

Are Ambiguity Attitudes Transferable?

A Test Using DOSE*

Sijia Zhang[†]

Alexander L. Brown[‡]

October 2026

[Click here for the latest version](#)

Abstract

We examine whether model-specific ambiguity-attitude parameters are stable across situations with varying ambiguity and can therefore support out-of-sample prediction. Under Source Theory, ambiguity attitudes are simply a transformation of first-order beliefs. The smooth ambiguity model and modified Hurwicz expected utility (HEU-B) additionally allow for confidence in these beliefs, i.e., second-order beliefs, which should make ambiguity attitudes more transferable. Approximating second-order beliefs in our benchmark experiment, we find initial support: estimated attitudes change substantially across situations under Source Theory but less so under the two alternatives. For a more complete test, the main experiment adapts Dynamically Optimized Sequential Experimentation (DOSE) to elicit second-order beliefs and ambiguity attitudes separately. We compare the models' out-of-sample predictions across situations. HEU-B outperforms the other two models, while the smooth ambiguity model does not outperform Source Theory. Thus, accounting for second-order beliefs does not by itself improve prediction; how the model incorporates them also matters.

JEL Classifications: D80; D81; D83.

Keywords: Ambiguity; Smooth ambiguity models; source theory; Hurwicz expected utility.

*We are grateful to Marco Casari, Marco Castillo, Enrico Diecidue, Catherine Eckel, Taisuke Imai, Yassine Kaouane, Emmanuel Kemel, Silvana Krasteva, Melvin Marti, Yusufcan Masatlioglu, and John Smith for helpful comments. We also thank participants at the 2026 Foundations of Utility and Risk (FUR) Conference, the 2026 Economic Science Association (ESA) Africa Conference, and the 2026 Caltech Summer Workshop in Theory-Based Experiments for valuable feedback. This research was partially funded by NSF grant #2125985. Computational support was provided by Texas A&M High Performance Research Computing.

[†]Department of Economics, Texas A&M University, College Station, TX 77843, United States. E-mail: teresasi-jia@tamu.edu.

[‡]Department of Economics, Texas A&M University, College Station, TX 77843, United States. Email: alexbrown@tamu.edu

1 Introduction

Ambiguity attitudes have been shown to matter for a wide variety of economically important decisions, including technology adoption (Barham et al., 2014; Bryan, 2019; Jozwik, 2026), insurance pricing (Hogarth and Kunreuther, 1989; Cabantous, 2007; Alary et al., 2013; Cecchi et al., 2024), health care decisions (Berger et al., 2013; Attema et al., 2018), portfolio choice (Dimmock et al., 2016b; Bianchi and Tallon, 2019) and stock-market participation (Dimmock et al., 2016a; Antoniou et al., 2015; Anantanasuwong et al., 2024). These results suggest that ambiguity models may be a promising addition to an empirical economist’s toolkit for predicting human behavior. However, nearly all existing literature compares ambiguity models’ predictive performance within the same or a similar ambiguous situation (Hey et al., 2010; Conte and Hey, 2013; Kothiyal et al., 2014; Hey and Pace, 2014; Qiu and Weitzel, 2016; Peysakhovich and Naecker, 2017; Georgalos, 2019). A more rigorous test would require a model’s estimated parameters to retain predictive value when the level of ambiguity changes. This paper provides the first examination of whether an ambiguity model calibrated in one ambiguous situation can predict behavior in another with a different level of ambiguity. We further ask which model performs best in such cross-situation prediction.

The importance of this question is straightforward. People rarely face exactly the same ambiguous situation twice. Empirical models of human behavior that incorporate ambiguity must be able to accommodate repeated observations by the same agent. This is particularly relevant in financial markets where conditions can change quickly and substantially (Brenner and Izhakian, 2018; Williams, 2015; Drechsler, 2013). How a model should do so is not so clear. Any model must specify how agents form beliefs over subjective uncertainty. This specification implicitly affects how such beliefs will be revised when moving from one ambiguous situation to another.

Different models make different assumptions about how beliefs are formed. We focus on three capable of prediction across different levels of ambiguity: Source Theory (Abdellaoui et al., 2011; Baillon et al., 2025), the smooth ambiguity model (Klibanoff et al., 2005), and a modified version of Hurwicz expected utility (Gul and Pesendorfer, 2015; Bleichrodt et al., 2023), henceforth HEU-B. Source Theory requires only a first-order belief, or expectation, over subjective uncertainty. In contrast, the smooth ambiguity model and HEU-B also incorporate second-order beliefs, capturing uncertainty about that expectation. The two models differ in how they represent this second-order belief. The smooth ambiguity model represents belief as a distribution over possible first-order beliefs, whereas HEU-B represents it as a range bounded by pessimistic and optimistic beliefs.

To illustrate these distinctions, consider a box containing 100 red, blue, and yellow balls with an unknown composition. After four draws with replacement, a decision maker may form a best guess about the box’s composition while remaining quite uncertain about it. Suppose instead that the DM observes twenty draws from a different box with the same color proportions as before. The DM may still form the same best guess, but would have greater confidence that the true composition of the box lies close to it. A *first-order belief* is this best guess, represented by one probability for each color. A *second-order belief* also captures how confident the decision maker is in that assessment, which we refer to as the decision maker’s *perceived ambiguity level*.

Our experimental environments fit closely with this example. In our initial, benchmark study, subjects make decisions under uncertainty with different amounts of information available. Specifically, they observe two and sixteen draws with replacement from two boxes of unknown composition, respectively. We elicit first-order beliefs and ambiguity attitudes to calibrate Source Theory. The estimated ambiguity-attitude parameters change significantly and substantially across the two information states (similar to Abdellaoui et al., 2025), which under Source Theory implies that ambiguity attitudes are not stable across situations. An alternative explanation is that attitudes are stable, but perceived ambiguity falls under high information even when first-order beliefs remain unchanged. Indeed, when we proxy for second-order beliefs using a Bayesian assumption, the resulting smooth ambiguity and HEU-B models produce more stable ambiguity attitudes across the two situations than Source Theory.

While these results provide suggestive evidence in favor of models that separately account for perceived ambiguity, a more complete test requires separately measuring second-order beliefs and ambiguity attitudes for each subject. There are limitations to the feasibility of this approach. Standard elicitation would require several hundred questions even with only two situations. As an alternative, we apply Dynamically Optimized Sequential Experimentation (DOSE; Chapman et al., 2025a), previously used with risk, loss aversion, and time preference, to belief and ambiguity-attitude elicitation. For a given model, DOSE selects the most informative next question for parameter estimation based on the subject’s previous responses. Our “DOSE-belief” and “DOSE-ambiguity” procedures therefore allow us to measure beliefs and ambiguity attitudes separately with a manageable number of questions.

In our main study, subjects again make decisions between two states that vary only in the level of information: this time, observing either exactly four or twenty draws. In the low-information state, our newly developed DOSE-belief procedure elicits each subject’s first- and second-order

beliefs over the box composition. Using these beliefs, DOSE-ambiguity estimates the subject’s underlying ambiguity attitudes. Because DOSE-ambiguity requires a model to guide question selection, subjects are randomly assigned to one of three treatments based on Source Theory, the smooth ambiguity model, or HEU-B. Thus, one of the three models is interactively calibrated to each subject in a between-subjects design. In the high-information state, we reuse DOSE-belief to elicit first- and second-order beliefs but do not run DOSE-ambiguity. We then combine the ambiguity attitude estimated in the low-information state with the beliefs elicited in the high-information state to make out-of-sample predictions for the held-out ambiguity questions.¹ The held-out questions are selected adaptively by a model-independent procedure that targets each subject’s indifference point, where the models’ predictions differ most.

Our main result shows that HEU-B provides better out-of-sample prediction than both Source Theory and the smooth ambiguity model. In our preregistered between-subject comparison, we benchmark each model’s squared prediction error against a coin flip that assigns equal probability to either choice.² Relative to this benchmark, HEU-B reduces the median mean squared prediction error across subjects by 20.4%, compared with 16.7% under Source Theory and 10.1% under the smooth ambiguity model.

We provide three additional analyses. First, no model dominates at the subject level: the smooth ambiguity model is best for 32.1% of subjects, and HEU-B and Source Theory each around 34%. HEU-B’s advantage is how rarely it ranks worst: 7.7% of subjects, compared with 44.8% for the smooth ambiguity model and 47.5% for Source Theory. Second, HEU-B remains the best-performing model whether a subject’s perceived ambiguity falls, as expected, or rises from the low- to the high-information state. Its advantage over Source Theory is smaller when perceived ambiguity rises, while the smooth ambiguity model ties with Source Theory when perceived ambiguity

¹It is important to distinguish out-of-sample prediction from model fitting. In sample, a nested model with more parameters will always fit at least as well on an unpenalized likelihood measure because it has additional degrees of freedom. This need not hold out of sample, where additional parameters can overfit the estimation data and reduce predictive accuracy. In our case, all three models carry two ambiguity-attitude parameters across situations. Under Source Theory, however, these parameters must also capture any effect of perceived ambiguity not contained in the first-order belief. If perceived ambiguity changes across situations, the estimated ambiguity-attitude parameters may therefore change with it and transfer poorly to a new situation. The smooth ambiguity model and HEU-B instead allow perceived ambiguity to enter through beliefs. Whether this improves prediction depends on whether the additional structure these models impose describes behavior well. Thus, it is not clear *ex ante* which model should best predict behavior in a different situation.

²The coin-flip benchmark is demanding in our setting because the held-out questions converge on each subject’s point of indifference, where choices are harder to predict. The absolute improvement over a coin flip therefore understates how much predictive content the models have; the relative performance across models is more informative for our purposes.

falls but performs significantly worse when it rises.³ Finally, calibrating all three models on the same subjects within each treatment shows that in-sample fit always favors the model that guided DOSE, but leaves the out-of-sample ranking unchanged.

Across all results, whether main analyses or robustness checks, HEU-B is the outright winner of this experiment. Moreover, its gain in predictive performance over Source Theory suggests that ambiguity attitudes can be transferable across different levels of ambiguity and retain predictive power out of sample. The qualification is the smooth ambiguity model, which carries a second-order belief yet predicts no better than Source Theory, which carries none. Thus, accounting for perceived ambiguity alone is not sufficient to improve prediction; how the model incorporates perceived ambiguity also matters. With second-order beliefs imputed by Bayes' rule, our benchmark experiment could not distinguish HEU-B from the smooth ambiguity model; both outperformed Source Theory. The main experiment separates them by directly eliciting second-order beliefs with DOSE and evaluating predictions on held-out questions placed where the models differ.

Our paper ties together three strands of literature on ambiguity attitudes. First, several papers compare the explanatory power of ambiguity models (Hey et al., 2010; Conte and Hey, 2013; Ahn et al., 2014; Kothiyal et al., 2014; Hey and Pace, 2014; Qiu and Weitzel, 2016; Peysakhovich and Naecker, 2017; Georgalos, 2019).⁴ Notably, Kothiyal et al. (2014) and Georgalos (2019) evaluate models based on out-of-sample performance, but still within closely related ambiguous situations.⁵ These studies reach different conclusions about which model performs best, and none test whether estimated ambiguity-attitude parameters retain predictive value when the situation itself changes. We add this missing dimension: whether ambiguity-attitude parameters estimated in one situation predict choices in another.

A second strand provides evidence that estimated ambiguity attitudes vary systematically with information or experience (Baillon et al., 2012; Ert and Trautmann, 2014; Baillon et al., 2018a; Klingebiel and Zhu, 2023). Most closely related, Abdellaoui et al. (2025) vary the number of draws observed from an ambiguous urn and find that model-free measures of ambiguity attitude move

³In both the smooth ambiguity model and HEU-B, differences in ambiguity attitudes matter less (more) when perceived ambiguity is low (high), so attitudes should be estimated more precisely when perceived ambiguity is higher. Parameters estimated under higher perceived ambiguity may therefore predict behavior under lower perceived ambiguity better than the reverse. Source Theory lacks this feature, and the results are broadly consistent with this intuition.

⁴A related experimental literature tests specific behavioral features that can distinguish among ambiguity models (Halevy, 2007; Baillon and Bleichrodt, 2015; Chew et al., 2017; Cubitt et al., 2020; Brown et al., 2026).

⁵Kothiyal et al. (2014), for instance, hold out some choices for evaluation, but both the estimation and held-out questions come from the same situation. Georgalos (2019) is more closely related to our work, comparing out-of-sample performance after additional information is provided. However, the information concerns the same underlying ambiguous source, and predictions depend jointly on the ambiguity model and a belief-updating rule.

toward neutrality as information increases. They also find systematic changes in ambiguity-attitude parameters estimated under the smooth ambiguity model.⁶ We ask whether this instability reflects changes in ambiguity attitudes or the model used to estimate them, by comparing models that do and do not separate perceived ambiguity from attitude. The final strand examines belief updating under ambiguity (Cohen et al., 2000; Dominiak et al., 2012; Moreno and Rosokha, 2016; Ngangoué, 2021; Bland and Rosokha, 2021; Aydogan, 2021; Aydogan et al., 2025) and documents systematic departures from Bayesian updating. We elicit beliefs directly in each situation rather than imposing an updating rule, and impose Bayesian updating only to approximate second-order beliefs in our benchmark experiment and as a robustness check of our main results.

Our paper also contributes to the literature on adaptive preference elicitation. Adaptive procedures have been developed to estimate preferences and discriminate among competing behavioral models (Cavagnaro et al., 2013a,b, 2016; Toubia et al., 2013; Bertani et al., 2026; Chapman et al., 2025a). In ambiguity experiments, bisection procedures are commonly used to elicit ambiguity attitudes (Dimmock et al., 2016a; Li et al., 2019; Abdellaoui et al., 2025; Gonzalez-Jimenez, 2026). We extend Dynamically Optimized Sequential Experimentation (DOSE; Chapman et al., 2025a) to individual-level elicitation of second-order beliefs and ambiguity attitudes.⁷ Our results also have implications for using DOSE in model comparison. Model-guided question selection affects in-sample comparisons, but out-of-sample rankings remain stable when competing models are evaluated using model-independent held-out questions. Thus, models can be compared out of sample even when DOSE question selection is guided by only one of them.

The paper proceeds as follows. Section 2 introduces the three main ambiguity models (Source Theory, the smooth ambiguity model, and HEU-B) and describes the elicitation and estimation procedures. Section 3 presents the designs of the benchmark and main experiments. Section 4 reports the results from the benchmark experiment, and Section 5 reports the results from the main experiment. Section 6 concludes the paper.

⁶Bleichrodt et al. (2023) provide a tractable source-theory specification, motivated by Hurwicz expected utility, that separates an ambiguity-attitude parameter from a source-specific parameter capturing perceived ambiguity. Cross-situation prediction additionally requires specifying how perceived ambiguity changes with beliefs; we provide this mapping.

⁷DOSE has been applied to elicit loss aversion, time preferences, preferences for autonomy, and charitable giving; see Chapman et al. (2025b), Imai and Camerer (2018), Freundt et al. (2023), and Mitani and Hanaki (2025).

2 Models and methods

In this section, we introduce the three ambiguity models considered in our main comparison: the Prelec specification of source theory (Source-Prelec), the smooth ambiguity model, and HEU-B, a modification of Hurwicz expected utility. We focus on how they differ in their representation of beliefs and what these differences imply for cross-situation prediction. We then describe our elicitation and estimation procedures.

2.1 Setup and matching probabilities

Throughout this section, we use an ambiguous box containing 100 balls with unknown composition to illustrate the models. Each ball is red (R), yellow (Y), or blue (B). We consider bets on the color of a ball randomly drawn from this ambiguous box. Let E denote an ambiguous event defined by the color of the drawn ball. In our setting, there are six such events: the three single-color events R , Y , and B , and the three two-color events RY , RB , and YB . For example, R denotes the event that the ball drawn from the ambiguous box is red, while RY denotes the event that it is either red or yellow.

We restrict attention to two-outcome prospects that pay a fixed prize $x > 0$ if the indicated event occurs and zero otherwise. We consider two types of prospects. For a risky prospect, the probability of receiving x is objectively known. For example, a risky prospect could pay x if a red ball is drawn from a box known to contain 60 red balls out of 100. We write (x, p) for a risky prospect that pays x with probability p and zero otherwise. For an ambiguous prospect, the probability of the indicated event is unknown. For example, an ambiguous prospect may pay x if a red ball is drawn from a box whose composition is unknown. We write (x, E) for an ambiguous prospect that pays x if the ambiguous event E occurs and zero otherwise.

We consider choices between a risky prospect (x, p) and an ambiguous prospect (x, E) that offer the same prize x . Under specific assumptions,⁸ utility over monetary outcomes does not affect the tradeoff between the two prospects. The comparison can therefore be expressed entirely on a probability scale. For the risky prospect, the relevant quantity is the objective probability p . The ambiguous prospect can be represented on the same scale through its matching probability. For an ambiguous event E , the matching probability $m(E)$ is the known probability that makes the

⁸The relevant assumptions differ across models and are specified below.

decision maker indifferent between the two prospects:

$$(x, m(E)) \sim (x, E).$$

Thus, $m(E)$ is the objective probability in the risky prospect at which the decision maker is indifferent between that prospect and betting on E .

Each ambiguity model introduced below provides a prediction for $m(E)$ as a function of the decision maker’s beliefs and ambiguity-attitude parameters. We express all three models in terms of $m(E)$ and examine how differences in their components and structure translate into different predictions across situations.

2.2 Ambiguity models and cross-situation prediction

We compare three ambiguity models: source theory, the smooth ambiguity model, and HEU-B. A key difference among them is how they represent beliefs under ambiguity. This distinction is particularly relevant for cross-situation prediction. Specifically, models differ in whether changes in how uncertain the decision maker is about an ambiguous situation are represented through beliefs or instead absorbed into ambiguity-attitude parameters. In the latter case, the estimated ambiguity attitudes may become more situation-specific and therefore less transferable to a new situation.

To describe this distinction across models, we define first-order and second-order beliefs. We use the term first-order belief to refer to a single probabilistic assessment of the ambiguous situation. In the box example, this can be represented by a probability assigned to each color. A second-order belief additionally contains information about how sure the decision maker is about that assessment.⁹ For example, a decision maker may assign the same probability to drawing a red ball after observing four draws and after observing twenty draws, implying the same first-order belief about that event, but be more sure after twenty draws that the assigned probability is close to the true proportion of red balls in the box. We refer to this additional “how sure” dimension as the decision maker’s perceived ambiguity level. It is captured by the second-order belief but not by the first-order belief. Source theory uses only the first-order belief, whereas the smooth ambiguity model and HEU-B incorporate this additional belief dimension in different ways.

⁹Our use of the term second-order belief is intended to distinguish this additional belief dimension from the first-order probabilistic assessment and may not coincide with every use of the term in the literature.

2.2.1 Source theory

We start with source theory (Abdellaoui et al., 2011; Baillon et al., 2025). Source theory represents beliefs about an ambiguous source using only a first-order belief, represented by a single subjective probability distribution. In our box example, the decision maker assigns a subjective probability p_C to each color $C \in \{R, Y, B\}$, with $p_R + p_Y + p_B = 1$. Let $P(\cdot)$ denote the probability function that maps each ambiguous event to its assigned subjective probability. For example, $P(R) = p_R$. Source theory captures ambiguity attitudes through a source-specific weighting function w_S , which maps these subjective probabilities into matching probabilities, placing ambiguous prospects on the same probability scale as risky prospects. Under the assumptions described in Section 2.1,¹⁰ the matching probability for event E is therefore

$$m(E) = w_S(P(E)).$$

Throughout the main text, we use the Prelec specification,

$$w_S(q) = \exp(-\beta(-\ln q)^\alpha), \quad \alpha, \beta > 0.$$

The corresponding matching probability is

$$m(E) = \exp(-\beta(-\ln P(E))^\alpha). \tag{1}$$

The parameters α and β capture the curvature and elevation of the ambiguity weighting function, respectively. When $\alpha = \beta = 1$, the decision maker is ambiguity neutral. In this case, the weighting function reduces to the identity and $m(E) = P(E)$. We refer to this specification as Source-Prelec. We also consider alternative source theory specifications, including Goldstein–Einhorn and neo-additive weighting functions, in the appendix; the main results are robust to these alternatives.

