COARSE BELIEF ELICITATION[†]

PAUL J. HEALY[*] & GREG LEO[**]

ABSTRACT. Researchers commonly elicit a precise beliefs about the probability of an event. But often, precision is not required, and it is sufficient to determine which of some set of probabilities is closest to a participant's belief. We show that, among the many suitable scoring rules for eliciting a precise belief, only quadratic rules are suitable for eliciting coarsely. However, scoring rules are not the only way to elicit a belief. We propose two types of simple menu procedures as an alternative to using a quadratic scoring rule to elicit coarsely. These procedures are simple, flexible, and incentive compatible under assumptions nearly universal to experimental economics.

Keywords: Belief Elicitation; Experimental Methods

JEL Classification: C90, D83

Preliminary
June 12, 2022

[†]Thanks y'all.

[*]Dept. of Economics, The Ohio State University; healy.52@osu.edu.

[**]Dept. of Economics, Vanderbilt University; g.leo@vanderbilt.edu.

I. INTRODUCTION

Many choices hinge on beliefs about an uncertain outcome. Bring an umbrella to work if it is likely enough to rain. Buy trip insurance if there is a good chance something will prevent travel. Go to the coffee shop if it is not likely to be busy.

In this paper, we focus on beliefs about the probability an uncertain event will occur. Measuring beliefs of this type is common in economics experiments. Experimental researchers often incentives belief elicitation using scoring rules.[1] A scoring rule maps the elicited belief and the true value (or sample from an unknown distribution) into compensation. For instance, *the quadratic scoring rule* pays an amount proportional to one minus the squared error of the prediction. If a participant reports they believe the likelihood of an event is 75% and the event occurs, the squared prediction error is $0.25^2 = 0.125$. If the maximum payment is \$10, then the rule pays $\$10 * (1-0.125) = \$8.75$. If the event does not occur, the squared prediction error is $0.75^2 = 0.5625$ and the rule pays $\$10 * (1-0.5625) \approx \$4.38$.

A scoring rule is **proper** if it is in a participant's best interest to report their true belief. A rule may be proper for a certain class of prefrences. For instance, the quadratic scoring rule is proper for risk-netural expected utility maximizers.

To expand the class of preferences for which a rule is proper, it is common practice in experimental economics to "binarize" scoring rules by using simple lotteries instead of certain amounts of money for compensation. A binarized version of the quadratic scoring rule pays a lottery over two amounts of money, where the probability of the larger amount is proportional to one minus the squared prediction error. This binarization process extends the class of preference for which a rule is proper from risk-neutral expected utility maximizers to all expected utility maximizers and some non-expected utility maximizers (Hossain and Okui, 2013).

There are many proper scoring rules for eliciting probabilistic beliefs (Gneiting and Raftery, 2007). However, until now, theoretical analysis and common usage has focused on eliciting a *precise belief*. However, precision is not always required. In many decision problems, optimal behavior a single or a small number of thresholds. For instance, the best response in a 2x2 coordination game is contingent on a single threshold of the belief about the opponent's action.

In empirical settings, it is common to use coarse response scales. A familiar form of likelihood elicitation in survey studies involves asking participants whether they think some event is "extremely unlikely", "unlikely", "neutral", "likely", "extremely likely" (Vagias, 2006). An analogous quantitative elicitation might attempt to determine which probability is closest to a subject's belief: $\{0.1, 0.3, 0.5, 0.7, 0.9\}$, thus partitioning beliefs into the intervals $[0.0, 0.2], [0.2, 0.4], [0.4, 0.6], [0.6, 0.8], [0.8, 1.0]$— a sort of quantitative Likert scale for beliefs.

If a participant can only select from some coarse set of beliefs, does every binarized proper scoring rule incentivize choosing the one closest to their true belief? *No.* For example, **the binarized spherical rule** pays a binary lottery where the probability of the "good" outcome is $t(p) = \frac{p}{(p^2+(1-p)^2)^{\frac{1}{2}}}$ if the event occurs and $f(p) = t(1-p)$ if the event does not occur. This is a proper scoring rule, so if a participant can reveal any belief in

[1]See Schotter and Trevino (2014); Schlag et al. (2015); Charness et al. (2021) for surveys of belief elicitation in experimental economics.

$[0,1]$, revealing their true belief $\tilde{p}$ will result in the largest expected probability of the good outcome. However, suppose the experimenter only needs to know which of $\{0, 0.5, 1\}$ a participant's belief is closest to. A participant with belief $\tilde{p} = \frac{2}{3}$ is better off claiming their belief is closest to 1, despite having a belief closer to 0.5.

