_j^{Wh} \geq m_i^{Wh}$, for any such j, $\frac{1-m_j^{Wh}}{(1-m_j^{RB})+(1-m_j^{Wh})} \leq \frac{1-m_i^{Wh}}{(1-m_i^{RB})+(1-m_i^{Wh})}$ and $\frac{m_j^{RB}}{m_j^{RB}+m_j^{Wh}} \leq \frac{m_i^{RB}}{m_i^{RB}+m_i^{Wh}}$. So, for any $j = i+k$ with $k \in \mathbb{N}$, $k \geq 1$, $|\Phi_j| \leq \left(\max\left\{\frac{1-m_i^{Wh}}{(1-m_i^{RB})+(1-m_i^{Wh})}, \frac{m_i^{RB}}{m_i^{RB}+m_i^{Wh}}\right\}\right)^k.|\Phi_i|$. So the sequence $[m_n^{Wh}, m_n^{RB}]$ is a bisection-like sequence of decreasing intervals (in the sense of containment), each of which contains $\sigma_1([\underline{p}(E), \overline{p}(E)])$. Moreover, by the previous observation, whenever a point $I_{p,q}$ is found in Wh with $\sigma_1(I_{p,q}) > 0$, then the sequence $|\Phi_n| = m_n^{RB} - m_n^{Wh} \to 0$ as $n \to \infty$, so $\sigma_1(I_{p_nq_n}) \to \sigma_1([\underline{p}(E), \overline{p}(E)])$ as required. (Recall that $0.5 \geq m_n^{RB} \geq m_n^{Wh} \geq 0$ for all n.)

We now separate two cases, according to whether $\sigma_1\left(\left[\underline{p}(E), \overline{p}(E)\right]\right) = 0$ or not. Suppose first that $\sigma_1\left(\left[\underline{p}(E), \overline{p}(E)\right]\right) = \delta > 0$. We show that the procedure will either arrive at a point with σ_1-value δ, or a point in Wh. At a stage i in the procedure where no points in Wh have been found, but a point in R-B has, $m_i^{Wh} = 0$ and $0.5 \geq m_i^{R-B} > 0$. At each subsequent stage, by Lemma E.1, either i. no point is found in Wh or R-B; ii. a point is found in Wh or R-B, and the next such point is in Wh; iii. a point is found in Wh or R-B, and the next such point is in R-B. In case ii., the claim is established; in case i., by Lemma E.1, the procedure is examining points with σ_1-value δ, and the claim is established. Assume for reductio that at all such stages, the σ_1-value of the explored points is not δ, but no point in Wh is found—i.e. we are always in case iii. Then, by the previous observations, for every $j = i+k$ with $k \in \mathbb{N}$, $k \geq 1$, $|\Phi_j| \leq \left(\frac{1-m_i^{Wh}}{(1-m_i^{RB})+(1-m_i^{Wh})}\right)^k.|\Phi_i| = \left(\frac{1}{2-m_i^{RB}}\right)^k.m_i^{RB}$. Hence $|\Phi_j| = m_j^{RB} \to 0$, contradicting the fact that there are no points with σ_1-value less that δ in R-B. Hence the procedure eventually finds a point in Wh. By the previous observation it follows that $\sigma_1([p_n, q_n]) \to \sigma_1([\underline{p}(E), \overline{p}(E)])$ as required.

Now consider the case where $\sigma_1\left(\left[\underline{p}(E), \overline{p}(E)\right]\right) = 0$. By Lemma E.1, whenever the procedure searches for a point on a line $\sigma_1(I_{p,q}) = m > 0$, it will find a point in R-B. Hence, by the previous argument, it produces a sequence of points $I_{p_nq_n}$ in R-B, defining Φ_n and associated $[m_n^{Wh}, m_n^{RB}]$, with $m_n^{Wh} = 0$ and $m_j^{RB} \to 0$, as required.