Source theory uses only the first-order belief $P(E)$ and has no separate belief component capturing the decision maker’s perceived ambiguity level. Conditional on $P(E)$, any systematic effect of perceived ambiguity on behavior, if present, must therefore be reflected in the shape of the source weighting function and hence in the estimated ambiguity-attitude parameters. If perceived ambi-

¹⁰For simplicity, suppose the decision maker is an expected-utility maximizer under risk and normalize $u(0) = 0$. An ambiguous prospect (x, E) is then valued as $V(x, E) = w_S(P(E))u(x)$, while a risky prospect (x, p) is valued as $V(x, p) = pu(x)$. Hence, $(x, m(E)) \sim (x, E)$ implies $m(E) = w_S(P(E))$. The same logic applies if risky prospects are evaluated using a probability-weighting function w_R ; in that case, ambiguity attitudes are represented through the transformation $w_R^{-1} \circ w_S$. See Baillon et al. (2025).

guity changes across situations, these parameters may therefore absorb situation-specific variation and become less transferable to a new situation.

2.2.2 Smooth ambiguity

We next consider the smooth ambiguity model (Klibanoff et al., 2005), which incorporates a second-order belief. In our box example, the decision maker considers multiple possible compositions of the box and assigns a subjective likelihood to each. Let Π denote this distribution over possible compositions, and let p_E denote the probability of event E under a particular composition. Following Abdellaoui et al. (2025), we use the average probability $\mathbb{E}_\Pi[p_E]$ to represent the first-order belief about event E . The full distribution Π , however, contains richer belief information and also captures the decision maker’s perceived ambiguity.

The smooth ambiguity model uses an increasing function ϕ to capture ambiguity attitudes. On the probability scale introduced above,¹¹ the matching probability for event E is

$$m(E) = \phi^{-1}(\mathbb{E}_\Pi[\phi(p_E)]). \quad (2)$$

Following Abdellaoui et al. (2025), we use the exponential-power specification

$$\phi(q) = \frac{1 - e^{-\theta q^\rho}}{1 - e^{-\theta}}.$$

The parameters θ and ρ determine the shape of ϕ and capture ambiguity attitudes. When $\theta \rightarrow 0$ and $\rho = 1$, the decision maker is ambiguity neutral. In this case, $\phi(q) = q$ and

$$m(E) = \mathbb{E}_\Pi[p_E],$$

which is exactly the first-order belief.

Unlike source theory, the smooth ambiguity model represents perceived ambiguity separately from ambiguity attitudes: Π captures the second-order belief, including perceived ambiguity, while ϕ captures ambiguity attitudes. When perceived ambiguity changes across situations, the model can therefore accommodate this change through the belief while holding the ambiguity-attitude

¹¹Suppose $u(0) = 0$ and normalize $u(x) = 1$. Under the smooth ambiguity model, the ambiguous prospect (x, E) is valued as

$$\phi^{-1}(\mathbb{E}_\Pi[\phi(p_E u(x) + (1 - p_E)u(0))]).$$

With this normalization, the expression reduces to $\phi^{-1}(\mathbb{E}_\Pi[\phi(p_E)])$, giving the matching probability above.

parameters fixed. This separation may make the estimated ambiguity-attitude parameters more transferable across situations. At the same time, incorporating the second-order belief requires the model to impose a particular structure on how beliefs and ambiguity attitudes jointly determine behavior. Specifically, how strongly ambiguity attitudes affect behavior depends on the second-order belief. As the second-order belief becomes more concentrated and perceived ambiguity falls, the influence of ambiguity attitudes on behavior becomes smaller. When the decision maker is certain about the box composition, Π assigns all probability to a single composition and $m(E) = p_E$, regardless of the shape of ϕ . In this case, ambiguity attitudes have no effect on behavior even if they are non-neutral.

2.2.3 Hurwicz expected utility and HEU-B

We next consider Hurwicz expected utility (HEU) (Gul and Pesendorfer, 2015), which belongs to the α -maxmin family. Under this framework, the decision maker considers a set of possible probabilities for an ambiguous event. When evaluating a prospect, only the most pessimistic and most optimistic probabilities in this set matter. For an event E , let $\underline{p}_E$ and $\bar{p}_E$ denote these lower and upper bounds. The parameter $\alpha \in [0, 1]$ determines the weight placed on the pessimistic belief, while $1 - \alpha$ is placed on the optimistic belief. On the probability scale introduced above, the matching probability can therefore be written as¹²

$$m(E) = \alpha \underline{p}_E + (1 - \alpha) \bar{p}_E. \quad (3)$$

HEU further imposes structure on how these lower and upper bounds are formed. Motivated by probabilistic sophistication within a source, as in source theory, the decision maker assigns a single subjective probability $P(E)$ to each ambiguous event. HEU then uses this probability to construct the lower and upper bounds that enter the decision rule. Specifically,

$$\underline{p}_E = \gamma(P(E)), \quad \bar{p}_E = 1 - \gamma(1 - P(E)), \quad (4)$$

¹²With $u(0) = 0$ and a normalization of $u(x) = 1$, an ambiguous prospect (x, E) is valued as

$$V(x, E) = \alpha [\underline{p}_E u(x) + (1 - \underline{p}_E) u(0)] + (1 - \alpha) [\bar{p}_E u(x) + (1 - \bar{p}_E) u(0)],$$

which gives the matching-probability representation in Equation (3).

where

$$\gamma(t) = \sum_{n=1}^{\infty} a_n t^n, \quad (5)$$

and the coefficients $a_1, a_2, \dots$ are nonnegative and sum to one. This form guarantees that the resulting lower and upper bounds satisfy desirable properties. In particular, $P(E)$ always lies between $\underline{p}_E$ and $\bar{p}_E$. If $P(E) = 0$, both bounds are zero, while if $P(E) = 1$, both bounds are one.

The general HEU representation allows γ to depend on the sequence of parameters $a_1, a_2, \dots$. Bleichrodt et al. (2023) consider a particular family of γ in which this sequence can be indexed by a single parameter β . Under this specification,

$$\gamma_{\beta}(p) = \frac{(1 - \beta)p}{(1 + \beta)(1 - p) + (1 - \beta)p}. \quad (6)$$

Together with α , this gives a two-parameter mapping from the single subjective probability $P(E)$ to the matching probability. For a fixed source, it can therefore also be represented as a particular source theory weighting specification. A key appeal of this parameterization is that its two parameters have distinct economic interpretations. The parameter α captures how the decision maker weights the pessimistic relative to the optimistic belief, while β determines the width of the belief range and is interpreted as perceived ambiguity. A larger β generates a wider gap between the lower and upper bounds.

This interpretation suggests that β may vary when the ambiguous situation changes, even if the underlying attitude represented by α remains stable. For cross-situation prediction, however, knowing that β may change is not sufficient: we also need a way to determine how it changes with beliefs. Without such a mapping, the perceived-ambiguity parameter cannot be updated from the belief information available in a new situation.

Suppose information about how uncertain the decision maker is is summarized by a second-order belief distribution Π over possible compositions. Let s measure the spread of this distribution. We assume that belief spread maps directly into the perceived-ambiguity parameter β , which in turn determines the gap between the pessimistic and optimistic beliefs. Specifically,

$$\beta = \tanh\left(\frac{\kappa s}{2}\right), \quad (7)$$

where κ is the parameter linking belief spread to perceived ambiguity.¹³

¹³Under the specification in Equation (6), the lower and upper bounds are symmetric around $P(E)$ in log-odds space. Each bound is shifted by $2 \operatorname{artanh}(\beta)$ from $P(E)$, so the total distance between the two bounds is $4 \operatorname{artanh}(\beta)$.

In our three-color setting, we define

$$s = \frac{1}{3} \sum_{c \in \{R, Y, B\}} \text{SD}_{\Pi} \left[\log \frac{p_{E_c}}{1 - p_{E_c}} \right]. \quad (8)$$

The expression is written using the three single-color events because each two-color event is the complement of a single-color event and therefore has the same standard deviation in log odds as its complement. Averaging over all six events would consequently give the same value of s . We refer to this specification as HEU-B.

HEU-B predicts the matching probability as

$$m(E) = \alpha \gamma_{\beta}(P(E)) + (1 - \alpha) \hat{\gamma}_{\beta}(P(E)), \quad (9)$$

where

$$\hat{\gamma}_{\beta}(p) = 1 - \gamma_{\beta}(1 - p), \quad \beta = \tanh\left(\frac{\kappa s}{2}\right).$$

The parameter α governs the relative weight placed on the pessimistic and optimistic beliefs, while κ governs the relationship between belief spread and perceived ambiguity. Within the α -maxmin component of the model, $\alpha = 0.5$ is the ambiguity-neutral benchmark, corresponding to equal weight on the pessimistic and optimistic beliefs.

Like the smooth ambiguity model, HEU-B imposes structure on how perceived ambiguity affects the behavioral influence of ambiguity attitudes. As belief spread s becomes smaller, β falls and the pessimistic and optimistic bounds move closer together. The influence of ambiguity attitudes on the matching probability therefore becomes smaller. In the limiting case in which perceived ambiguity vanishes, $\beta = 0$, the two bounds coincide with $P(E)$ and hence $m(E) = P(E)$ regardless of α . Thus, even a non-neutral ambiguity attitude has no effect on behavior when there is no perceived ambiguity.

To provide an intuitive illustration of the consequences of incorporating this additional belief dimension, Figure 1 compares the mappings from first-order belief $P(E)$ to matching probability $m(E)$ implied by the three models. In each panel, the ambiguity-attitude parameters are held fixed. Under Source-Prelec, perceived ambiguity does not enter separately, so the mapping remains unchanged when perceived ambiguity changes. Under the smooth ambiguity model and HEU-B, by contrast, the mapping can change with perceived ambiguity even when the first-order belief and

Equation (7) therefore links this distance directly to belief spread.

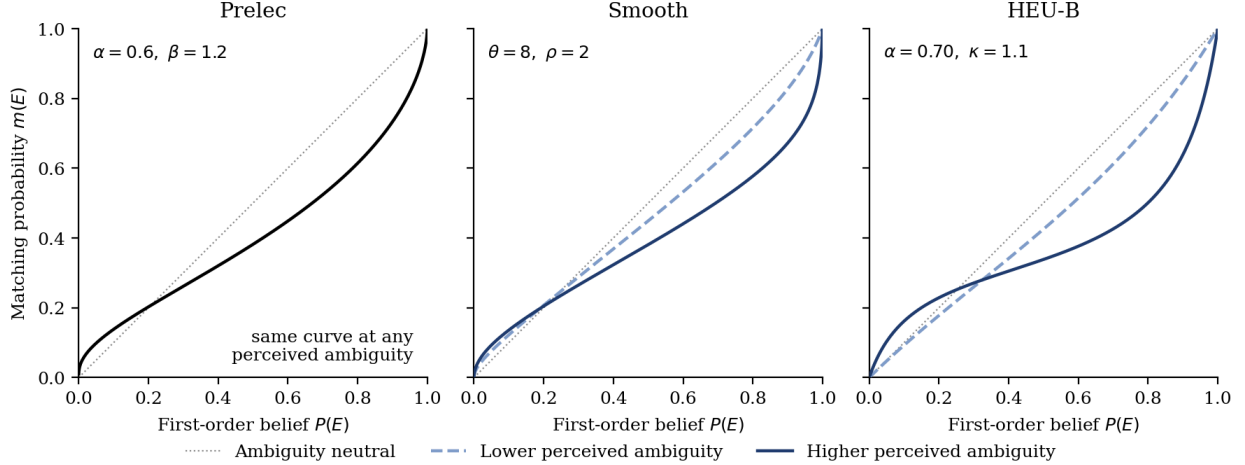


Figure 1: First-order belief, perceived ambiguity, and matching probabilities across models. The horizontal axis gives the first-order belief $P(E)$ and the vertical axis gives the predicted matching probability $m(E)$. The dotted 45-degree line represents $m(E) = P(E)$. For the smooth ambiguity model and HEU-B, the dashed and solid curves illustrate lower and higher perceived ambiguity, respectively, holding ambiguity-attitude parameters fixed. Source-Prelec has no separate perceived-ambiguity component, so the same mapping applies at different levels of perceived ambiguity. Parameter values are chosen for illustration.

ambiguity-attitude parameters are held fixed.

The central implication for our analysis is that the three models may differ in the transferability of their estimated ambiguity-attitude parameters. Because Source-Prelec uses only the first-order belief, behavioral differences associated with changes in perceived ambiguity may be absorbed into its estimated ambiguity-attitude parameters, making those parameters more situation-specific. The smooth ambiguity model and HEU-B separate perceived ambiguity from ambiguity attitudes, which may allow their estimated attitude parameters to remain more stable when the situation changes. This potential advantage comes at the cost of additional structure: both models restrict how changes in perceived ambiguity alter the behavioral effect of a fixed ambiguity attitude. If this imposed relationship does not capture behavior well, more transferable parameters do not necessarily lead to better prediction. Which model provides the best cross-situation prediction is therefore an empirical question.

Evaluating these models at the individual level requires estimating both beliefs and ambiguity attitudes. For the smooth ambiguity model and HEU-B, the relevant belief representation is richer than a single subjective probability and therefore requires eliciting second-order beliefs. Jointly eliciting these richer beliefs and ambiguity attitudes can require a large number of choices. We next discuss this elicitation challenge and conventional approaches before introducing our procedure.

2.3 Elicitation challenges and conventional methods

As discussed above, the matching probability $m(E)$ places the ambiguous prospect (x, E) on the same probability scale as the risky prospect (x, p) . For the binary choices we consider, it therefore summarizes the tradeoff between the two prospects. Previous ambiguity studies have commonly elicited matching probabilities using either a multiple price list (MPL) or a bisection procedure. Both methods aim to identify the probability p at which the decision maker is indifferent between (x, p) and (x, E) , but they present the choices in different ways.

An MPL presents a series of choices on the same page. Each row contains the same ambiguous prospect (x, E) , but the probability p in the risky prospect (x, p) increases across rows. When p is low, the ambiguous prospect may be more attractive, while sufficiently high values of p make the risky prospect more attractive. The matching probability is inferred from the point in the list at which the preferred option changes. For example, if the decision maker chooses the ambiguous prospect when $p = 0.4$ but the risky prospect when $p = 0.5$, the matching probability lies between 0.4 and 0.5. Using finer increments in p gives a narrower interval for the matching probability but requires more rows.

The bisection procedure instead presents one tradeoff between an ambiguous prospect and a risky prospect at a time, with each subsequent question determined by the previous answer. It usually begins with $p = 0.5$. Suppose the decision maker prefers the risky prospect $(x, 0.5)$ to the ambiguous prospect (x, E) . The matching probability is then taken to lie below 0.5, so the next question uses $p = 0.25$. If instead the ambiguous prospect is preferred in the first question, the matching probability is taken to lie above 0.5, and the next question uses $p = 0.75$. Each answer therefore halves the remaining range in which the matching probability can lie. Repeating this process progressively narrows that range until the desired level of precision is reached.

Both methods have advantages and potential drawbacks. MPLs require decision makers to understand how their choices across the entire list reveal a switching point. Misunderstanding this task, or relying on features of the list rather than the underlying tradeoff, can distort the elicited matching probability. In particular, decision makers may be drawn toward switching near the middle of the list, generating a midpoint or compromise effect (Beauchamp et al., 2020). Binary choices may therefore be easier to understand and less susceptible to this feature of the elicitation format (Dave et al., 2010; Freeman et al., 2019). Bisection presents such binary choices and has been widely used in ambiguity experiments, including studies with broad population samples (von

Gaudecker et al., 2025). Its adaptive structure, however, relies heavily on the correctness of previous responses. If the decision maker makes an error early in the sequence, the procedure may narrow the range in the wrong direction and place subsequent questions away from the underlying matching probability. Even if the estimation procedure allows for choice error, an early error can still reduce the information contained in the elicited choices because it affects where the later questions are asked.

For our purpose, another important consideration is that the matching probabilities themselves are not the ultimate objects of interest. In many ambiguity applications, matching probabilities are elicited for the full set of relevant events because commonly used measures of ambiguity attitude are constructed directly from them. For example, the a - and b -indexes in our three-color setting use the matching probabilities for all six single- and two-color events. For structural estimation, by contrast, our objective is to recover the underlying belief and ambiguity-attitude parameters. Different matching-probability questions may provide different amounts of information about these parameters. It may therefore be unnecessary to estimate every matching probability to the same predetermined level of precision.

A full calibration of the models also requires belief elicitation. For the smooth ambiguity model and HEU-B, this requires richer information about the decision maker’s second-order belief, beyond a single subjective probability. Existing choice-based approaches recover this information from several indifference conditions involving bets on possible compositions of the ambiguous source, and differ in how these conditions are constructed.¹⁴ Regardless of the particular implementation, recovering a sufficiently rich second-order belief requires several additional elicitation sequences on top of those used to identify ambiguity attitudes. Jointly recovering beliefs and ambiguity attitudes at the individual level can therefore become highly question-intensive.

Taken together, recovering both belief and ambiguity-attitude parameters at the individual level using conventional elicitation procedures can require a large number of choices. This leaves scope for improving elicitation efficiency: it may be possible to recover these parameters with comparable or greater precision using fewer questions. We therefore adopt an adaptive elicitation procedure that allows us to recover the required information within a more manageable number of questions.

¹⁴For example, Abdellaoui et al. (2011) construct exchangeable partitions recursively, so later elicitation steps use boundaries identified in earlier steps. This avoids having to specify those boundaries in advance, but makes later elicitations depend on earlier ones. Abdellaoui et al. (2025) instead use exchangeability conditions based on boundaries specified independently of earlier elicited values, in part to reduce potential error propagation.

2.4 DOSE: elicitation and estimation

To improve the efficiency of our elicitation, we use Dynamically Optimized Sequential Experimentation (DOSE), developed by Chapman et al. (2025a). DOSE is an adaptive, model-guided procedure that selects each question according to how informative it is expected to be about the parameters of interest.

We now describe the basic logic of the procedure. DOSE begins with a prior distribution over possible parameter values. For every question that could be asked, the structural model implies a probability of each possible response under each parameter combination. Using these predicted responses and the current distribution over the parameters, DOSE evaluates how much information each question is expected to provide. It asks the question with the highest expected information gain. After observing the answer, the distribution over the parameters is updated using Bayes' rule, and the procedure is repeated.