Under what conditions will a binarized scoring rule provide incentive for participants to *always* reveal the belief closest to their own? We show that, despite the huge variety of proper scoring rules, only quadratic rules have this property. This provides a surprising new justification for the binarized quadratic scoring rule.

However, this result also has a downside. Recently, Danz et al. (2020) and Dustan et al. (2021) have presented evidence of accuracy issues in studies using the binarized quadratic scoring rule, especially eliciting beliefs away from 50%. If, indeed, the quadratic scoring rule has replicable accuracy issues, there is no other proper scoring rule to turn to that can elicit coarsely.

Ultimately, applying rules that *can* elicit precisely to elicit coarsely does not let the experimenter leverage their willingness to reduce precision to improve incentives. To provide an alternative to using a coarse version of the quadratic scoring rule, we present simpler procedures for eliciting coarsely.

The menu procedures we present are easy to implement, easy to explain, take little experiment time. Our novel ternary menu procedure can categorize beliefs into any symmetric partition of the unit interval using a number of menus roughly half the number of sets in that partition. For instance, in our Likert scale example above, we proposed classifying beliefs into the intervals $[0.0, 0.2], [0.2, 0.4], [0.4, 0.6], [0.6, 0.8], [0.8, 1.0]$. This can be accomplished by asking participants only two questions: "Would you rater have \$10 if the event happens, \$10 if it does not happen, or \$10 with a 80% chance?" and "Would you rater have \$10 if the event happens, \$10 if it does not happen, or \$10 with a 60% chance?". Furthermore, these procedures have straight-forward incentives. In fact, they are incentive compatible under preference assumptions that are essentially universal to economics experiments.

In section II of this paper, we focus on identifying which scoring rules, when binarized, can be used for coarse elicitation of the belief about the probability p of some event e. In section III, we present menu procedures that serve as an alternative to coarse elicitation through scoring rules. Section IV concludes.

II. Coarse Scoring

Scoring Rules and Definitions

In a binarized scoring rule, a participant who states their belief about the probability of the event is p is compensated with a lottery that pays the "good outcome" with probability $t(p)$ if e occurs and pays the good outcome with probability $f(p)$ if it does not occur.

If the participant treats subjective risk and objective risk identically and reduces compound lotteries, the participant is indifferent between the subjective lottery resulting from the binarized scoring rule and one that pays the good outcome with probability: $\tilde{p}t(p) + (1-\tilde{p})f(p)$.

A scoring rule is **strictly proper** is $p = \tilde{p}$ is the unique maximizer of the above expression. Experimental economists often use the **binarized quadratic scoring rule** where $t(p) = 1-(1-p)^2$ and $f(p) = t(1-p) = 1-p^2$ (see, for example, Babcock et al., 2017 or Dianat et al., 2018). However, there are other strictly incentive compatible choices for $t()$ and $f()$, for instance, the **the spherical rule** we mentioned in the introduciton: $t(p) = \frac{1-p}{(p^2+(1-p)^2)^{\frac{1}{2}}}$ and $f(p) = t(1-p)$.

Both of these rules (and many others) are incentive compatible (strictly proper) for eliciting a *precise* belief. But, suppose the experimenter only allows the participant to submit a belief from some finite set $P = \{p_1, ..., p_n\}$. We refer to such a set P as a **coarse belief set**. A desirable property of a scoring rule is that, when reports are limited to a coarse belief set, participants will choose the report closest to their true belief. We formalize this as follows:

Let $p^* = Arg.Min_{p \in P}(|\tilde{p} - p|)$. A scoring rule **elicits the nearest belief** for coarse belief set P if for all $\tilde{p} \in [0,1]$:

$$p^* = Arg.Max_{p \in P}[\tilde{p}t(p) + (1-\tilde{p})f(p)]$$

A scoring rule is **proper for coarse beliefs** if is elicits the nearest belief for any coarse belief set P. Note that, eliciting a coarse belief set P, also categorizes participants beliefs into a finite set of closed intervals $\mathscr{P}$ such that $\bigcup \mathscr{P} = [0,1]$ and for any two $S, S' \in \mathscr{P}$, $\#\left(S \cap S'\right) \leq 1$.. That is, $\mathscr{P}$ covers the unit interval and the sets overlap at most one point. We refer to such sets of intervals that cover $[0,1]$ as ***coarse belief categorizations***. In the case of eliciting a nearest belief from a coarse belief set P, the endpoints of the intervals (besides 0 and 1) in the induced coarse belief categorization are the points halfway between each pair of closest elements in the set P. For instance, eliciting the coarse belief set $P = \{0, 0.375, 0.625, 1\}$ induces the coarse belief categorization $[0,0.25], [0.25,0.5], [0.5,0.75], [0.75,1]$.