□

Lemma E.4. *Let $[p_{Wh}, q_{Wh}]$ be a point in Wh, with $\sigma_1([p_{Wh}, q_{Wh}]) = m_{Wh}$ and suppose that the* *line* $\sigma_1([p,q]) = m_{R-B}$ *contains a point in R-B but not in Wh. Then, for any point* $[p_{R-B}, q_{R-B}] \in R-B$ *with* $\sigma_1([p_{R-B}, q_{R-B}]) = m_{R-B}$

$$\sigma_1([\frac{p_{Wh+}p_{R-B}}{2}, \frac{q_{Wh}+q_{R-B}}{2}]) \in \left[\frac{2m_{Wh}.m_{R-B}}{m_{Wh}+m_{R-B}}, \frac{m_{Wh}(1-m_{R-B})+m_{R-B}(1-m_w)}{(1-m_{R-B})+(1-m_w)}\right]$$

Moreover, the same holds for a given point $[p_{R-B}, q_{R-B}] \in R-B$ *and any point* $[p_{Wh}, q_{Wh}] \in Wh$ *on the line* $\sigma_1([p,q]) = m_{Wh}$.

Proof. We begin by noting for reference that the inverse map of σ is given by:

$$\sigma^{-1}(m,\alpha) = \begin{cases} [2\alpha m, 2\alpha(1-m)] & \alpha \leq \frac{1}{2} \\ [(2-2\alpha)m + (2\alpha-1), (2-2\alpha)(1-m) + (2\alpha-1)] & \alpha > \frac{1}{2} \end{cases} \tag{15}$$

We first restrict attention to points $[p,q]$ with $p < 1-q$ (or, in the polar-style coordinate system, $\alpha < \frac{1}{2}$). For any points $[p_1, q_1]$ and $[p_2, q_2]$, written in polar-style coordinate system as (m_1, α_1) and (m_2, α_2), by (13) and (15), the midpoint (in Cartesian coordinates), $\frac{1}{2}[p_1, q_1] + \frac{1}{2}[p_2, q_2]$ is $\left(\frac{\alpha_1 m_1 + \alpha_2 m_2}{\alpha_1+\alpha_2}, \frac{\alpha_1+\alpha_2}{2}\right)$ in the polar system. Written in the polar coordinate system, let $[p_{Wh}, q_{Wh}]$ be (m_{Wh}, α_{Wh}); the points on the line $\sigma_1([p,q]) = m_{R-B}$ are (m_{R-B}, α), for varying α. Note that, by Proposition 1, $m_{R-B} > m_{Wh}$. It follows from representation 1 that $(z,[p',q'],0) \prec (z,[p,q],0)$ whenever $q' < q$ and $p' < p$, whence, since $[p_{Wh}, q_{Wh}] \in Wh$, we have that $(z,[p',q'],0) \prec (z,E,0)$ for all $q' < q_{Wh}$ and $p' < p_{Wh}$, so such points are not in R-B. So any point $[p,q]$ on $\sigma_1([p,q]) = m_{R-B}$ which is in R-B is such that $p \geq p_{Wh}$. By a similar argument (using the fact that $(0,[p',q'],z) \prec (0,E,z)$ for all $q' > q_{Wh}$ and $p' > p_{Wh}$), any point $[p,q]$ on $\sigma_1([p,q]) = m_{R-B}$ which is in R-B is such that $q \leq q_{Wh}$. So any point $[p,q]$ on $\sigma_1([p,q]) = m_{R-B}$ which is in R-B has $\alpha > \frac{\alpha_{Wh} m_{Wh}}{m_{R-B}}$ (where, by (15), this is in the α of the point on $\sigma_1([p,q]) = m_{R-B}$ with $p = p_{Wh} = 2\alpha_{Wh} m_{Wh}$); similarly, any such point has $\alpha < \frac{\alpha_{Wh}(1-m_{Wh})}{(1-m_{RB})}$. Plugging these bounds into the expression for the midpoint yields the result. Similar calculations yield the same result for the cases of $p > 1-q$ for some or all of the point considered. Finally, analogous arguments establish the conclusion for $[p_{R-B}, q_{R-B}] \in R-B$ fixed and $[p_{Wh}, q_{Wh}] \in Wh$ on the line $\sigma_1([p,q]) = m_{Wh}$. □