Formally, let ψ denote a candidate parameter vector and let $w_t(\psi)$ denote the weight placed on it after the first t responses. For a binary ambiguity question Q , let $P_\psi(Q)$ denote the probability of choosing the ambiguous prospect under parameter vector ψ . Given the current weights, the predicted probability of choosing the ambiguous prospect for question Q is

$$\bar{P}_t(Q) = \sum_{\psi} w_t(\psi) P_\psi(Q).$$

DOSE evaluates each candidate question using its expected information gain,

$$\text{EIG}_t(Q) = \sum_{\psi} w_t(\psi) \left[P_\psi(Q) \log \frac{P_\psi(Q)}{\bar{P}_t(Q)} + (1 - P_\psi(Q)) \log \frac{1 - P_\psi(Q)}{1 - \bar{P}_t(Q)} \right]. \quad (10)$$

The question with the highest expected information gain is selected. After the decision maker answers, the weights on the candidate parameter vectors are updated using Bayes' rule, and the expected information gain is recalculated to select the next question. In this way, each question is chosen based on the information already collected from that decision maker and on what remains uncertain about the parameters.

This gives DOSE a broader scope for adaptive question selection than bisection. In a bisection elicitation, the researcher first specifies the object to be elicited. For example, once the matching probability for event E has been chosen as the target, E remains fixed and the procedure varies the probability q in the risky prospect to locate the corresponding indifference point. DOSE instead

selects questions according to the parameters we ultimately want to estimate. For ambiguity choices, for example, it can choose both which ambiguous event to ask about and which risky probability to pair with that event. It can therefore search across a larger set of possible questions rather than separately locating each matching probability to a predetermined level of precision.

Another useful feature of DOSE is that response error can be incorporated directly into question selection. Consider an ambiguity choice between the ambiguous prospect (x, E) and the risky prospect (x, q) . For a given parameter vector ψ , let $m_\psi(E)$ denote the matching probability implied by the model. We specify the probability of choosing the ambiguous prospect as

$$P_\psi(E, q) = \Phi\left(\frac{m_\psi(E) - q}{\sigma}\right),$$

where σ captures response error. Choices are therefore less deterministic when q is close to the predicted matching probability and more deterministic when the two are farther apart. An answer that is unlikely under a particular parameter vector lowers the weight placed on that vector rather than eliminating it. Possible response errors are thus taken into account when the posterior is updated and when subsequent questions are selected.

The same general procedure can be applied to belief elicitation. Belief questions involve choices between bets on events concerning the possible composition of the ambiguous box rather than choices between an ambiguous prospect and a risky prospect. A candidate set of belief parameters determines the relative attractiveness of the two options and therefore the predicted probability of each response. DOSE can then evaluate the possible belief questions using the same expected-information-gain criterion. The precise belief specification and the candidate ambiguity and belief questions used in our experiments are described in the next section.

The DOSE procedure itself relies on Bayesian updating: after each response, Bayes' rule updates the weights placed on the candidate parameter values, and these updated weights are used to select the next question. We use the same Bayesian approach for structural estimation throughout the paper. The resulting posterior distribution reflects both which parameter values are better supported by the observed choices and how much uncertainty remains about them. When the choices provide strong information about the parameters, the posterior becomes concentrated; when a range of parameter values explains the choices similarly well, the posterior remains more diffuse.

3 Experimental design

This section describes our benchmark and main experiments. In both experiments, subjects face two ambiguous boxes, each containing 100 balls in three colors: red, yellow, and blue. The composition of each box is unknown to subjects. Before making decisions involving each box, subjects observe a number of draws with replacement from that box. By varying the number of observed draws, we change the amount of information subjects have about the box’s composition. The two experiments differ in how beliefs and ambiguity attitudes are elicited, as described below.

Both experiments were preregistered and conducted online using oTree (Chen et al., 2016), with subjects recruited through Prolific.¹⁵ Subjects received a \$3 participation fee and could earn an additional \$5 bonus. Both experiments included comprehension and attention checks.¹⁶

3.1 Benchmark Experiment

In the benchmark experiment, each subject makes decisions involving two ambiguous boxes, labeled Box A and Box B. Before making decisions about each box, subjects observe either two or sixteen draws with replacement, with the order of the two information conditions randomized across subjects.¹⁷

Subjects answer 48 binary questions in total, completing 24 questions for one box before proceeding to the other. In each question, a winning event is specified. For example, the winning event could be drawing a red ball, which we call a single-color event, or drawing either a red or a blue ball, which we call a two-color event. Subjects choose which of two boxes the ball will be drawn from: a box with a known composition or the corresponding ambiguous box. The probability of the winning event is known for the first option but unknown for the second, corresponding to the risky and ambiguous prospects introduced in Section 2.1. Both options pay \$5 if the winning event occurs and zero otherwise. Figure 2 illustrates a question in which the winning event is drawing a red ball. Subjects choose between a known box containing 50 red balls out of 100 and the ambiguous box (Box A in this example).

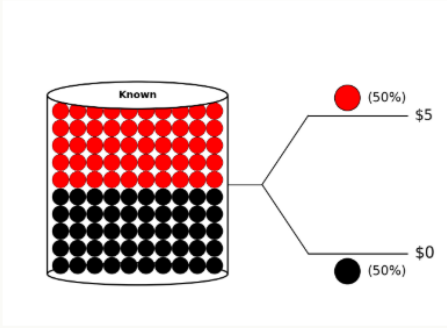
¹⁵Eligible participants were U.S. residents aged 18–80, fluent in English, with an approval rate of at least 99% and at least 100 approved submissions. Participants who had taken part in an earlier experiment in this project were excluded from subsequent experiments.

¹⁶Comprehension questions were administered after the instructions. Subjects who answered the same comprehension question incorrectly twice were not allowed to proceed and were asked to return their submission. Subjects who failed the attention checks were excluded from the analysis.

¹⁷Subjects were randomly assigned to one of two composition families: $\{10, 30, 60\}$ or $\{35, 30, 35\}$. Both boxes for a given subject belong to the assigned family, with the ball counts randomly assigned to colors.

Section 1 · Question 1 of 24 — Which option do you prefer?

OPTION 1: Known box — 50 out of 100 balls are red. Win \$5 if a ball drawn from the box is red, \$0 otherwise.

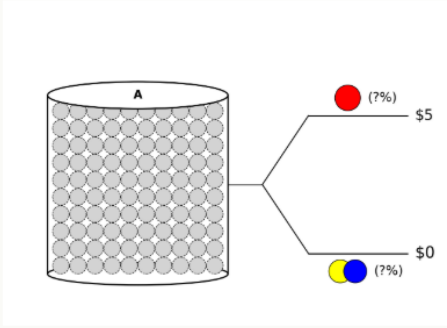


CHOOSE OPTION 1

You saw 16 example draws from Box A:

● Red × 8
 ● Blue × 5
 ● Yellow × 3

OPTION 2: Box A — an unknown number out of 100 balls are red. Win \$5 if a ball drawn from Box A is red, \$0 otherwise.



CHOOSE OPTION 2

SUBMIT

Figure 2: An MP question in the benchmark experiment. Option 1 pays \$5 if a ball drawn from a known box with 50 red balls out of 100 is red; Option 2 pays \$5 if a ball drawn from Box A, whose composition is unknown, is red. The panel above Option 2 shows the draws the subject has seen from Box A.

For each box, we consider six winning events: three single-color events and three two-color events. We use a bisection procedure to elicit the matching probability for each event. Specifically, subjects answer four binary questions for each event. The first question offers a 50% winning probability in the known box, and the probability in the subsequent three questions is adjusted based on the subject's previous responses. The order of the six event blocks is randomized.

To determine the bonus payment, one of the 48 question positions is randomly selected and fixed before the experiment begins. Subjects receive a meaningless four-letter code uniquely associated with the selected position but are not informed which question it represents. At the end of the experiment, the subject's choice at that position is implemented: a ball is drawn from the box they selected, and they receive the \$5 bonus if the winning event occurs. We collected responses from 255 subjects, of whom four failed the attention checks, leaving 251 subjects for analysis.

Hypotheses. Based on the findings of Abdellaoui et al. (2025), we expect estimated ambiguity attitudes to move toward ambiguity neutrality as subjects observe more draws. We preregistered the following hypotheses:

Hypothesis 1 *From the 2-draw box to the 16-draw box, both the a -index and the b -index (defined in Section 4.1) move toward zero.*

Hypothesis 2 *From the 2-draw box to the 16-draw box, both Source-Prelec parameters, α and β , move toward their ambiguity-neutral values.*¹⁸

As discussed in Section 2.4, our main experiment uses DOSE to elicit both beliefs and ambiguity attitudes efficiently. To validate its performance, we conducted a DOSE validation experiment comparing it with the conventional bisection procedure. Subjects were randomly assigned to one of the two procedures. Both treatments used the same set of eligible events for elicitation, with additional events reserved exclusively for evaluation and unavailable to either procedure. Each subject answered 32 elicitation questions followed by 16 held-out questions. We used the same estimation method for both treatments and evaluated their performance using out-of-sample predictions for the held-out choices. The results show that DOSE provides more precise parameter estimates and achieves comparable predictive performance with fewer elicitation questions. The full experimental design and results are reported in Appendix A.

3.2 Main Experiment

In the main experiment, subjects again make decisions involving two ambiguous boxes, Box A and Box B. Subjects first observe four draws with replacement from Box A and complete questions about that box. They then observe twenty draws with replacement from Box B and complete questions about Box B.¹⁹ We estimate ambiguity-attitude parameters using choices from Box A and combine these estimates with beliefs elicited about Box B to predict held-out choices in Box B.

Subjects answer 52 binary questions in total, with 24 questions for Box A and 28 for Box B. There are two types of questions: belief questions, which elicit subjects’ beliefs about the composition of a box, and ambiguity questions, which involve the same type of choices between

¹⁸We also preregistered and tested corresponding hypotheses for alternative source theory specifications, including neo-additive and Goldstein–Einhorn. The results are reported in Appendix Table B.1.

¹⁹Both boxes have compositions from the $\{10, 30, 60\}$ family, with the assignment of ball counts to colors randomly determined for each subject and shared across the two boxes.

known and ambiguous boxes as in the benchmark experiment. The questions are organized into four consecutive blocks. For Box A, subjects answer a block of 12 belief questions followed by a block of 12 ambiguity questions. For Box B, subjects answer a block of 12 belief questions followed by a block of 16 ambiguity questions.

Questions are selected adaptively within each block. For the belief questions in both boxes, we use a common belief specification to guide DOSE question selection. For the ambiguity questions in Box A, subjects are randomly assigned to one of three between-subject treatments, with DOSE question selection guided by Source-Prelec, the smooth ambiguity model, or HEU-B. Finally, the 16 ambiguity questions in Box B are selected using a model-independent adaptive procedure and are held out exclusively for evaluating predictions.

In the belief-question blocks, subjects face a different type of choice from those in the benchmark experiment: they choose between two bets on the composition of the ambiguous box. Specifically, they are shown two ranges for the number of balls of a particular color and choose which range they believe is more likely to contain the actual number.²⁰ Each option pays \$5 if the actual number falls within the chosen range and zero otherwise. Figure 3 illustrates a belief question about the number of blue balls in Box A. Subjects choose between winning \$5 if Box A contains 0–28 blue balls and winning \$5 if it contains 29–100 blue balls.

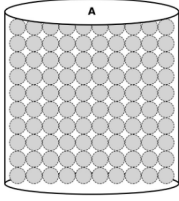
Belief questions allow us to identify perceived ambiguity, which is needed for the smooth ambiguity model and HEU-B. Unlike ambiguity questions, belief questions compare two bets on the composition of the same ambiguous box, so ambiguity-attitude parameters do not affect the trade-off between them. This allows us to elicit beliefs about Box A and Box B separately from ambiguity attitudes, accounting for differences in beliefs between the two boxes while keeping the ambiguity choices in Box B entirely held out for prediction. We can therefore conduct a clean out-of-sample test of whether ambiguity-attitude parameters estimated from Box A can predict behavior in Box B.

To guide DOSE question selection in the belief blocks, we use a common belief representation. Specifically, we represent subjects’ beliefs about the composition of each box using a Dirichlet distribution over the proportions of the three colors.²¹ This distribution is characterized by three parameters: m_R , m_Y , and K . The first two represent subjects’ mean beliefs about the proportions

²⁰The two ranges do not necessarily cover all possible numbers of balls. If the actual number falls outside both ranges, neither option pays the bonus.

²¹Previous studies have used Beta distributions to represent beliefs about the composition of two-color boxes (Abdellaoui et al., 2024, 2025). The Dirichlet distribution is a natural extension to the three-color case.

Question 1 of 52 — Which do you think is more likely?



This question is about the number of blue balls in Box A.
Among 4 example draws from Box A:

Red × 2 Blue × 1 Yellow × 1

OPTION 1

Win \$5 if the number of blue balls in Box A (out of 100) is:

- (i) at least 0, AND
- (ii) no more than 28.

Otherwise, you win \$0.

CHOOSE OPTION 1

OPTION 2

Win \$5 if the number of blue balls in Box A (out of 100) is:

- (i) at least 29, AND
- (ii) no more than 100.

Otherwise, you win \$0.

CHOOSE OPTION 2

SUBMIT

Figure 3: A belief question in the main experiment. The subject chooses between two bets on the number of blue balls in Box A, one paying \$5 if the count lies in the lower range and one if it lies in the upper range. The four draws seen from Box A are shown above the options. This is the first belief question of the training box, identical for every subject; later questions adapt to the answers given.

of red and yellow balls, respectively, with $m_B = 1 - m_R - m_Y$ representing the mean belief about blue balls. The parameter K captures how concentrated the belief distribution is around these mean proportions. A higher K indicates that the subject is more sure that the actual composition is close to their best guess. Figure 4a illustrates two possible belief distributions after observing four and twenty draws with the same observed color proportions.²² Although the mean beliefs about the proportion of red balls are similar, the distribution after twenty draws has a higher K and is more concentrated, reflecting lower perceived ambiguity. This illustrates how the Dirichlet representation captures both first-order beliefs and the additional dimension of how sure the decision maker is

²²The distributions illustrate beliefs that a decision maker would hold under Bayesian updating. We use them only for illustration and do not assume that subjects update their beliefs according to Bayes' rule. Instead, we elicit their beliefs about each box.

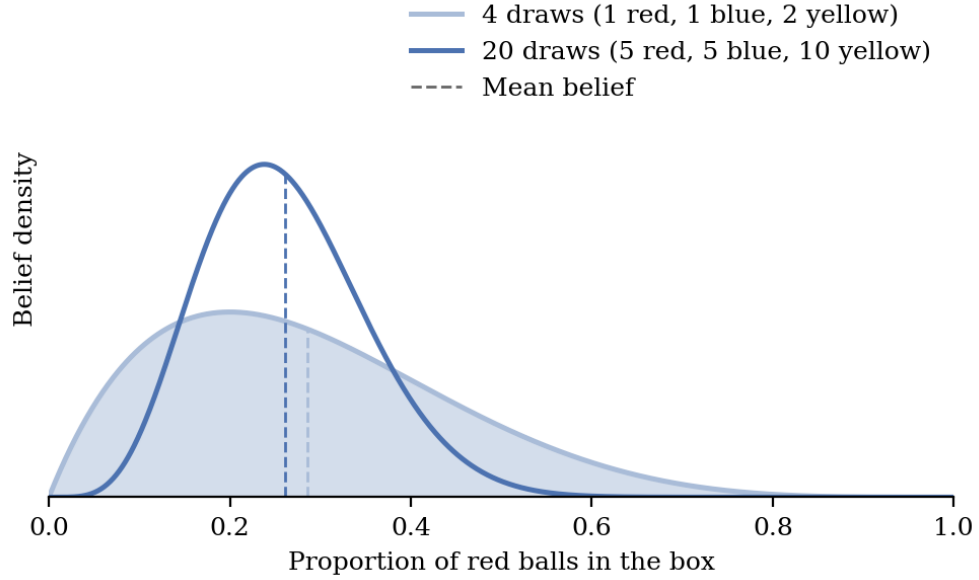
about those probabilities.

We use this common belief representation to construct the belief components required by each model, as illustrated in Figure 4b. For Source-Prelec, we use the mean beliefs from the Dirichlet distribution as the first-order beliefs. The smooth ambiguity model uses the full Dirichlet distribution as its second-order belief. For HEU-B, we use the mean beliefs together with a measure of belief spread derived from the Dirichlet distribution to construct the lower and upper probabilities entering its decision rule, as described in Section 2.2.3. In this way, we can use the same DOSE procedure to select belief questions across all three treatments, while preserving the different ways the models incorporate belief information in estimation and prediction.

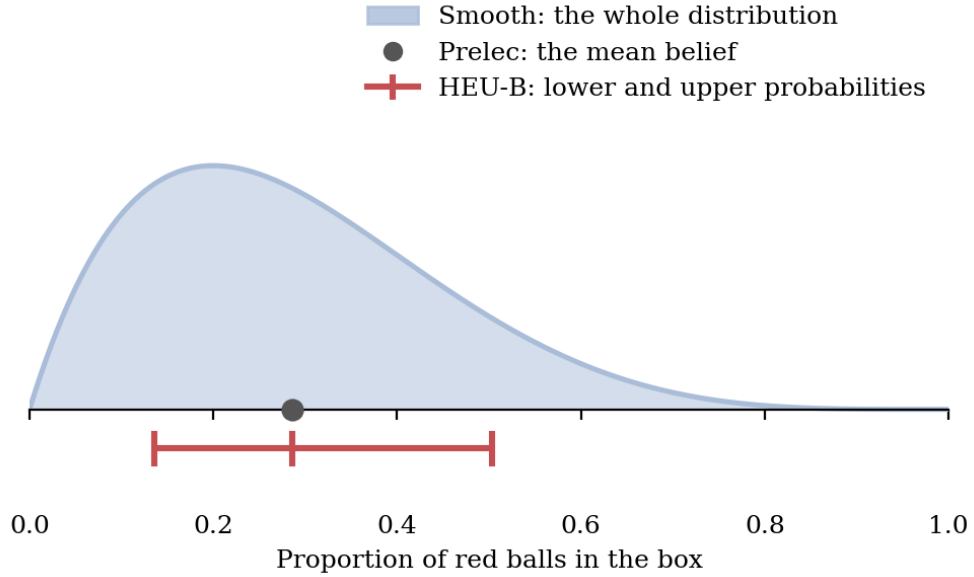
In the belief-question blocks, DOSE selects each question to maximize expected information gain about the Dirichlet belief parameters, m_R , m_Y , and K , as well as a response-error parameter. After each response, the posterior distribution over these parameters is updated, and the next question is selected based on the updated posterior. The same adaptive belief-elicitation procedure is used in both boxes and across all three treatments. For the ambiguity-question block in Box A, DOSE starts with the belief information obtained from the preceding belief-question block, together with a prior distribution over the ambiguity-attitude parameters. Each question is then selected to maximize expected information gain about the ambiguity-attitude parameters, the belief parameter K where relevant, and a response-error parameter.²³ The guiding model depends on the subject’s assigned treatment: Source-Prelec, the smooth ambiguity model, or HEU-B.

The 16 held-out ambiguity questions in Box B follow a different adaptive selection procedure to ensure that the evaluation does not favor any particular model. We randomly select one of the three pairs of complementary events, such as drawing a red ball and drawing a yellow or blue ball, and ask eight questions about each event. Rather than relying on any of the three ambiguity models, the procedure selects questions to obtain information about the subject’s matching probability for each event. Similar to bisection, it adjusts the known winning probability based on previous responses, but explicitly accounts for response error. Importantly, the procedure starts with a prior that does not depend on any responses collected before the held-out block. Responses within this block are used to select subsequent held-out questions and evaluate predictions, but are never used to estimate the parameters of the three models. This design allows us to compare the models

²³For computational feasibility in real-time question selection, the mean belief parameters m_R and m_Y are fixed at their posterior means from the belief-question block, while the posterior distributions over K , where relevant, and the ambiguity-attitude and response-error parameters are updated sequentially. In the final estimation, all model-relevant parameters are jointly updated using the full set of responses from Box A.