Scoring Rules Proper for Coarse Beliefs are Quadratic

In this section, we demonstrate that among the many incentive compatible scoring rules, *only quadratic rules are proper for coarse beliefs*.

Define $G(p) = pt(p) + (1-p)f(p)$, which represents the expected payment function under truth telling. A scoring rule is proper if and only if $G()$ is convex (Gneiting and Raftery, 2007). However, only rules with quadratic and convex $G()$ are proper for coarse beliefs.

Proposition 1. A strictly incentive compatible scoring rule is proper for corase beliefs if and only if $G(p)$ is quadratic and convex.

Proof. Any proper scoring rule must have:

$$pt'(p) + (1-p)f'(p) = 0. \tag{1}$$

Consider a grid $P = \{p_1, p, \ldots\}$.[2] Incentive compatibility for coarse beliefs requires that an agent with belief $\frac{p_1+p}{2}$ is indifferent between announcing p_1 and announcing p. Thus,

$$\frac{p_1+p}{2}t(p_1) + (1 - \frac{p_1+p}{2})f(p_1) = \frac{p_1+p}{2}t(p) + (1 - \frac{p_1+p}{2})f(p). \tag{2}$$

[2] p_1 may be equal to zero.

This must be true for any p, so take derivatives of both sides with respect to p and rearrange to get:

$$p_1 t'(p) + (1-p_1) f'(p) = t(p_1) - f(p_1) + f(p) - t(p) - \left[p t'(p) + (1-p) f'(p) \right]. \tag{3}$$

By incentive compatibility the term in brackets evaluates to zero, giving

$$p_1 t'(p) + (1-p_1) f'(p) = t(p_1) - f(p_1) + f(p) - t(p). \tag{4}$$

Now consider a grid $P = \{\ldots, p, p_n\}$.[3] Now an agent with belief $\frac{p_n + p}{2}$ must be indifferent between announcing p and p_n:

$$\frac{p_n + p}{2} t(p) + \left(1 - \frac{p_n + p}{2}\right) f(p) = \frac{p_n + p}{2} t(p_n) + \left(1 - \frac{p_n + p}{2}\right) f(p_n). \tag{5}$$

Taking a derivative gives

$$p_n t'(p) + (1-p_n) f'(p) = t(p_n) - f(p_n) + f(p) - t(p) - \left[p t'(p) + (1-p) f'(p) \right] \tag{6}$$

$$= t(p_n) - f(p_n) + f(p) - t(p). \tag{7}$$

Subtracting equations (6) from (4) and defining $T = t(p_n) - t(p_1)$ and $F = f(p_1) - f(p_n)$ gives

$$t'(p) - f'(p) = \frac{T+F}{p_n - p_1}.$$

Since $G(p)$ is differentiable, we have

$$t(p) = G(p) + (1-p) G'(p) \quad \text{and} \tag{8}$$

$$f(p) = G(p) - p G'(p), \tag{9}$$

or

$$t'(p) = (1-p) G''(p) \quad \text{and} \tag{10}$$

$$f'(p) = -p G''(p). \tag{11}$$

Substituting these into equation (8) gives

$$G''(p) = \frac{T+F}{p_n - p_1}. \tag{12}$$

We therefore have that G is quadratic and convex.

Let $G(p) = a + bp + cp^2$ with $c > 0$, which gives

$$t(p) = (a+b) + 2cp - cp^2 \quad \text{and} \tag{13}$$

$$f(p) = a - cp^2. \tag{14}$$

Note that $t(p)$ runs from $a+b$ at $p = 0$ up to $a+b+c$ at $p = 1$, while $f(p)$ runs from a at

[3] p_n may be equal to one.

$p = 0$ down to $a - c$ at $p = 1$. Any a, b, c combination such that these extreme values lie in $[0,1]$ is acceptable.

It is not hard to verify that any scoring rule of the form given by equation (13) indeed satisfies incentive compatibility for coarse beliefs. □

III. Limitations of Coarse Scoring

IV. Menus Procedures for Coarse Belief Elicitation

As defined in the previous section, any rule that is incentive compatible for coarse beliefs induces a coarse belief categorization. While the previous section showed that quadratic rules are the only *binarized scoring rules* that can elicit any coarse belief categorization, there are other types of procedures that can be used. In this section, we consider directly eliciting coarse belief categorizations through simple menu procedures.