(a) Dirichlet belief distributions after four and twenty draws



(b) How the three models use belief information

Figure 4: Illustration of the common belief representation. Panel (a) shows two possible belief distributions with similar means but different concentrations, illustrating how perceived ambiguity can differ even when first-order beliefs are similar. Panel (b) illustrates how Source-Prelec uses the mean belief, the smooth ambiguity model uses the full distribution, and HEU-B uses the mean belief and belief spread to construct lower and upper probabilities.

on questions selected independently of them, while obtaining informative responses for evaluating their relative predictive performance with a limited number of questions.

To implement DOSE, we need to setup prior (the possible set of parameters and their corresponding probability) so DOSE has starting point to update on. All parameter values we used for prior are assumed to be equally likely. For m_r and m_y belief parameters, we allow parameters to take value between $[0.025, 0.95]$, with increment of 0.025. For K parameters, we take 14 points, which are log-spaced from 3 to 100. For ambiguity attitudes parameters, we take 14 points as well for each parameter. All parameters' value are chosen based on the benchmark experiment, more details about the exact values of these parameters can be found in Appendix Table C.1. Since the first belief questions and the first held-out question are purely using prior, we build the question tree for these questions. That is, we pre-compute what possible choice should lead what question to ask for each question and record them. For the MP question for the first box, because information from the same individual's belief questions would be needed for DOSE procedure as well, we let the program to do the real-time calculation after collecting each question. Due to this constraint (we can not let subjects to wait for a long time), have 14 possible values for each ambiguity parameters. But this does not affect the performance of DOSE, we just to ensure DOSE is asking questions that are good enough for parameter identification, for later analyses, we can use a lot finer parameter set (like 28 per parameter) for better identification. Chapman et al. (2025a) did in the similar way: select question using coarse parameter set then estimate using finer parameter set.

And note that since DOSE needs to select question to update for prior, we also need to have question pool for DOSE to choose from. For belief questions, we have question pool of 12699 per box. Since question's format is always about which of the two ranges include the true number of balls of one specific color, the variation for the question includes color and what are the two ranges. We don't think the question that provides range that is too thin is meaningful, as when the range is too thin, the error might be high enough to offset any information that might be provided by the choice. For each color, if we represent the two ranges provided by the choice question as $[\text{lower number}, t]$ and $[t, \text{higher number}]$. We make sure that (1) all lower number and higher number are on the 5 grid (2) t is always between lower number and higher number, and t is at least 10 over lower number and 10 below higher number. Overall, we have 12699 possible belief questions for three colors. For MP question, we vary both the event and the probability provided in the known box side. There are six events, and we provide probabilities with increment of 1%, as a result, we have 606 possible questions each box.

For payments, we follow the PRINCE mechanism (Johnson et al., 2021). Before the experiment begins, one question is randomly selected from the full pool of possible questions and fixed for payment. Subjects receive a meaningless code corresponding to this question but are not informed which question it represents. If the selected question is asked during the experiment, the subject’s response to that question determines the bonus payment. Otherwise, the selected question is presented at the end of the experiment, and the subject’s response is implemented. Because the payment question is determined independently of subjects’ responses, subjects have no incentive to misreport their choices to influence which questions are subsequently asked.

We recruited 391 subjects for the main experiment. Three failed the attention checks, and an additional 24 were excluded under the preregistered criterion for nearly constant response patterns.²⁴ The final sample consists of 364 subjects: 117 in the Source-Prelec treatment, 119 in the smooth ambiguity treatment, and 128 in the HEU-B treatment.

To evaluate out-of-sample predictive performance, we estimate each model’s ambiguity-attitude parameters using choices from Box A and combine these estimates with beliefs elicited about Box B to predict the 16 held-out ambiguity choices. We evaluate predictions using the quadratic score. Let p denote the probability a model assigns to the option actually chosen by the subject. The model receives a score of $-(1 - p)^2$ for each held-out question. We average these scores across the 16 questions to obtain each subject’s out-of-sample prediction score for each model. Higher scores indicate better predictive performance, with a maximum possible score of 0.

We next present three hypotheses about the models’ out-of-sample predictive performance. Both the smooth ambiguity model and HEU-B incorporate perceived ambiguity into beliefs, allowing differences in perceived ambiguity across boxes to be accounted for separately from ambiguity attitudes. We therefore expect their estimated ambiguity-attitude parameters to be more transferable across boxes, potentially leading to better predictions than Source-Prelec.

Hypothesis 3 *The smooth ambiguity model achieves a higher out-of-sample prediction score than Source-Prelec.*

Hypothesis 4 *HEU-B achieves a higher out-of-sample prediction score than Source-Prelec.*

Although both models incorporate perceived ambiguity, they differ in how they use this information. The smooth ambiguity model evaluates the full distribution of possible probabilities, whereas

²⁴The main ranking of the three models’ predictive performance remains unchanged when these subjects are included.

HEU-B focuses on pessimistic and optimistic probabilities constructed from the belief distribution. We conjecture that the latter approach better captures behavior, motivating Hypothesis 5.

Hypothesis 5 *HEU-B achieves a higher out-of-sample prediction score than the smooth ambiguity model.*

4 Results from the Benchmark Experiment

The benchmark experiment provides an initial examination of how transferable estimated ambiguity-attitude parameters are across situations with different levels of information. We first examine how model-free measures of ambiguity attitudes and Source-Prelec parameters change between the two-draw and sixteen-draw boxes. We then approximate second-order beliefs under a Bayesian-updating assumption and compare the stability of estimated ambiguity attitudes across the three models. Finally, we evaluate each model’s ability to predict choices in one box using parameters estimated from the other. We find that models incorporating second-order beliefs yield more stable ambiguity-attitude parameters, but their advantage in cross-situation prediction depends on the model and the direction of prediction.

4.1 Changes in Estimated Ambiguity Attitudes

We begin by examining whether estimated ambiguity attitudes change with the amount of information available. Following Baillon et al. (2018b), we use two model-free measures of ambiguity attitudes: the b-index, which captures ambiguity aversion, and the a-index, which captures ambiguity-generated likelihood insensitivity (a-insensitivity). Let $\bar{m}_s$ and $\bar{m}_c$ denote the average matching probabilities for the three single-color and three two-color events, respectively. The indexes are defined as

$$b = 1 - \bar{m}_s - \bar{m}_c, \quad a = 1 - 3(\bar{m}_c - \bar{m}_s).$$

A positive b-index indicates ambiguity aversion, while a positive a-index indicates likelihood insensitivity. Both indexes equal zero under ambiguity neutrality. We calculate these indexes separately for each box, using the midpoint of the final interval from the bisection procedure as the matching probability for each event.²⁵

²⁵This approximation assumes that subjects make no errors during the bisection procedure. An incorrect response may cause the final interval to exclude the underlying matching probability.

We first examine whether the changes in ambiguity attitudes documented by Abdellaoui et al. (2025) also arise in our three-color setting. Figure 5 compares the two indexes between the two-draw and sixteen-draw boxes. Both indexes move significantly toward ambiguity neutrality as subjects observe more draws. The median b-index decreases from 0.152 to 0.050 (Wilcoxon signed-rank test, $p < 0.001$), indicating less ambiguity aversion. Similarly, the median a-index decreases from 0.805 to 0.560 ($p < 0.001$), indicating greater sensitivity to event likelihoods. These findings are consistent with Abdellaoui et al. (2025), who also find that both indexes move toward neutrality as more information becomes available.

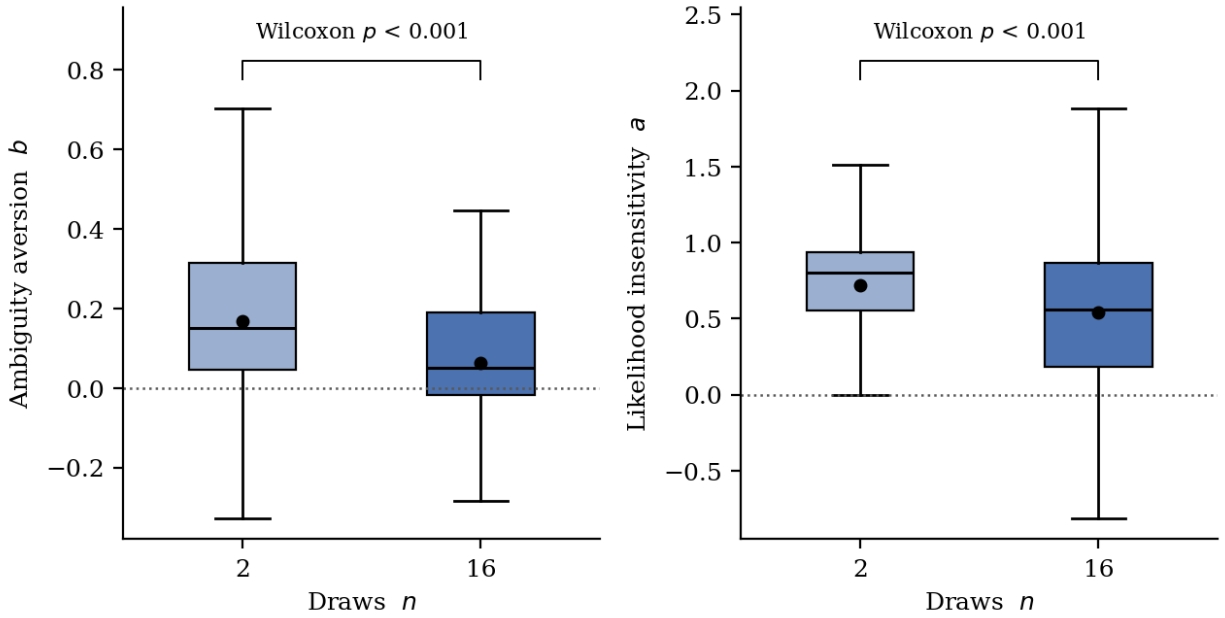


Figure 5: Model-free ambiguity attitude across the two boxes. Box, median line and mean dot, with the dotted line marking ambiguity neutrality. The bracket above each pair carries the Wilcoxon signed-rank test on per-subject differences. $N = 251$.

The a- and b-indexes measure ambiguity attitudes without requiring a particular structural model. However, under the neo-additive specification of source theory, they map directly into the parameters of the weighting function, which imposes an affine relationship between matching probabilities and first-order beliefs. Moreover, the model-free approach treats elicited matching probabilities as exact, although they may contain errors from the bisection procedure. We therefore turn to structural estimation, allowing for a more flexible weighting function and response errors, to examine whether estimated ambiguity attitudes still exhibit similar changes. We focus on the Prelec specification introduced in Section 2.2.1 and estimate its two ambiguity-attitude parameters

jointly with first-order beliefs using the grid-based Bayesian procedure that also underlies our DOSE estimation method.²⁶

Figure 6 compares the estimated Source-Prelec parameters across the two boxes. Both parameters change significantly, with their median values moving toward ambiguity neutrality. The median of α increases from 0.152 to 0.440 ($p < 0.001$), while the median of β decreases from 1.031 to 1.002 ($p = 0.033$). Under the Prelec specification, ambiguity neutrality corresponds to $\alpha = \beta = 1$. Thus, the changes observed in the model-free indexes are also reflected in the structural estimates under source theory.

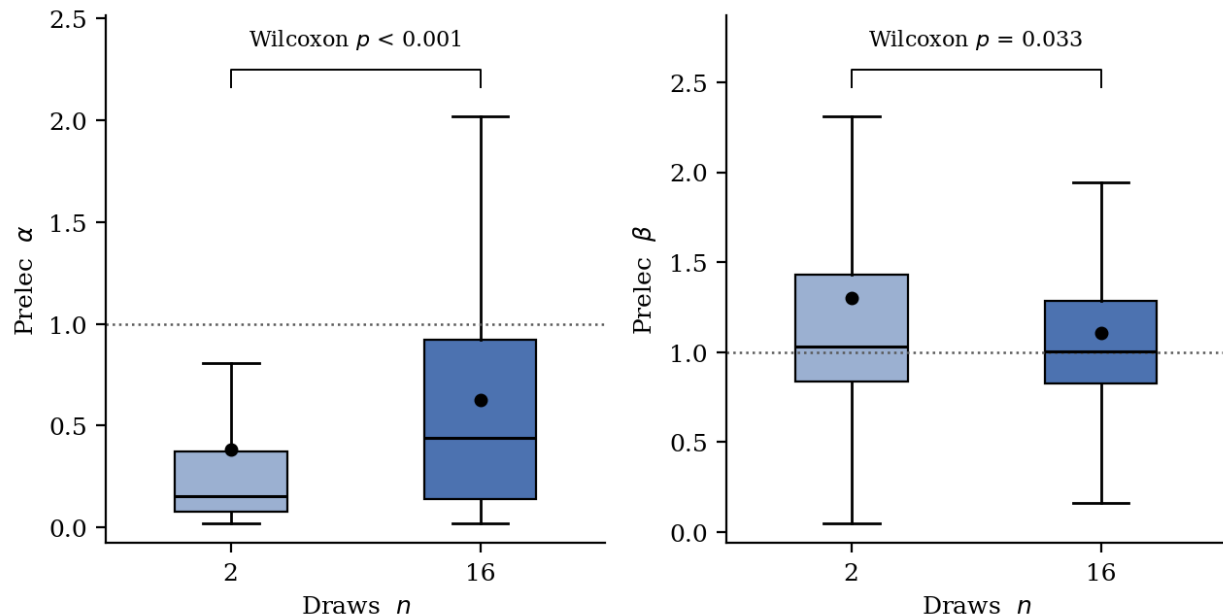


Figure 6: Prelec parameters across the two boxes, grid-Bayesian posterior means with the belief estimated jointly rather than assumed. Box, median line and mean dot, with the dotted line marking ambiguity neutrality ($\alpha = \beta = 1$). The bracket above each pair carries the Wilcoxon signed-rank test on per-subject differences. $N = 251$.

The observed movement toward ambiguity neutrality does not necessarily imply that underlying ambiguity attitudes have changed. Under the smooth ambiguity model and HEU-B, behavior can move closer to ambiguity neutrality as perceived ambiguity decreases, even when ambiguity-attitude parameters remain unchanged. If this mechanism drives the behavioral response to additional

²⁶Maximum likelihood estimation (MLE) may produce poorly identified parameter estimates when subjects' matching probabilities exhibit little variation across events. In our benchmark experiment, six matching probabilities are used to estimate five parameters: two belief parameters, two ambiguity-attitude parameters, and one error parameter. The grid-based Bayesian procedure provides posterior estimates for all 251 subjects, with less informative observations yielding estimates that remain more influenced by the prior. We also conduct MLE on the subsample of subjects whose data meet the identification criteria. The MLE results, reported in the appendix, show the same direction of parameter changes.

draws, source theory may instead attribute these changes to its ambiguity-attitude parameters, making them move toward neutrality. This suggests that models separating perceived ambiguity from ambiguity attitudes may yield more stable parameter estimates, which we examine next.

4.2 Parameter Stability with Approximated Second-Order Beliefs

As discussed in Section 2, the binary choices used to elicit matching probabilities in the benchmark experiment cannot separately identify second-order beliefs and ambiguity attitudes. Thus, while source theory allows us to estimate ambiguity-attitude parameters jointly with first-order beliefs, estimating these parameters under the smooth ambiguity model and HEU-B requires additional information about second-order beliefs.

To provide a preliminary comparison across models, we approximate second-order beliefs by assuming that subjects update according to Bayes’ rule from a uniform prior. Specifically, we assume that subjects begin with a Dirichlet(1, 1, 1) prior over the possible compositions of the box, which assigns equal prior density to all possible color proportions. After observing D draws, of which D_c are of color c , the posterior belief follows Dirichlet($1 + D_R, 1 + D_Y, 1 + D_B$). The corresponding first-order belief for color c is the mean of the marginal distribution,

$$m_c = \frac{1 + D_c}{3 + D},$$

and the concentration parameter is $K = 3 + D$. A higher K indicates a more concentrated second-order belief and therefore lower perceived ambiguity. Under this assumption, we can estimate the ambiguity-attitude parameters of the smooth ambiguity model and HEU-B. For comparability, we also estimate Source-Prelec using the same approximated beliefs, rather than estimating first-order beliefs jointly with ambiguity attitudes as in Section 4.1.

Table 1 reports the estimated ambiguity-attitude parameters for all three models in the two-draw and sixteen-draw boxes. To examine parameter stability, we calculate the correlation of each parameter’s estimates across the two boxes at the individual level. For Source-Prelec, the correlations are 0.255 for α and 0.310 for β . In contrast, HEU-B has correlations of 0.561 for α and 0.491 for κ , while the corresponding correlations for the smooth ambiguity model are 0.464 for θ and 0.565 for ρ . Thus, both models incorporating second-order beliefs exhibit higher cross-box parameter correlations than Source-Prelec, suggesting that their estimated ambiguity attitudes may be more stable even when second-order beliefs are only approximated.

These findings suggest that separating perceived ambiguity from ambiguity attitudes may yield more stable parameter estimates. However, greater parameter stability does not necessarily translate into better prediction. We next compare the models’ cross-situation predictions using the same approximated beliefs.

Table 1: Ambiguity attitude parameters across the two boxes, benchmark experiment

		All on the assumed Bayesian belief					
		Prelec α	Prelec β	HEU-B α	HEU-B κ	Smooth θ	Smooth ρ
2 draws	Median	0.17	1.01	0.59	4.85	34.97	3.15
	Std	0.90	1.29	0.15	3.34	18.73	2.28
16 draws	Median	0.40	0.99	0.56	2.98	16.98	2.58
	Std	0.69	0.48	0.14	4.84	21.32	2.00
Median difference		+0.127***	-0.019***	-0.030***	-0.173	-10.892***	-0.664***
Correlation across boxes		0.255	0.310	0.561	0.491	0.464	0.565

Notes: Parameter estimates are individual-level posterior means from grid-based Bayesian estimation, assuming that subjects update a uniform Dirichlet(1, 1, 1) prior according to Bayes’ rule. Medians and standard deviations are calculated across subjects. The median difference is the median of within-subject differences (16 draws minus 2 draws). We use the Wilcoxon signed-rank test to compare parameter estimates across the two boxes. The correlation across boxes is the Pearson correlation of individual parameter estimates. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. HEU-B’s κ is weakly identified when the implied $\beta = \tanh(\kappa s/2)$ approaches one, as larger values of κ generate nearly identical predictions. Evidence on the cross-box stability of κ is therefore only suggestive. $N = 251$.

4.3 Cross-Situation Prediction

To examine whether the stability differences documented in Section 4.2 translate into predictive performance, we estimate each model’s ambiguity-attitude parameters using choices from one box and combine them with the approximated Bayesian beliefs from the other box to predict choices in that box. We evaluate predictions using the average quadratic score over the 24 binary choices in the prediction box. For each choice, the score is $-(1 - p)^2$, where p is the model’s predicted probability of the observed choice. A higher score indicates better predictive performance. We conduct this exercise in both directions, estimating parameters from the two-draw box to predict choices in the sixteen-draw box and vice versa.

Table 2 reports the results. For each prediction direction, we use Source-Prelec’s prediction score as the benchmark and compare it with those of HEU-B and the smooth ambiguity model using the Wilcoxon signed-rank test. When parameters are estimated from the two-draw box and used to predict choices in the sixteen-draw box, both HEU-B and the smooth ambiguity model outperform Source-Prelec. The median improvement in the quadratic score relative to Source-Prelec is 0.0220 for HEU-B ($p < 0.001$) and 0.0268 for the smooth ambiguity model ($p = 0.020$). The difference

between HEU-B and the smooth ambiguity model is not statistically significant. However, the results change when we reverse the prediction direction. When parameters are estimated from the sixteen-draw box and used to predict choices in the two-draw box, HEU-B is no longer significantly different from Source-Prelec ($p = 0.170$), while the smooth ambiguity model performs significantly worse ($p = 0.003$).