Our menu procedures have the benefit of requiring weaker assumptions for incentive-compatibility. The incentive compatibility condition for binarized proper scoring rules is based on the maximization the expected probability of the good outcome from a compound lottery. Thus, binarized scoring rules rely on the reduction of compound lotteries for incentive compatibility. On the other hand, our menu procedures are incentive compatibility under statewise-monotonicity, a weaker dominance relation on compound lotteries that is essentially universal in experimental economics (Azrieli et al., 2018).

Menu Procedures

A choice-from-sets procedure is a set of menus $\{M_1, ..., M_n\}$. Participants choose one objects from each menu. A menu is randomly chosen, and the participant is compensated with their choice from that chosen menu.

A choice-from-sets experiment ***elicits a belief categorization*** if for every set in the belief categorization, participants with beliefs in the interior of the set choose differently than participants with beliefs on the interior of any other set.

A *binary belief procedure* is a choice experiment involving several ***binary menus*** of the form:

a if E	a with p

A *ternary belief procedure* is a choice experiment involving several ***ternary menus*** of the form:

a if E	a if E'	a with p

In the above menus, a should be understood as some form of compensation for instance, some amount of money, and the alternative is implicitly some outcome b that is less preferred. For instance, a might be \$10 and b might be \$0. Since a, b are fixed across menus, p is the only parameter of each menu. Thus, either type of procedure can be characterized by the set objective probabilities appearing in the menus: $\{\pi_1, ..., \pi_n\}$.

In the following sections, we demonstrate that any coarse belief categorization can be elicited using a binary belief procedure with $\#(\mathscr{P}) - 1$ menus (one less than the number of intervals in $\mathscr{P}$). Furthermore, define ***coarse belief categorization*** to be ***symmetric*** if and only if $[p, p'] \in \mathscr{P} \Leftrightarrow [1 - p', 1 - p] \in \mathscr{P}$. We demonstrate that a symmetric belief elicitation can be elicited using a ternary belief procedure with $\lceil \frac{\#(\mathscr{P})-1}{2} \rceil$. That is, roughly half the number of menus.

Preference Assumptions

Let $\{a,b\}$ be two outcomes such that $a \succ b$. $e \in 2^{\mathscr{S}}$ in which $\mathscr{S} = \{...,s,...\}$ are states of the world. Our menu methods make use of the following lottery types:

Simple subjective lotteries of the form $S = \left(A \circ e, B \circ e'\right)$.
Simple objective lotteries of the form $O = (A \circ p, B \circ (1-p))$.
Simple mixtures *of the form* $M = (p_1 \circ L_1, p_2 \circ L_2, ..., p_n \circ L_n)$ *where each* L_i *is either a simple subjective or simple objective lottery,* $p_i \in [0,1]$ *and* $\sum_{i=1}^{n} p_i = 1$.

We make the following assumptions about $\succsim$, the preference relation on simple mixtures:

1. ***Order:*** $\succsim$ is complete and transitive.
2. ***Replacement:*** $\forall e \in E, \exists p \in [0,1]$: $\left(a \circ e, b \circ e'\right) \sim (a \circ p, b \circ (1-p))$.
3. ***Complementarity:*** $\left(a \circ e, b \circ e'\right) \sim (a \circ p, b \circ (1-p)) \Leftrightarrow \left(a \circ e', b \circ e\right) \sim (a \circ (1-p), b \circ p)$.
4. ***Monotonicity:*** $(a \circ p, b \circ (1-p)) \succsim \left(a \circ p', b \circ \left(1-p'\right)\right) \Leftrightarrow p > p'$.
5. ***State-wise Monotonicity:*** $(p_1 \circ L_1, ..., p_i \circ L_i, ..., p_n \circ L_n) \succ \left(p_1 \circ L_1, ..., p_i \circ \tilde{L_i}, ..., p_n \circ L_n\right) \Leftrightarrow L_i \succ \tilde{L_i}$

Eliciting Coarse Categorizations with Binary Menus

A coarse belief categorization can be characterized by the interval endpoints. Formally this set of endpoints is:

$$P = \{p | p = max(S)\ for\ some\ S \in \mathscr{P}\ \&\ p \in [0,1)\}.$$

Order elements of P into the sequence $(p_1, ..., p_n)$ such that for $i > j$, $p_i > p_j$. This is the sequence of *upper interval thresholds.* Note that there is one upper interval threshold for each set in $\mathscr{P}$ except the set that has endpoint 1. Thus, there are $\#(\mathscr{P})-1$ upper interval thresholds.