Table 2: Cross-box prediction, benchmark experiment

	Prelec	HEU-B	Smooth
<i>Fit 2 draws, predict 16 draws</i>			
Median	-0.2132	-0.1757	-0.1929
Std	0.1168	0.1058	0.0884
Difference from Prelec	—	+0.0220***	+0.0268**
<i>Fit 16 draws, predict 2 draws</i>			
Median	-0.2088	-0.2087	-0.2178
Std	0.1185	0.1201	0.0952
Difference from Prelec	—	+0.0015	-0.0118***

Note: The quadratic score is $-(1-p)^2$, where p is the predicted probability of the observed choice, averaged over the 24 binary choices in the prediction box. Higher scores indicate better prediction. Each model is estimated using one box and evaluated on the other. For each subject, we calculate the difference between each model’s score and Prelec’s score. The table reports the median of these differences across subjects. Stars indicate significance from Wilcoxon signed-rank tests of these differences: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Beliefs are approximated under Bayesian updating for all models. $N = 251$.

One possible explanation for this directional difference comes from how perceived ambiguity affects parameter identification under the smooth ambiguity model and HEU-B. As discussed in Section 2, the influence of ambiguity attitudes on behavior becomes weaker as perceived ambiguity decreases, making ambiguity-attitude parameters harder to distinguish from response errors. Our Bayesian approximation implies lower perceived ambiguity in the sixteen-draw box than in the two-draw box. If subjects perceive a similar reduction in ambiguity, estimating parameters from the sixteen-draw box may yield less precise estimates of ambiguity attitudes under these two models, potentially reducing predictive performance when these parameters are applied to the two-draw box.

Overall, these results suggest that incorporating second-order beliefs can improve cross-situation prediction, but the advantage depends on the model and prediction direction. HEU-B is the only model that is not significantly outperformed by Source-Prelec in either direction. However, this comparison remains preliminary because second-order beliefs are approximated rather than directly elicited. Previous studies have documented departures from Bayesian updating under ambiguity

(Epstein and Halevy, 2024; Aydogan et al., 2025; Liang, 2025), and observing more draws does not necessarily reduce every subject’s perceived ambiguity. Our main experiment addresses these limitations by separately eliciting second-order beliefs and ambiguity attitudes, allowing for a more complete comparison of the models’ cross-situation predictive performance.

5 Results from the Main Experiment

We now turn to the results of the main experiment, in which beliefs are directly elicited. We begin by comparing the three models’ out-of-sample predictive performance using both between-treatment and within-subject comparisons. We then validate our elicitation and explore what may account for differences in the models’ predictive performance. Finally, we discuss the implications of using DOSE for model comparison.

5.1 Out-of-Sample Model Comparison

We begin with our main result, comparing the three models’ out-of-sample predictive performance. For each model, we combine ambiguity-attitude parameters estimated from Box A with beliefs elicited from Box B to predict subjects’ choices on the 16 held-out ambiguity questions. We evaluate predictive performance using the quadratic score, with higher scores indicating better prediction. Figure 7 reports the results from both between-treatment and within-subject comparisons.

Panel A of Figure 7 presents the between-treatment comparison, in which each model is evaluated using subjects from its corresponding DOSE treatment. HEU-B achieves the highest median quadratic score (-0.199), followed by Source-Prelec (-0.208) and the smooth ambiguity model (-0.225). HEU-B significantly outperforms both Source-Prelec (Mann–Whitney test, $p = 0.040$) and the smooth ambiguity model ($p = 0.001$), while the difference between the latter two models is not statistically significant ($p = 0.194$).

Panel B presents the within-subject comparison. For each subject, we fit all three models using the same choice data and combine their estimated ambiguity-attitude parameters with the subject’s elicited Box B beliefs to predict the same held-out choices. This allows us to compare the models while holding the data used for estimation and evaluation fixed. We report the results separately for each DOSE treatment and for the pooled sample. The ranking is robust across all comparisons: HEU-B significantly outperforms both alternatives regardless of which model guided question selection, while the smooth ambiguity model and Source-Prelec do not differ significantly. In the

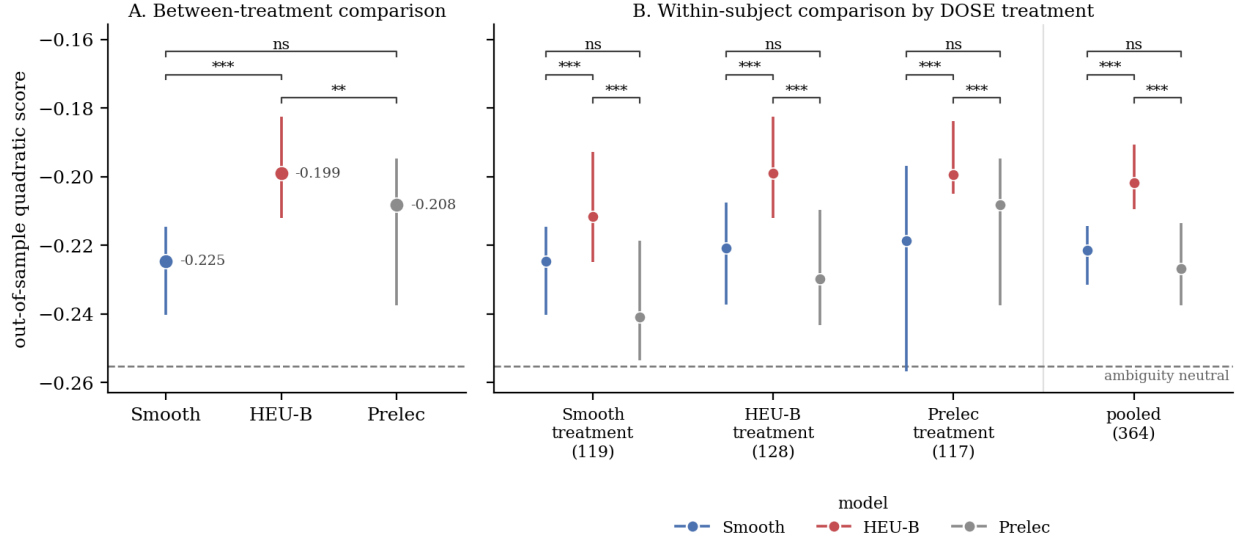


Figure 7: Out-of-sample predictive performance across models and treatments. Dots report median quadratic scores, with vertical lines indicating 95% bootstrap confidence intervals obtained by resampling subjects. Higher scores indicate better prediction. Panel A presents the between-treatment comparison, in which each model is evaluated on subjects from its corresponding treatment (119 Smooth, 128 HEU-B, and 117 Prelec). Panel B presents the within-subject comparison, in which all three models are evaluated on the same subjects, separately by treatment and pooled across all 364 subjects. Brackets report two-sided Mann–Whitney tests in Panel A and Wilcoxon signed-rank tests of paired differences in Panel B. The dashed line indicates the pooled median score (-0.255) under ambiguity neutrality, where matching probabilities equal subjects’ elicited first-order beliefs. $***p < 0.01$, $**p < 0.05$, $*p < 0.10$; ns indicates no statistical significance at the 10% level.

pooled sample, HEU-B outperforms both models (Wilcoxon signed-rank tests, both $p < 0.001$), while the difference between the smooth ambiguity model and Source-Prelec remains insignificant ($p = 0.543$).

Importantly, all three models significantly outperform an ambiguity-neutral benchmark, under which matching probabilities equal subjects’ elicited first-order beliefs. The dashed line in Figure 7 indicates the pooled median score of this benchmark. The improvements are statistically significant in both the between-treatment and pooled within-subject comparisons (Appendix Table C.2). This suggests that the estimated ambiguity attitudes under all three models retain predictive value in a new situation, although to different degrees.

Taken together, these results suggest that accounting for second-order beliefs does not necessarily improve predictive performance over a model that does not incorporate them. The smooth ambiguity model does not outperform Source-Prelec, whereas HEU-B significantly outperforms both. HEU-B’s advantage suggests that incorporating second-order beliefs through pessimistic and

optimistic probability bounds may better capture how these beliefs affect behavior across ambiguous situations.

5.2 Validation

Having shown that the estimated ambiguity attitudes improve out-of-sample prediction relative to ambiguity neutrality, we now examine whether the elicited beliefs respond meaningfully to each subject’s observed draws. Recall that beliefs are elicited using the same Dirichlet specification across treatments, characterized by mean probabilities for each color and a concentration parameter K . We first consider the elicited mean probabilities, which are used as first-order beliefs under source theory. We regress the elicited mean probability for each color on the corresponding Bayesian approximation based on the observed draws. Table 3 reports the results separately for Box A and Box B. The estimated coefficients are 0.935 and 0.818, respectively, both statistically significant at the 1% level. These results suggest that elicited mean beliefs closely track Bayesian approximations based on subjects’ observed draws.

Table 3: Elicited first-order beliefs and Bayesian approximations

	Box A, 4 draws	Box B, 20 draws
<i>Dependent variable: Elicited first-order belief</i>		
Bayesian approximation from draws	0.935*** (0.033)	0.818*** (0.023)
R^2	0.650	0.706
Subjects	364	364
Observations	1,092	1,092

Notes: The dependent variable is the elicited first-order belief about the proportion of each color in the box. The explanatory variable is the corresponding belief approximated from the observed draws, assuming Bayesian updating from a uniform prior. Regressions are estimated separately for the two boxes, with three observations per subject (one for each color). Standard errors clustered at the subject level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

We next examine the concentration parameter K , which captures an additional dimension of belief relevant to the smooth ambiguity model and HEU-B. Holding mean beliefs fixed, higher K indicates a more concentrated second-order belief and therefore lower perceived ambiguity. Consistent with observing more draws, K is significantly higher in Box B than in Box A (Wilcoxon signed-rank test, $p < 0.001$). Since ambiguity level depends on both K and the mean beliefs, we apply the measure of ambiguity level proposed by Izhakian (2017, 2020) to our Dirichlet specification,

$p \sim \text{Dirichlet}(m_1K, m_2K, m_3K)$, yielding

$$\mathcal{U}^2 = \frac{1}{K+1} \sum_{i=1}^3 m_i^2(1 - m_i), \quad (11)$$

where m_i denotes the mean probability for color i . A higher value of $\mathcal{U}^2$ indicates a higher ambiguity level. The median index decreases from 0.0140 in Box A to 0.0092 in Box B (Wilcoxon signed-rank test, $p < 0.001$). However, perceived ambiguity does not decrease for every subject: it increases for 151 of the 364 subjects. We examine the implications of this heterogeneity for predictive performance in Section 5.3.

5.3 Further Analysis of Predictive Performance

Our main results compare the models using beliefs directly elicited from both boxes. In practical applications, however, decision makers' beliefs may not be directly observable and may instead need to be approximated using available information. It is therefore important to examine whether our model ranking is robust to using approximated rather than directly elicited beliefs. To do so, we approximate each subject's beliefs using Bayesian updating based on the observed draws. For each model, we re-estimate ambiguity-attitude parameters using Box A choices under these approximated beliefs, and then combine the estimates with the Bayesian approximations for Box B to predict the held-out choices. Figure 8 reports the results from both between-treatment and within-subject comparisons. In the between-treatment comparison, HEU-B achieves the highest median quadratic score (-0.194), followed by Source-Prelec (-0.217) and the smooth ambiguity model (-0.228). HEU-B significantly outperforms Source-Prelec (Mann-Whitney test, $p = 0.036$) and the smooth ambiguity model ($p < 0.001$), while the difference between the latter two models is not statistically significant ($p = 0.173$). The within-subject comparisons yield the same overall ranking: HEU-B significantly outperforms both alternatives regardless of which model guided question selection, while the smooth ambiguity model and Source-Prelec do not differ significantly in the pooled comparison. Thus, HEU-B's predictive advantage remains robust even when beliefs are approximated rather than directly elicited.²⁷

²⁷We conduct a similar analysis in the benchmark experiment, where beliefs are not directly elicited. When parameters are estimated using the two-draw box and used to predict choices in the sixteen-draw box, the smooth ambiguity model and HEU-B perform similarly, and both significantly outperform Source-Prelec. This may partly reflect difficulties in identifying ambiguity-attitude parameters, particularly under Source-Prelec, from the two-draw data. With relatively few observed draws, subjects may assign similar probabilities to the three colors, generating similar matching probabilities across events and providing less information for parameter estimation.

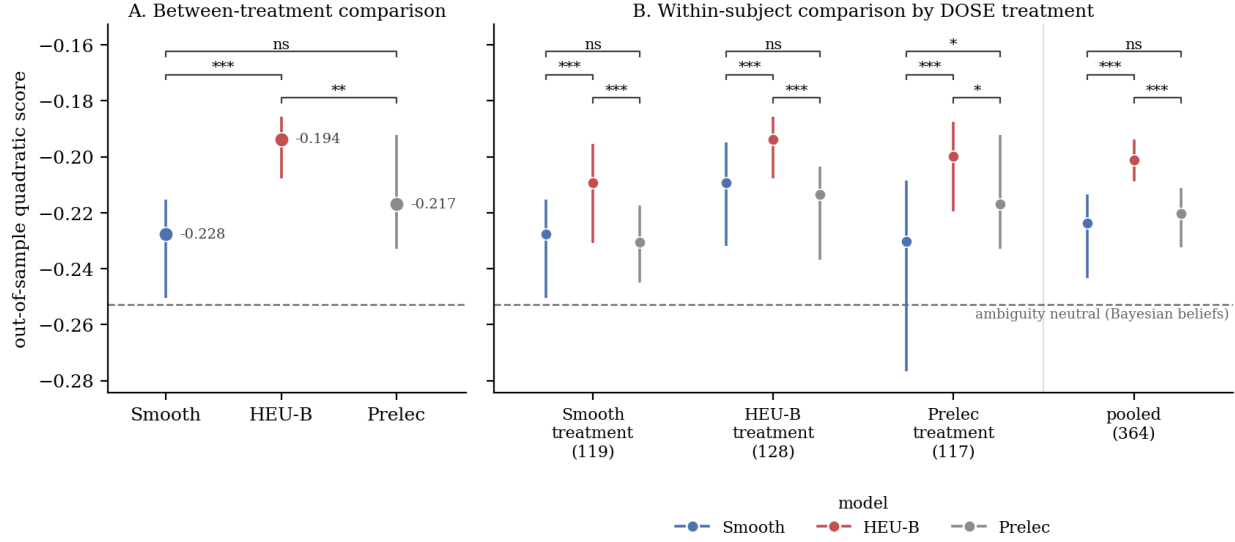


Figure 8: Out-of-sample predictive performance under Bayesian updating. Beliefs in both boxes are approximated using Bayesian updating from a uniform prior based on each subject’s observed draws, rather than directly elicited. Each model’s ambiguity-attitude parameters are estimated using Box A choices under these approximated beliefs and then used to predict held-out choices in Box B. Panel A presents the between-treatment comparison, while Panel B presents the within-subject comparison, separately by DOSE treatment and for the pooled sample. Dots report median quadratic scores, with vertical lines indicating 95% bootstrap confidence intervals obtained by resampling subjects. Higher scores indicate better prediction. Brackets report two-sided Mann-Whitney tests in Panel A and Wilcoxon signed-rank tests of paired differences in Panel B. The dashed line indicates the pooled median score (-0.253) under ambiguity neutrality, where matching probabilities equal the Bayesian first-order beliefs. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; ns indicates no statistical significance at the 10% level.

In both our main comparison and the analysis using Bayesian approximations, the differences in out-of-sample quadratic scores across models may appear relatively small. However, this partly reflects the structure of our evaluation questions rather than necessarily small differences in the models’ predictions. Recall that the held-out questions are selected adaptively to locate subjects’ matching probabilities. As the questions approach a subject’s point of indifference, models that reasonably capture the subject’s behavior tend to assign probabilities closer to 0.5 to each choice. Consequently, differences in predicted choice probabilities, and therefore in quadratic scores across models, tend to become smaller. Thus, the magnitude of differences in quadratic scores may not fully reflect how much the choice of ambiguity model matters.

To assess the magnitude of these differences more directly, we compare the matching probabilities predicted by all three models for the same subjects and events, using each subject’s elicited Box B beliefs for all three models. Thus, only the model specifications and corresponding estimated

ambiguity-attitude parameters differ. For each subject, we calculate the range of predicted matching probabilities across the three models for each event and average these ranges across events. The median of these subject-level averages is 11.8 percentage points, and 59% of subjects have an average difference exceeding 10 percentage points. These results show that different ambiguity models can generate substantially different predictions even for the same subjects under identical beliefs.

The comparison above uses beliefs elicited from the particular ambiguous situations in our experiment. To examine how model differences may vary under other beliefs, we consider one subject whose estimated ambiguity-attitude parameters are closest to the sample medians. Holding these estimates fixed, we compare the models' predicted matching probabilities across a range of first-order beliefs and levels of perceived ambiguity, using identical beliefs for all three models in each case. Specifically, we vary the first-order belief about red²⁸ and consider two levels of belief concentration, $K = 4$ and $K = 60$. Figure 9 presents the results. When $K = 4$, HEU-B and Source-Prelec generate relatively similar predictions, while the smooth ambiguity model differs more from the other two. When $K = 60$, the predictions of the smooth ambiguity model and HEU-B move closer to ambiguity neutrality, whereas those of Source-Prelec remain unchanged, leading to greater differences between Source-Prelec and the other two models. This illustrates how the consequences of choosing different ambiguity models may depend on the situation in which they are applied, particularly the level of perceived ambiguity.

We next explore what may account for HEU-B's superior predictive performance. We begin by examining how frequently each model provides the best prediction for a given subject. Figure 10 ranks the three models according to their out-of-sample quadratic scores for each of the 364 subjects and reports the shares for which each model ranks first, second, or third. The shares for which each model performs best are similar: 32.1% for the smooth ambiguity model, 34.3% for HEU-B, and 33.5% for Source-Prelec. However, the models differ substantially in how often they perform worst. HEU-B ranks last for only 7.7% of subjects, compared with 44.8% for the smooth ambiguity model and 47.5% for Source-Prelec. Instead, HEU-B ranks second for 58.0% of subjects. Thus, HEU-B's overall advantage appears to come more from how rarely it performs worst than from how often it performs best.

Another potentially important factor is the direction of change in perceived ambiguity across

²⁸For each value of the first-order belief about red, the remaining probability is divided equally between the other two colors.

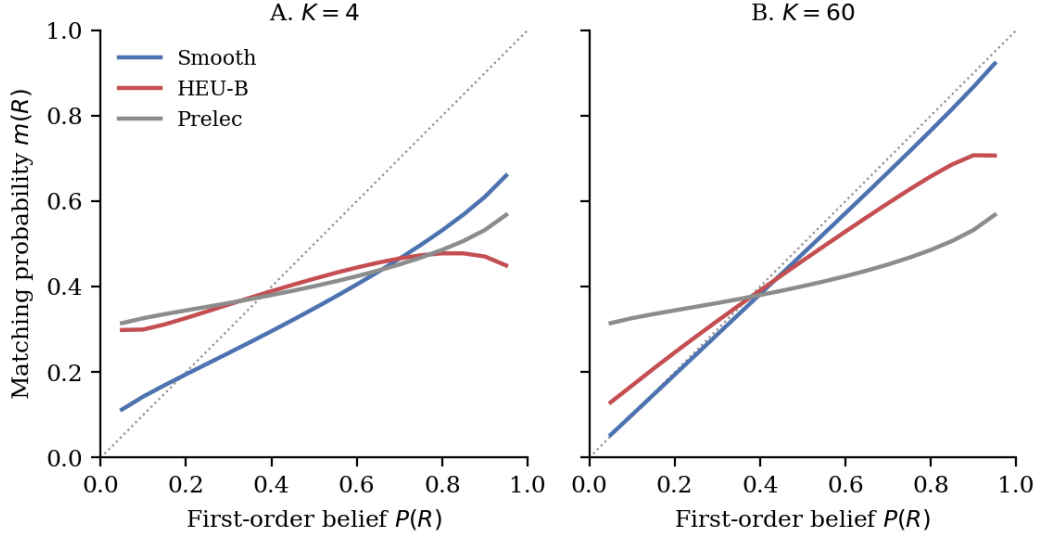


Figure 9: Models' predicted matching probabilities under common beliefs. For each model, ambiguity-attitude parameters are estimated from the same subject's Box A choices. The subject is selected as the one whose posterior-mean ambiguity-attitude estimates are closest to the sample medians. Using these estimates, we calculate each model's predicted matching probability for red under identical hypothetical beliefs. The first-order belief $P(R)$ varies along the horizontal axis, with the remaining probability divided equally between the other two colors. The concentration parameter is fixed at $K = 4$ in Panel A and $K = 60$ in Panel B. The dotted diagonal represents ambiguity neutrality, where $m(R) = P(R)$.

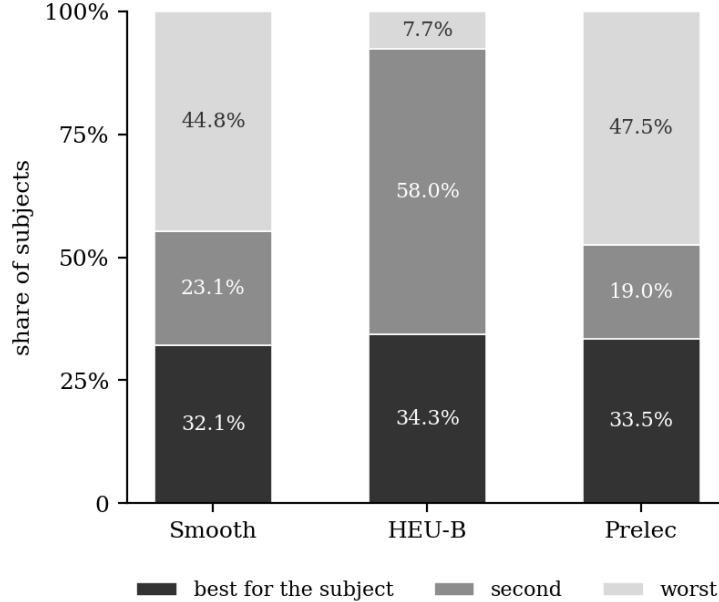


Figure 10: How the models rank across subjects. For each subject, the three models are ranked by their out-of-sample quadratic scores on the same 16 held-out questions. Each bar reports the share of the 364 subjects for whom the corresponding model ranks first, second, or third.

the two boxes. Recall that both the smooth ambiguity model and HEU-B impose structures under which differences in ambiguity attitudes have a smaller effect on behavior when perceived ambiguity is lower. Consequently, ambiguity-attitude parameters may be more difficult to identify when perceived ambiguity is relatively low, because differences in attitudes generate smaller differences in observed choices. Estimating these parameters in a situation with relatively low perceived ambiguity and using them to predict behavior in a situation with higher perceived ambiguity may therefore be particularly challenging. By contrast, Source-Prelec does not incorporate perceived ambiguity as a separate belief dimension. Relative model performance may therefore depend on whether and how models incorporate perceived ambiguity, as well as the direction in which perceived ambiguity changes. Our benchmark experiment also provides evidence that prediction direction may matter for relative model performance, further motivating an examination of this issue in the main experiment.

Although subjects observe more draws in Box B than in Box A, their perceived ambiguity does not necessarily decrease. As reported in Section 5.2, the ambiguity-level index $\mathcal{U}^2$ decreases for 213 subjects but increases for the remaining 151 subjects. The latter pattern contrasts with the direction suggested by Bayesian updating, but is not necessarily unreasonable. For example, subjects may observe draws that conflict with their prior beliefs, making them less confident in their assessment of the box’s composition. Alternatively, observing more draws may make it more difficult for subjects to assess the possible compositions. We also find evidence supporting the credibility of the elicited beliefs among these subjects. Appendix Figure C.1 compares out-of-sample predictions using elicited beliefs with those using Bayesian approximations, separately for subjects whose perceived ambiguity decreased or increased. Even for subjects who perceived more ambiguity after observing additional draws, predictions based on elicited beliefs are significantly better under all three models. This suggests that these subjects’ elicited beliefs meaningfully reflect their behavior, even though their perceived ambiguity changes in the opposite direction from that suggested by Bayesian updating.

We then compare the models’ out-of-sample predictive performance separately for subjects whose perceived ambiguity decreased and those whose perceived ambiguity increased. Figure 11 reports the within-subject comparisons for these two groups. HEU-B significantly outperforms both alternatives in either group, demonstrating that its predictive advantage is robust to the direction of change in perceived ambiguity. The relative performance of the smooth ambiguity model and Source-Prelec, however, differs between the groups. When perceived ambiguity decreases, their

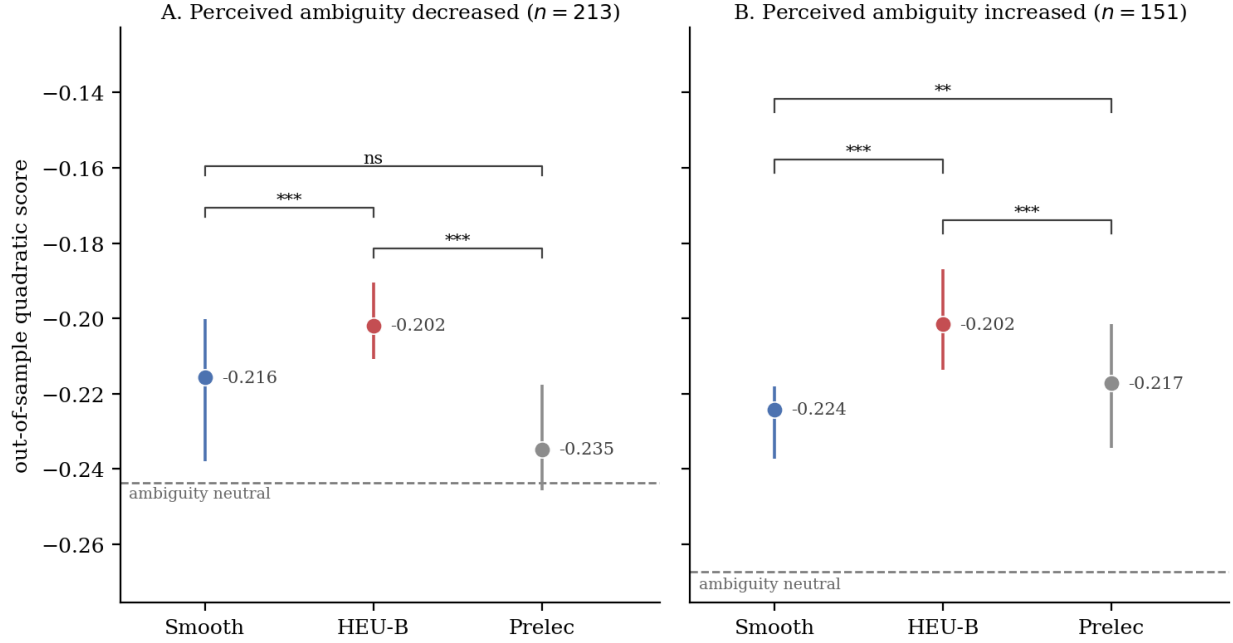


Figure 11: Predictive performance by perceived ambiguity change. Subjects are divided according to whether their ambiguity-level index $\bar{U}^2$ decreased (Panel A) or increased (Panel B) from Box A to Box B. All three models are evaluated on each subject within each group. Dots report median out-of-sample quadratic scores, with vertical lines indicating 95% bootstrap confidence intervals obtained by resampling subjects. Brackets report two-sided Wilcoxon signed-rank tests of paired differences. The dashed lines indicate the median ambiguity-neutral scores within each group (-0.244 where $\bar{U}^2$ decreased and -0.267 where it increased). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; ns indicates no statistical significance at the 10% level.

predictive performance does not differ significantly. When perceived ambiguity increases, Source-Prelec significantly outperforms the smooth ambiguity model. This pattern is consistent with the intuition that estimating ambiguity-attitude parameters under relatively low perceived ambiguity is more challenging for models that incorporate second-order beliefs.

5.4 Implications for Model-Guided Elicitation

Our results also have implications for using DOSE to compare competing models. Because DOSE adaptively selects questions to obtain information about the parameters of a particular model, the questions asked, and hence the responses available for estimation, depend on which model guides the elicitation. This raises a concern about whether model-guided question selection affects the outcome of model comparisons. Our experimental design, with three DOSE treatments guided by different models, allows us to examine this issue.

We begin by comparing the models' in-sample fit. For each subject, we estimate all three

models using responses to the same 12 Box A ambiguity questions and evaluate their fit on those responses. Figure 12 reports the results separately by DOSE treatment and for the pooled sample. The relative in-sample performance varies across treatments, with each model generally performing more favorably when it guides question selection. This suggests that DOSE tends to select questions that favor the guiding model in terms of in-sample fit.

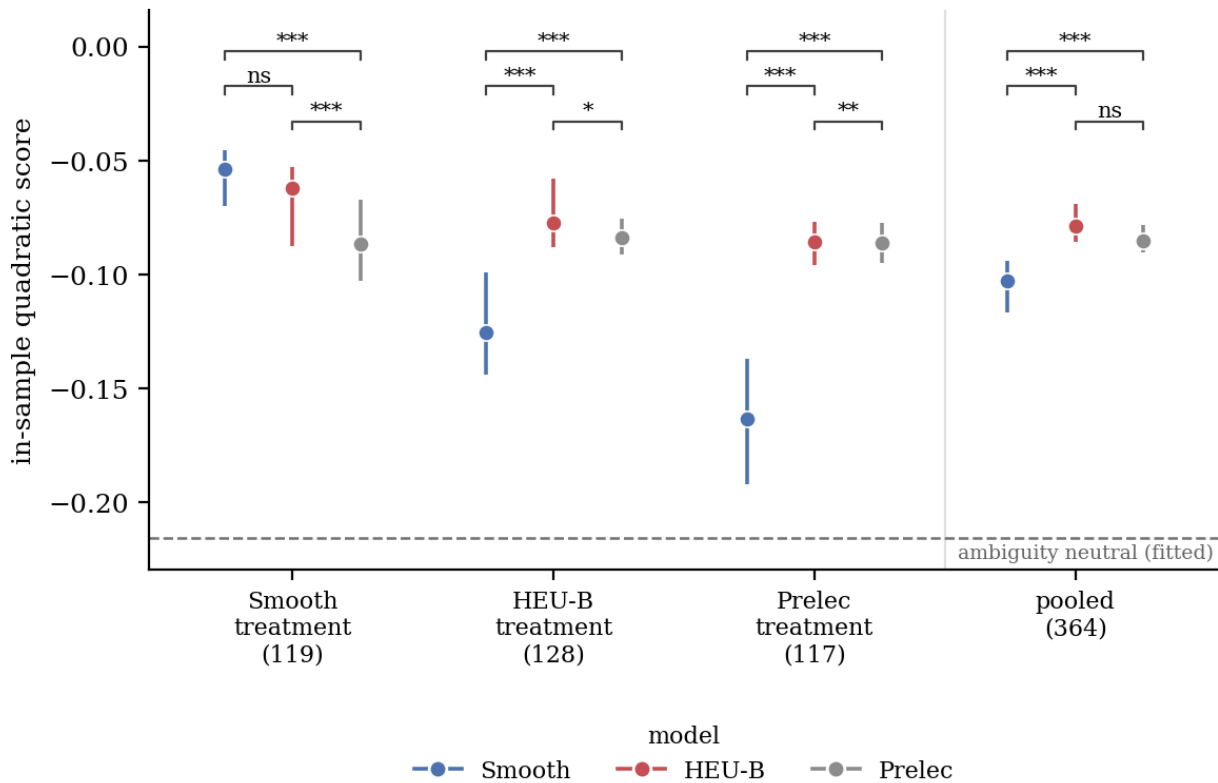


Figure 12: In-sample model comparison by DOSE treatment. For each subject, all three models are fitted and evaluated using responses to the same 12 Box A ambiguity questions. Results are reported separately by DOSE treatment and for the pooled sample. Dots report median in-sample quadratic scores, with vertical lines indicating 95% bootstrap confidence intervals obtained by resampling subjects. Higher scores indicate better fit. Brackets report two-sided Wilcoxon signed-rank tests of paired differences. The dashed line indicates the pooled median in-sample score (-0.215) of the fitted ambiguity-neutral benchmark. $***p < 0.01$, $**p < 0.05$, $*p < 0.10$; ns indicates no statistical significance at the 10% level.

However, the in-sample advantage of the guiding model does not necessarily carry over to out-of-sample comparisons. In our experiment, the 16 held-out ambiguity questions in Box B are selected using a model-independent adaptive procedure, so the evaluation questions are not tailored to any of the competing models. As shown in Figure 7, HEU-B significantly outperforms both alternatives regardless of which model guided the DOSE questions in Box A. Thus, despite

the treatment-dependent in-sample comparisons, HEU-B’s out-of-sample advantage remains robust even when its parameters are estimated using questions selected by a competing model.

These findings have broader implications for using DOSE in model comparison. If model-guided question selection substantially affected out-of-sample comparisons even when evaluation questions were model-independent, establishing robust comparisons might require a separate DOSE treatment for each competing model, substantially increasing the required sample size. Moreover, evaluating newly proposed models using existing data could be problematic if those data were collected using questions guided by other models. Our findings suggest that these concerns may be less severe once models are evaluated on separate, model-independent questions. In our experiment, estimating competing models using questions guided by a single model yields a consistent out-of-sample ranking across treatments. This suggests that DOSE can be used to compare multiple models without necessarily requiring a separate elicitation treatment for each, and that data collected under one guiding model may also be useful for evaluating newly proposed models.

6 Conclusion

In this paper’s main experiment, we systematically evaluate three models of ambiguity based on their out-of-sample predictive power in a different environment with a different level of ambiguity. The first model, Source Theory, represents an agent’s subjective perception of ambiguity entirely through a first-order belief. The other two models, the smooth ambiguity model and a modified version of Hurwicz expected utility (HEU-B), allow for second-order beliefs that capture perceived ambiguity, that is, how confident an agent is in the first-order expectation. The smooth ambiguity model represents perceived ambiguity as a distribution over possible first-order beliefs; HEU-B instead represents it through the gap between pessimistic and optimistic first-order beliefs. HEU-B performs significantly better than the other two models in predictive accuracy, while Source Theory and the smooth ambiguity model are statistically indistinguishable from each other. The superior performance of HEU-B is robust to analyses that rank the models subject by subject, split subjects by changes in perceived ambiguity, and approximate second-order beliefs through Bayesian updating rather than direct elicitation.

The results provide qualified insights into the structure of ambiguity models. Both the smooth ambiguity model and HEU-B represent perceived ambiguity as part of belief, which was expected to increase the transferability of ambiguity attitudes across ambiguous environments. However,

only HEU-B performs better than Source Theory, suggesting that how perceived ambiguity is incorporated into the model matters at least as much as whether it is incorporated at all. Our HEU-B specification is a first attempt to map the measured second-order belief directly into perceived ambiguity in Hurwicz expected utility. Future research could consider alternative mappings and whether other models in the α -maxmin family can further improve predictive performance.

As the comparison to our benchmark experiment demonstrates, these conclusions would not have been reached without the newly developed DOSE-ambiguity procedure. Without DOSE, directly eliciting second-order beliefs is impractical, so we instead impute them by Bayes' rule from observed draws. Under this approach, both HEU-B and the smooth ambiguity model appear to outperform Source Theory in one direction of prediction, with less consistent results in the other. Only the main study, with elicited second-order beliefs and questions chosen to discriminate between models, reveals that HEU-B consistently outperforms the other models. Out-of-sample prediction is similarly important: comparing models on their own treatment's in-sample fit, or on pooled in-sample fit, would not have led to the same conclusions.

Pairing out-of-sample prediction with DOSE also has a methodological implication for how DOSE is used in model comparison. When evaluation is based on model-independent out-of-sample questions, the model ranking is robust to which model guides DOSE question selection. Future applications therefore need not assign a separate treatment to each candidate model, substantially reducing the sample required for model comparison. Our design also provides a way to construct an out-of-sample question block that retains adaptive question selection while remaining model independent.

Our experimental design compares two situations that differ only in the amount of information available, and hence in the level of ambiguity. This provides a clean setting to isolate the role of perceived ambiguity in the relative performance of models. Recent research has focused on ambiguity attitudes toward natural sources (e.g., Abdellaoui et al., 2011; Li et al., 2018; Baillon et al., 2018b; Anantanasuwong et al., 2024; Gutierrez and Kemel, 2024; Lotito et al., 2024) and has documented substantial heterogeneity in ambiguity attitudes across sources. However, almost all of this work uses Source Theory, under which differences in perceived ambiguity may be absorbed into estimated ambiguity attitudes. This leaves open how transferable ambiguity attitudes are across sources under models that account for perceived ambiguity, and how much of the observed behavioral difference instead reflects differences in perceived ambiguity. The answer is particularly important for financial applications. The same investor may make decisions across stocks, bonds,

futures, and other assets, so it is important to know how much information is needed to predict future behavior. In particular, can an ambiguity attitude estimated for one asset be applied to others, or must ambiguity attitudes be measured separately for each type of asset?

References

- Abdellaoui, M., A. Baillon, L. Placido, and P. P. Wakker (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *American Economic Review* 101(2), 695–723.
- Abdellaoui, M., H. Bleichrodt, and C. Gutierrez (2024). Unpacking overconfident behavior when betting on oneself. *Management Science* 70(10), 7042–7061.
- Abdellaoui, M., B. Hill, E. Kemel, and H. Maafi (2025). Learning under ambiguity: An experimental investigation. *Journal of Economic Theory* 230, 106093.
- Ahn, D., S. Choi, D. Gale, and S. Kariv (2014). Estimating ambiguity aversion in a portfolio choice experiment. *Quantitative Economics* 5(2), 195–223.
- Alary, D., C. Gollier, and N. Treich (2013). The effect of ambiguity aversion on insurance and self-protection. *The Economic Journal* 123(573), 1188–1202.
- Anantanasuwong, K., R. Kouwenberg, O. S. Mitchell, and K. Peijnenburg (2024). Ambiguity attitudes for real-world sources: field evidence from a large sample of investors. *Experimental Economics* 27(3), 548–581.
- Antoniou, C., R. D. F. Harris, and R. Zhang (2015). Ambiguity aversion and stock market participation: An empirical analysis. *Journal of Banking & Finance* 58, 57–70.
- Attema, A. E., H. Bleichrodt, and O. L’Haridon (2018). Ambiguity preferences for health. *Health Economics* 27(11), 1699–1716.
- Aydogan, I. (2021). Prior beliefs and ambiguity attitudes in decision from experience. *Management Science* 67(11), 6934–6945.
- Aydogan, I., A. Baillon, E. Kemel, and C. Li (2025). How much do we learn? measuring symmetric and asymmetric deviations from Bayesian updating through choices. *Quantitative Economics* 16(1), 329–365.

- Baillon, A. and H. Bleichrodt (2015). Testing ambiguity models through the measurement of probabilities for gains and losses. *American Economic Journal: Microeconomics* 7(2), 77–100.
- Baillon, A., H. Bleichrodt, U. Keskin, O. l’Haridon, and C. Li (2018a). The effect of learning on ambiguity attitudes. *Management Science* 64(5), 2181–2198.
- Baillon, A., H. Bleichrodt, C. Li, and P. P. Wakker (2025). Source theory: A tractable and positive ambiguity theory. *Management Science* 71(10), 8767–8782.
- 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(2), 115–147.
- Baillon, A., Z. Huang, A. Selim, and P. P. Wakker (2018b). Measuring ambiguity attitudes for all (natural) events. *Econometrica* 86(5), 1839–1858.
- Barham, B. L., J.-P. Chavas, D. Fitz, V. Ríos Salas, and L. Schechter (2014). The roles of risk and ambiguity in technology adoption. *Journal of Economic Behavior & Organization* 97, 204–218.
- Beauchamp, J. P., D. J. Benjamin, D. I. Laibson, and C. F. Chabris (2020). Measuring and controlling for the compromise effect when estimating risk preference parameters. *Experimental Economics* 23(4), 1069–1099.
- Berger, L., H. Bleichrodt, and L. Eeckhoudt (2013). Treatment decisions under ambiguity. *Journal of Health Economics* 32(3), 559–569.
- Bertani, N., E. Diecidue, P. Perny, and P. Viappiani (2026). Fast and simple adaptive elicitations. *Management Science*. Articles in Advance, 12 June 2026.
- Bianchi, M. and J.-M. Tallon (2019). Ambiguity preferences and portfolio choices: Evidence from the field. *Management Science* 65(4), 1486–1501.
- Bland, J. R. and Y. Rosokha (2021). Learning under uncertainty with multiple priors: experimental investigation. *Journal of Risk and Uncertainty* 62(2), 157–176.
- Bleichrodt, H., S. Grant, and J. Yang (2023). Testing Hurwicz expected utility. *Econometrica* 91(4), 1393–1416.

- Brenner, M. and Y. Izhakian (2018). Asset pricing and ambiguity: Empirical evidence. *Journal of Financial Economics* 130(3), 503–531.
- Brown, A. L., H. Guo, and H. Je (2026). Preferences for the resolution of risk and ambiguity. *Journal of Economic Theory* 232, 106151.
- Bryan, G. (2019). Ambiguity aversion decreases the impact of partial insurance: Evidence from African farmers. *Journal of the European Economic Association* 17(5), 1428–1469.
- Cabantous, L. (2007). Ambiguity aversion in the field of insurance: Insurers’ attitude to imprecise and conflicting probability estimates. *Theory and Decision* 62(3), 219–240.
- Cavagnaro, D. R., G. J. Aranovich, S. M. McClure, M. A. Pitt, and J. I. Myung (2016). On the functional form of temporal discounting: An optimized adaptive test. *Journal of Risk and Uncertainty* 52(3), 233–254.
- Cavagnaro, D. R., R. Gonzalez, J. I. Myung, and M. A. Pitt (2013a). Optimal decision stimuli for risky choice experiments: An adaptive approach. *Management Science* 59(2), 358–375.
- Cavagnaro, D. R., M. A. Pitt, R. Gonzalez, and J. I. Myung (2013b). Discriminating among probability weighting functions using adaptive design optimization. *Journal of Risk and Uncertainty* 47(3), 255–289.
- Cecchi, F., R. Lensink, and E. Slingerland (2024). Ambiguity attitudes and demand for weather index insurance with and without a credit bundle: Experimental evidence from Kenya. *Journal of Behavioral and Experimental Finance* 41, 100885.
- Chapman, J., E. Snowberg, S. W. Wang, and C. Camerer (2025a). Dynamically optimized sequential experimentation (DOSE) for estimating economic preference parameters. Working paper, September 25, 2025.
- Chapman, J., E. Snowberg, S. W. Wang, and C. Camerer (2025b). Looming large or seeming small? attitudes towards losses in a representative sample. *Review of Economic Studies* 92(5), 2893–2922.
- Chen, D. L., M. Schonger, and C. Wickens (2016). oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9, 88–97.
- Chew, S. H., B. Miao, and S. Zhong (2017). Partial ambiguity. *Econometrica* 85(4), 1239–1260.

- Cohen, M., I. Gilboa, J. Y. Jaffray, and D. Schmeidler (2000). An experimental study of updating ambiguous beliefs. *Risk, Decision and Policy* 5(2), 123–133.
- Conte, A. and J. D. Hey (2013). Assessing multiple prior models of behaviour under ambiguity. *Journal of Risk and Uncertainty* 46(2), 113–132.
- Cubitt, R., G. van de Kuilen, and S. Mukerji (2020). Discriminating between models of ambiguity attitude: a qualitative test. *Journal of the European Economic Association* 18(2), 708–749.
- Dave, C., C. C. Eckel, C. A. Johnson, and C. Rojas (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty* 41(3), 219–243.
- Dimmock, S. G., R. Kouwenberg, O. S. Mitchell, and K. Peijnenburg (2016b). Ambiguity aversion and household portfolio choice puzzles: Empirical evidence. *Journal of Financial Economics* 119(3), 559–577.
- Dimmock, S. G., R. Kouwenberg, and P. P. Wakker (2016a). Ambiguity attitudes in a large representative sample. *Management Science* 62(5), 1363–1380.
- Dominiak, A., P. Duersch, and J.-P. Lefort (2012). A dynamic Ellsberg urn experiment. *Games and Economic Behavior* 75(2), 625–638.
- Drechsler, I. (2013). Uncertainty, time-varying fear, and asset prices. *Journal of Finance* 68(5), 1843–1889.
- Epstein, L. G. and Y. Halevy (2024). Hard-to-interpret signals. *Journal of the European Economic Association* 22(1), 393–427.
- Ert, E. and S. T. Trautmann (2014). Sampling experience reverses preferences for ambiguity. *Journal of Risk and Uncertainty* 49(1), 31–42.
- Freeman, D. J., Y. Halevy, and T. Kneeland (2019). Eliciting risk preferences using choice lists. *Quantitative Economics* 10(1), 217–237.
- Freundt, J., H. Herz, and L. Kopp (2023). Intrinsic preferences for autonomy. CESifo Working Paper 10342, CESifo.
- Georgalos, K. (2019). An experimental test of the predictive power of dynamic ambiguity models. *Journal of Risk and Uncertainty* 59(1), 51–83.

- Gonzalez-Jimenez, D. (2026). Learning gaps? attitudes and beliefs under described and experienced signals. Working paper, March 25, 2026.
- Gul, F. and W. Pesendorfer (2015). Hurwicz expected utility and subjective sources. *Journal of Economic Theory* 159, 465–488.
- Gutierrez, C. and E. Kemel (2024). Measuring natural source dependence. *Experimental Economics* 27(2), 379–416.
- Halevy, Y. (2007). Ellsberg revisited: An experimental study. *Econometrica* 75(2), 503–536.
- Hey, J. D., 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(2), 81–111.
- Hey, J. D. and N. Pace (2014). The explanatory and predictive power of non two-stage-probability theories of decision making under ambiguity. *Journal of Risk and Uncertainty* 49(1), 1–29.
- Hogarth, R. M. and H. Kunreuther (1989). Risk, ambiguity, and insurance. *Journal of Risk and Uncertainty* 2(1), 5–35.
- Imai, T. and C. F. Camerer (2018). Estimating time preferences from budget set choices using optimal adaptive design. Working paper, under revision; cited as LMU Munich mimeo in Chapman et al. (2025).
- Izhakian, Y. (2017). Expected utility with uncertain probabilities theory. *Journal of Mathematical Economics* 69, 91–103.
- Izhakian, Y. (2020). A theoretical foundation of ambiguity measurement. *Journal of Economic Theory* 187, 105001.
- Johnson, C., A. Baillon, H. Bleichrodt, Z. Li, D. van Dolder, and P. P. Wakker (2021). Prince: An improved method for measuring incentivized preferences. *Journal of Risk and Uncertainty* 62, 1–28.
- Jozwik, J. (2026). Ambiguity and technology adoption under index insurance: Experimental evidence. *Food Policy* 141, 102854.
- Klibanoff, P., M. Marinacci, and S. Mukerji (2005). A smooth model of decision making under ambiguity. *Econometrica* 73(6), 1849–1892.

- Klingebiel, R. and F. Zhu (2023). Ambiguity aversion and the degree of ambiguity. *Journal of Risk and Uncertainty* 67(3), 299–324.
- Kothiyal, A., V. Spinu, and P. P. Wakker (2014). An experimental test of prospect theory for predicting choice under ambiguity. *Journal of Risk and Uncertainty* 48(1), 1–17.
- Li, C., U. Turmunkh, and P. P. Wakker (2019). Trust as a decision under ambiguity. *Experimental Economics* 22(1), 51–75.
- Li, Z., J. Müller, P. P. Wakker, and T. V. Wang (2018). The rich domain of ambiguity explored. *Management Science* 64(7), 3227–3240.
- Liang, Y. (2025). Learning from unknown information sources. *Management Science* 71(5), 3873–3890.
- Lotito, G., A. Maffioletti, and M. Santoni (2024). Testing source influence on ambiguity reaction: Preference and insensitivity. *Journal of Risk and Uncertainty* 69(3), 349–411.
- Mitani, Y. and N. Hanaki (2025). Pay a lot to a few instead of a bit to all! Evidence from online donation experiments. ISER Discussion Paper 1273, Institute of Social and Economic Research, The University of Osaka.
- Moreno, O. M. and Y. Rosokha (2016). Learning under compound risk vs. learning under ambiguity – an experiment. *Journal of Risk and Uncertainty* 53(2-3), 137–162.
- Ngangoué, M. K. (2021). Learning under ambiguity: An experiment in gradual information processing. *Journal of Economic Theory* 195, 105282.
- Peysakhovich, A. and J. Naecker (2017). Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity. *Journal of Economic Behavior & Organization* 133, 373–384.
- Qiu, J. and U. Weitzel (2016). Experimental evidence on valuation with multiple priors. *Journal of Risk and Uncertainty* 53(1), 55–74.
- Toubia, O., E. Johnson, T. Evgeniou, and P. Delquié (2013). Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. *Management Science* 59(3), 613–640.

von Gaudecker, H.-M., A. Wogrolly, and C. Zimpelmann (2025). The distribution and relevance of ambiguity attitudes. Working paper, March 3, 2025.

Williams, C. D. (2015). Asymmetric responses to earnings news: A case for ambiguity. *The Accounting Review* 90(2), 785–817.

A The DOSE validation experiment

The DOSE validation experiment compares DOSE with a conventional bisection procedure under the same question budget. Subjects were recruited through Prolific and randomly assigned to one of the two treatments. Each subject faced an ambiguous box containing balls of four colors and observed ten draws with replacement from the box. Subjects then answered 32 binary choices between a bet on an ambiguous event and a bet with a known winning probability. To ensure comparability, both procedures were restricted to the same eight ambiguous events: four single-color and four three-color events. Under bisection, each event received four questions. Under DOSE, the 32 questions were selected adaptively from a pool of 808 possible questions combining these eight events with different known winning probabilities (Section 2.4).

After these 32 questions, subjects answered 16 held-out questions randomly selected from a separate pool involving events excluded from both procedures' question pools. The selection of held-out questions was independent of treatment. We estimated the Source-Prelec model using the 32 elicitation choices, jointly recovering ambiguity-attitude parameters, first-order beliefs, and a response-error parameter σ . Both treatments used the same grid-based Bayesian estimation procedure. We then evaluated the resulting estimates based on their out-of-sample predictions for the 16 held-out choices. Thus, the treatments differed only in which elicitation questions were asked.

Of the 246 subjects who completed the experiment, 243 passed the attention checks. We further excluded subjects who exhibited near-constant responding.²⁹ The resulting sample contains 236 subjects: 120 in the DOSE treatment and 116 in the bisection treatment.

²⁹We excluded subjects who chose the same option in at least 30 of the 32 elicitation questions or at least 15 of the 16 held-out questions.

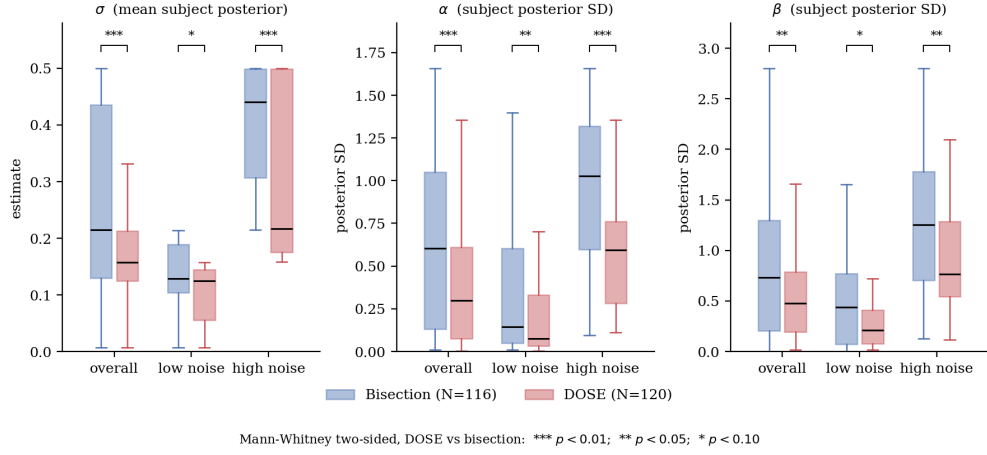


Figure A.1: Response noise and parameter precision in the DOSE validation experiment. The left panel reports the posterior mean of the response-error parameter σ , while the middle and right panels report the posterior standard deviations of the Source-Prelec parameters α and β . Lower values indicate less response noise (σ) or greater precision in parameter estimates (α , β). Results are shown for all subjects and separately for low- and high-noise subjects, classified using the within-treatment median of estimated σ . Boxes represent the interquartile range, with horizontal lines marking the median. Whiskers extend to the most extreme non-outlying observations; outliers are omitted. The estimation grid restricts $\sigma \leq 0.5$. Stars indicate two-sided Mann-Whitney tests comparing DOSE and bisection: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. $N = 236$ (120 DOSE; 116 bisection).

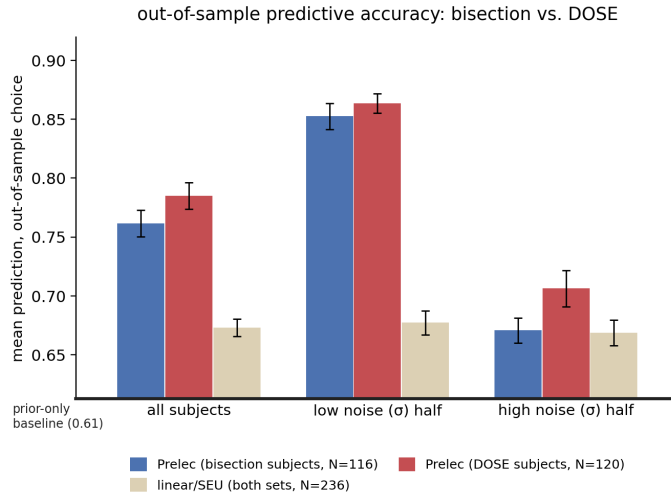


Figure A.2: Out-of-sample predictive accuracy in the DOSE validation experiment. Bars report the mean predicted probability assigned to subjects' actual choices, averaged over the 16 held-out questions and then across subjects. Blue and red bars represent the Source-Prelec model estimated using choices from the bisection and DOSE treatments, respectively. Beige bars represent an ambiguity-neutral benchmark, estimated using the same belief and response-error grids. The vertical axis begins at 0.61, the predictive accuracy obtained using only the prior over the estimation grid, without incorporating subjects' elicitation choices. Results are shown for all subjects and separately for low- and high-noise subjects, classified using the within-treatment median of estimated σ . Error bars represent ± 1 standard error of the mean. Two-sided Mann-Whitney tests comparing DOSE and bisection yield $p = 0.078$ for all subjects, $p = 0.579$ for low-noise subjects, and $p = 0.008$ for high-noise subjects. The observed advantage of DOSE is larger among high-noise subjects. $N = 236$ (120 DOSE; 116 bisection).

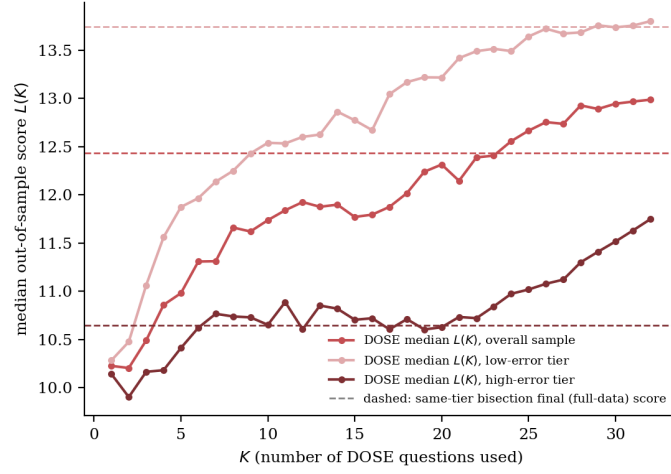


Figure A.3: Out-of-sample predictive performance by number of DOSE questions in the validation experiment. For each K , parameters are re-estimated using only the first K elicitation choices of each subject in the DOSE treatment. The out-of-sample score $L(K)$ (y-axis) is the sum of the predicted probabilities assigned to the subject's actual choices across the 16 held-out questions, with a maximum of 16. Solid lines report the median $L(K)$ across DOSE subjects, for all subjects and separately for low- and high-noise subjects, classified using the within-treatment median of estimated σ . Dashed lines in the same colors report the median out-of-sample scores for the corresponding bisection sample or subgroup, using all 32 elicitation questions. $N = 236$ (120 DOSE; 116 bisection).

B Additional results from the benchmark experiment

Table B.1: Parameter changes under alternative specifications, benchmark experiment

Specification	Parameter	Grid-based Bayesian, $N = 251$			MLE, $N = 114$		
		2 draws	16 draws	Median difference	2 draws	16 draws	Median difference
Prelec	α	0.152	0.440	+0.125***	0.292	0.639	+0.194***
	β	1.031	1.002	-0.021**	1.043	0.995	-0.089***
Neo-additive	a	0.767	0.559	-0.141***	0.565	0.303	-0.182***
	b	0.151	0.056	-0.079***	0.145	0.029	-0.087***
Goldstein–Einhorn	γ	0.166	0.404	+0.110***	0.330	0.594	+0.080**
	δ	0.738	0.896	+0.113***	0.704	0.932	+0.203***
HEU	α	0.611	0.581	-0.035***	0.621	0.546	-0.057***
	β	0.845	0.605	-0.124***	0.919	0.728	-0.026***

Notes: The first three specifications use different source-theory weighting functions: Prelec; neo-additive, $w(p) = (1 - a)p + (a - b)/2$, where a captures likelihood insensitivity and b ambiguity aversion; and Goldstein–Einhorn, $w(p) = \delta p^\gamma / (\delta p^\gamma + (1 - p)^\gamma)$, where γ governs curvature and δ elevation. The fourth specification is Hurwicz expected utility (HEU), following Bleichrodt et al. (2023), where α captures ambiguity attitude and β perceived ambiguity. Beliefs and model parameters are estimated jointly. Columns report the medians of individual parameter estimates for the 2-draw and 16-draw boxes, and the median of within-subject differences (16 draws minus 2 draws). For the grid-based Bayesian estimates, individual estimates are posterior means ($N = 251$). Limited variation in matching probabilities can make it difficult to obtain reliable MLE estimates of the five model parameters. We therefore restrict the MLE analysis to subjects whose six elicited matching probabilities have a variance of at least 0.005 and take at least three distinct values in each box. Of the 251 subjects, 131 meet these criteria in the 2-draw box, 191 in the 16-draw box, and 114 in both boxes. Stars indicate two-sided Wilcoxon signed-rank tests of within-subject differences: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

C Additional details and results from the main experiment

Table C.1: Prior grids used by DOSE, main experiment

Model	Parameter	Grid values
All	m_R, m_Y	0.025 to 0.95 in steps of 0.025 (741 belief points on the simplex)
All	K	3, 3.929, 5.145, 6.738, 8.825, 11.557, 15.135, 19.821, 25.958, 33.996, 44.521, 58.306, 76.358, 100
All	σ	0.02, 0.05, 0.09, 0.15, 0.22, 0.33, 0.5
Source-Prelec	α	0.01, 0.02, 0.032, 0.051, 0.082, 0.131, 0.209, 0.335, 0.535, 0.856, 1.368, 2.189, 3.5, 5
	β	0.05, 0.3, 0.384, 0.491, 0.628, 0.803, 1.027, 1.314, 1.681, 2.15, 2.75, 3.518, 4.5, 10
Smooth	θ	-50, -25, -18.727, -12.455, -6.182, 0.091, 6.364, 12.636, 18.909, 25.182, 31.455, 37.727, 44, 50
	ρ	0.05, 0.09, 0.134, 0.199, 0.295, 0.438, 0.651, 0.967, 1.437, 2.135, 3.172, 4.712, 7, 10
HEU-B	α	0, 0.077, 0.154, 0.231, 0.308, 0.385, 0.462, 0.538, 0.615, 0.692, 0.769, 0.846, 0.923, 1
	κ	0.02, 0.25, 0.344, 0.475, 0.654, 0.901, 1.242, 1.711, 2.358, 3.249, 4.477, 6.169, 8.5, 20

Notes: For each model, we assign equal prior probability to every possible combination of the listed parameter values. The belief grid consists of 741 points (m_R, m_Y, m_B) on the simplex, with each share at least 0.025. σ is the response-error parameter. For each model-specific parameter except HEU-B's α , the grid contains 12 core points spanning approximately the central 90–95% of individual estimates from earlier data, including the benchmark experiment, plus one extreme value at each end to extend the range. Core points are evenly spaced for Smooth θ and log-spaced for the other parameters. HEU-B's α is bounded on $[0, 1]$ and takes 14 evenly spaced values.

Table C.2: Predictive value of elicited ambiguity attitudes, main experiment

<i>Panel A. By treatment: each treatment's own model, its own subjects</i>					
model	n	fitted	neutral	gain	p
Smooth	119	-0.2246	-0.2481	+0.0123	< 0.001
HEU-B	128	-0.1990	-0.2495	+0.0207	< 0.001
Prelec	117	-0.2083	-0.2668	+0.0110	0.005
<i>Panel B. Pooled: every model on all 364 subjects, one shared neutral</i>					
model	n	fitted	neutral	gain	p
Smooth	364	-0.2214	-0.2553	+0.0129	< 0.001
HEU-B	364	-0.2019	-0.2553	+0.0190	< 0.001
Prelec	364	-0.2269	-0.2553	+0.0139	< 0.001

Notes: The table compares out-of-sample quadratic scores from each fitted model with an ambiguity-neutral benchmark that uses elicited beliefs without ambiguity-attitude parameters. The ambiguity-neutral benchmark is identical across the three models. Higher scores indicate better predictive performance. Panel A evaluates each model using subjects from its corresponding DOSE treatment, while Panel B evaluates all three models on the pooled sample of 364 subjects. The fitted and neutral columns report median scores across subjects. Gain is the median within-subject difference (fitted minus neutral). The p -values are from two-sided Wilcoxon signed-rank tests of within-subject differences. $N = 364$.

Table C.3: Out-of-sample model comparisons under alternative specifications, main experiment

	HEU-B vs source	HEU-B vs Smooth	Smooth vs source
<i>Panel A. Between treatment (difference of medians)</i>			
Source-Prelec (main specification)	+0.0093**	+0.0256***	-0.0163
Neo-additive in place of Prelec	+0.0047	+0.0256***	-0.0209**
Goldstein-Einhorn in place of Prelec	+0.0055**	+0.0256***	-0.0201
HEU in place of Prelec	+0.0012	+0.0256***	-0.0245**
Finer attitude grid (20 per axis)	+0.0056*	+0.0246***	-0.0190
<i>Panel B. Within subject, pooled (364 subjects; median paired difference)</i>			
Source-Prelec (main specification)	+0.0106***	+0.0052***	+0.0013
Neo-additive in place of Prelec	+0.0038***	+0.0052***	-0.0085***
Goldstein-Einhorn in place of Prelec	+0.0082***	+0.0052***	-0.0029
HEU in place of Prelec	+0.0011***	+0.0052***	-0.0077***
Finer attitude grid (20 per axis)	+0.0083***	+0.0050***	-0.0029

Notes: Each row reports pairwise differences in out-of-sample quadratic scores under an alternative specification. The second through fourth rows replace the Prelec weighting function with the neo-additive, Goldstein-Einhorn, or the Hurwicz-motivated weighting function of Bleichrodt et al. (2023) (Table B.1), while keeping the HEU-B and smooth ambiguity models unchanged. The fifth row increases the number of grid points for each attitude parameter from 14 to 20, keeping the same ranges. “Source” refers to the source-theory specification used in each row. Panel A compares each model using subjects from its corresponding DOSE treatment and reports differences in median scores, with two-sided Mann-Whitney tests. Panel B evaluates all models on the pooled sample of 364 subjects and reports median within-subject differences, with two-sided Wilcoxon signed-rank tests. Positive values favor the first model in each comparison. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

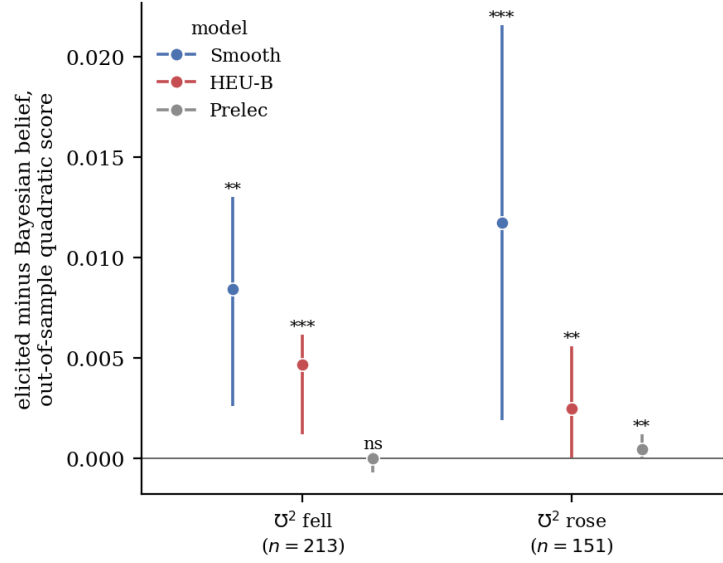


Figure C.1: Out-of-sample predictive performance using elicited versus Bayesian-approximated beliefs. For each model, the figure reports the difference in out-of-sample quadratic scores using subjects' elicited beliefs and Box B beliefs approximated by Bayesian updating from a uniform prior based on the observed draws (elicited minus Bayesian). Positive values indicate better prediction using elicited beliefs. All three models are evaluated on each subject. Subjects are divided according to whether their elicited ambiguity-level index $\mathcal{U}^2$ decreases ($N = 213$) or increases ($N = 151$) from Box A to Box B. Dots represent median differences across subjects, and vertical lines indicate 95% bootstrap confidence intervals obtained by resampling subjects. Stars indicate two-sided Wilcoxon signed-rank tests of within-subject differences against zero: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; ns indicates no statistical significance at the 10% level.

D Instructions

The pages below are the shipped oTree templates of each experiment rendered as a participant saw them, in order. Figure 2 and Figure 3 in the main text show the question pages.

D.1 Benchmark experiment

Instructions

Throughout the study, you will make a series of bets on the color of a ball drawn from a box.

The two boxes

You will be assigned two boxes, labeled A and B, each containing 100 colored balls. Each ball is either **red**, **blue**, or **yellow**.

The color composition of each box (how many balls of each color it contains) is *unknown* to you during the study. The two boxes will very likely have different compositions. At the end of the study, you will be told the true composition of both.

What you'll do

The study has two sections: one for Box A and one for Box B. In each section you will (1) see some example draws from that section's box, then (2) answer 24 binary-choice questions about that box. Each question asks you to choose between two options, Option 1 and Option 2.

For each example draw, one ball is randomly drawn from the box, its color is shown, and the ball is then returned before the next draw. Example draws can help: they give you some information about the box's true color composition. Only a few draws might not give the full picture, but the more draws you see, the closer the observed colors usually are to the true proportions.

NEXT

Figure D.1: Benchmark experiment, instructions page 1 of 4: the two boxes and the structure of the study.

Questions and How You're Paid

Every question in the main study follows the same format. Here's what to expect.

The question format

Every question shows you two options and asks which you prefer. Each option corresponds to drawing one ball from a box. The \$5 is a potential bonus on top of your base payment.

- Option 1: A known box — the number of balls of the specified color (out of 100) IS known. Win \$5 if the drawn ball is the specified color, \$0 otherwise.
- Option 2: An unknown box — the number of balls of the specified color (out of 100) IS NOT known. Win \$5 if the drawn ball is the specified color, \$0 otherwise.

The questions ahead will share the same Option 1 vs. Option 2 format, but the details of each option change from question to question. For example, the color that pays \$5, or how many of that color are in the known box, may differ. **Please read each question carefully before answering.**

Your bonus

In addition to your base payment, you can earn a \$5 bonus. Here's how:

1. At the beginning of the study, one question will be selected at random from all the questions you'll answer. You won't learn which one until the end.
2. At the end of the study, the option you chose on that question will be executed: one ball is drawn from the box you chose, and you win \$5 if that ball is of the specified color (and \$0 otherwise).

Because every question is equally likely to be the selected one, it is in your best interest to choose the option you genuinely prefer on every question — that way, no matter which question is chosen, you always get your preferred bet.

WHY CHOOSING WHAT YOU ACTUALLY PREFER IS BEST FOR YOU

On the next page, we'll ask a few comprehension questions before the main study begins.

NEXT

Figure D.2: Benchmark experiment, instructions page 2 of 4: the question format and how the bonus is paid.

Comprehension Questions

Please answer the comprehension questions below to ensure you understand the task.

You will be able to proceed only when all answers are correct.

Please read each question carefully. If you submit two wrong answers to the same question, you will be asked to return your submission on Prolific.

Click the "Instructions" button below to review the instructions.

▶ INSTRUCTIONS

1. How many of your choices will determine your bonus payment?

☐ (a) All of my choices contribute to my bonus.

☐ (b) The last question in each section determines my bonus.

☐ (c) One single question, randomly selected, determines my bonus.

2. For each section, you will see some example draws from the corresponding box. After each draw, the ball is:

☐ (a) kept aside and not used again.

☐ (b) put back into the box before the next ball is drawn.

☐ (c) replaced with a new ball of a different color.

3. Suppose the question that determines your payment is:

- Option 1: One ball is drawn from a known box (30 out of 100 RED). You win \$5 if the ball is RED.
- Option 2: One ball is drawn from an unknown box. You win \$5 if the ball is RED.

Suppose you chose Option 2. The ball drawn from the known box turns out to be RED, and the ball drawn from the unknown box turns out to be BLUE. Do you win the \$5 bonus?

☐ (a) Yes, because the ball drawn from the known box (Option 1) is red.

☐ (b) No, because the ball drawn from the unknown box (Option 2) is blue.

☐ (c) Yes, because at least one of the two balls is red.

4. Suppose you want to learn the composition of an unknown box of 100 balls. Recall that after each example draw, the ball is put back into the box before the next draw. Which reveals more information about the box's true composition?

☐ (a) Seeing 4 example draws from the box.

☐ (b) Seeing 16 example draws from the box.

☐ (c) Both would be about equally informative.

SUBMIT

Figure D.3: Benchmark experiment, instructions page 3 of 4: the comprehension questions. A second wrong answer to the same question ends the session.

Your Bonus Question

Your bonus will be determined by one of the questions you'll answer in this study. We are randomly selecting which one now, before the task begins.

Click the button below to lock in this random selection. The chosen question will be recorded on our server and shown to you as a short reference code (a meaningless sequence of characters). You can write it down to verify the same code appears at the end of the study, when we reveal what it refers to alongside your answer and bonus.

LOCK IN MY BONUS QUESTION

CONTINUE →

Figure D.4: Benchmark experiment, instructions page 4 of 4: the bonus question is drawn here, before the first question is seen. The page is shown before the draw.

D.2 Main experiment

Instructions

The two ambiguous boxes

You will be assigned two *ambiguous* boxes, labeled A and B, each containing 100 colored balls. Each ball is **red**, **blue**, or **yellow**.

The color composition of each box (the number of red, blue, and yellow balls) is *unknown* to you during the study. At the end of the study, you will be told the true composition of each box.

What you'll do

The study has two sections, one for each box. In each section you will (1) see some example draws from that section's box, then (2) answer questions about that box. This study contains 52 questions in total.

For each example draw, one ball is randomly drawn from the box, its color is shown, and the ball is then returned before the next draw. Example draws can help: they give you some information about the box's true color composition. Only a few draws might not give you the full picture. But as more balls are drawn, the mix of colors you see tends to look more like the true mix in the box.

On the next page, we describe the two types of question you will see.

NEXT

Figure D.5: Main experiment, instructions page 1 of 5: the two boxes and the structure of the study.

The Two Types of Question

In each section you will answer two types of question. Both types ask you to choose between two options, each with a possible \$5 bonus on top of your base payment.

Type 1: A bet on the box's composition

Your bet is about how many balls of one specific color are in the ambiguous box. For each question, you are given two ranges, and you pick the range you think is more likely to contain the actual number of balls of that color.

Here is an example:

- Option 1: Win \$5 if the number of BLUE balls in the ambiguous box is: (i) at least 0, AND (ii) no more than 60.
- Option 2: Win \$5 if the number of BLUE balls in the ambiguous box is: (i) at least 61, AND (ii) no more than 100.

The composition of the box will be shown to you at the end of the study. You win \$5 if the actual number of balls of that color falls in the range you chose, and \$0 otherwise.

Type 2: A bet on the color of a drawn ball

Your bet is about the color of a ball randomly drawn from a box. Each question marks one or more colors as winning colors. You can choose which box you want the ball drawn from: a known box (where you know exactly how many of its balls are a winning color) or the ambiguous box (where you do not).

Here is an example. The winning color for this question is YELLOW.

- Option 1: One ball is drawn from a known box (70 out of 100 YELLOW). You win \$5 if the ball is YELLOW.
- Option 2: One ball is drawn from the ambiguous box. You win \$5 if the ball is YELLOW.

One ball will be randomly drawn from your chosen box. You win \$5 if it is a winning color, and \$0 otherwise.

The details change from question to question. **Please read each question carefully before answering.**

NEXT

Figure D.6: Main experiment, instructions page 2 of 5: the two types of question.

How Your Bonus Is Determined

In addition to your base payment, you have a chance to earn a \$5 bonus.

At the beginning of the study, before you answer any question, the computer will select one question at random from a large collection of possible questions. This question will be used to determine whether you earn the bonus. At the end of the study, we will reveal which question was selected.

How the selected question decides your bonus

- If the selected question is one you answered during the study, that answer determines your bonus.
- If you were not asked it during the study, you will be asked to answer it at the end, and that answer determines your bonus.

Given how the bonus works, it is in your best interest to answer each question based on what you truly prefer. That way, no matter which question is selected, you always get the option you wanted.

WHY ANSWERING HONESTLY IS BEST FOR YOU

On the next page, we'll ask a few comprehension questions before the main study begins.

NEXT

Figure D.7: Main experiment, instructions page 3 of 5: how the bonus is determined.

Comprehension Questions

Please answer the comprehension questions below.

You will be able to proceed only when all answers are correct.

Please read each question carefully. If you submit two wrong answers to the same question, you will be asked to return your submission on Prolific.

You can click the "Instructions" button below to review the instructions.

▶ INSTRUCTIONS

1. How is your bonus payment determined?

☐ (a) All of my answers contribute to my bonus.

☐ (b) The last question I answer determines my bonus.

☐ (c) One single question, selected at random before the questions begin, determines my bonus.

2. Suppose the question that determines your payment is this question:

- Option 1: One ball is drawn from a known box (30 out of 100 RED). You win \$5 if the ball is RED.
- Option 2: One ball is drawn from an ambiguous box. You win \$5 if the ball is RED.

Suppose you chose Option 2. At the end of the study, one ball is drawn from the ambiguous box and it turns out to be BLUE. Do you win the \$5 bonus?

☐ (a) Yes, because a ball drawn from the known box (Option 1) could have been RED.

☐ (b) No, because the ball drawn from the box you chose (Option 2, the ambiguous box) is BLUE, not RED.

☐ (c) It depends on the ball that would have been drawn from the known box (Option 1).

3. Suppose the question that determines your payment is this question:

- Option 1: Win \$5 if the number of RED balls in the ambiguous box is: (i) at least 0, AND (ii) no more than 40.
- Option 2: Win \$5 if the number of RED balls in the ambiguous box is: (i) at least 41, AND (ii) no more than 100.

Suppose you chose Option 1. In which of the following cases do you win the \$5 bonus?

☐ (a) At the end of the study, the true composition of the ambiguous box is revealed, and it contains 23 red balls.

☐ (b) At the end of the study, the true composition of the ambiguous box is revealed, and it contains 66 red balls.

☐ (c) At the end of the study, one ball is drawn from the ambiguous box, and it is red.

4. Before each section's questions, you will see some example draws from that section's box. After each draw, the ball is:

☐ (a) kept aside and not used again.

☐ (b) put back into the box before the next ball is drawn.

☐ (c) replaced with a new ball of a different color.

5. Suppose you want to learn the composition of a box of 100 balls. Which reveals more information about the box's true composition?

☐ (a) Seeing 4 example draws from the box.

☐ (b) Seeing 20 example draws from the box.

☐ (c) Both would be about equally informative.

SUBMIT

Figure D.8: Main experiment, instructions page 4 of 5: the comprehension questions.

66

Your Bonus Question

Your bonus will be determined by one question, selected at random from a large collection of possible questions. We are making that selection now, before the task begins.

Click the button below to lock in this random selection. The chosen question will be recorded on our server and shown to you as a short reference code (a meaningless sequence of characters). At the end you will know which question the code refers to.

LOCK IN MY BONUS QUESTION

CONTINUE →

Figure D.9: Main experiment, instructions page 5 of 5: the bonus question is drawn here, before the first question is seen. The page is shown before the draw.