Example 1: For $\mathscr{P} = \{[0,0.25],[0.25,0.5],[0.5,0.75],[0.75,1]\}$, *the upper interval thresholds are:* $(0.25, 0.5, 0.75)$.

Example 2: For $\mathscr{P} = \{[0,0.2],[0.2,0.4],[0.4,0.6],[0.6,0.8],[0.8,1]\}$, *the upper interval thresholds are:* $(0.2, 0.4, 0.6, 0.8)$.

Proposition 2. Under assumptions 1,2,4,5 a belief categorization characterized by upper interval thresholds $(p_1, ..., p_n)$ can be elicited by a ternary lottery precedure with objective probabilities $(p_1, ..., p_n)$.

Proof in Appendix.

For instance, the belief categorization $[0,0.1],[0.1,0.3],[0.3,0.7],[0.7,1]$ *can be elicited with the following menus:*

\$x if E	\$x if E'	\$x with 70%
\$x if E	\$x if E'	\$x with 30%
\$x if E	\$x if E'	\$x with 10%

Eliciting Symmetric Coarse Categorizations with Ternary Menus

To characterize a symmetric coarse belief categorization, it is sufficient to enumerate the upper interval thresholds at least as large as 0.5.

Example 1: For $\mathscr{P} = \{[0,0.25],[0.25,0.5],[0.5,0.75],[0.75,1]\}$*, the upper interval threshold at least as large as* 0.5 *are:* $(0.5,0.75)$.

Example 2: For $\mathscr{P} = \{[0,0.2],[0.2,0.4],[0.4,0.6],[0.6,0.8],[0.8,1]\}$*, the upper interval thresholds at least as large as* 0.5 *are:* $(0.6,0.8)$.

Proposition 3. Under assumptions 1-5, a symmetric belief categorization characterized by upper interval thresholds at least as large as 0.5: $(p_1,...,p_n)$ can be elicited by a ternary lottery precedure with objective probabilities $(p_1,...,p_n)$.

Proof in Appendix.

For instance, the belief categorization $[0,0.25],[0.25,0.5],[0.5,0.75],[0.75,1]$ *can be elicited with the following menus:*

\$x if E	\$x if E'	\$x with 75%
\$x if E	\$x if E'	\$x with 50%

Comparison to the Binarized Quadratic Scoring Rule

Scoring rules are inherently precise. On the other hand, the precision of our menu procedures is adjustable. It is interesting to highlight one precise relationship between these procedures. Suppose you increase the precision of a symmetric belief categorization by adding and shrinking the size of the intervals. In doing this, the precision of the procedure increaes. In the limit, the menu procedure becomes a precise rule. In fact, is converges to the binaraized quadratic scoring rule.

To see this, note that in each menu procedure along the way to this limit, the participant chooses all objective lotteries that pay more than $\tilde{p}$ and otherwise prefers the subjective lottery (that pays with probability $\tilde{p}$).

As the number of lotteries grows, the probability one of the menus where the participant chose the objective lottery approaches $2(1-p)$. The expected probability of the good prize in these lotteries approaches $p+\frac{1-p}{2}$. The remainder of the lotteries pay the subjective probability p and occur with $2p-1$. Thus, as the number of menus grows, the expected probability of the good prize under truth telling (that is, the function $G()$ is given by:

$$G(\tilde{p}) = (2-2p)\left(p+\frac{1-p}{2}\right)+(2p-1)p = \tilde{p}^2-\tilde{p}+1$$

The binarized quadratic scoring rule pays the good outcome with probability $t(p) = 1-(1-p)^2$ if the even occurs and $f(p) = 1-p^2$ if it does not. Thus the expected probability of the good outcome under truth-telling of the quadratic binarized scoring rule is:

$$G(\tilde{p}) = \tilde{p}\left(1-(1-\tilde{p})^2\right) + (1-\tilde{p})\left(1-\tilde{p}^2\right) = \tilde{p}^2 - \tilde{p} + 1$$

V. Conclusion

In this paper, we have provided the first formal analysis of coarse belief elicitation. In many settings, precision is not needed, and a coarse elicitation will suffice- eliciting which of a small set of probailibites is closest to a participant's true belief. It is possible to adapt the commonly used binarized quadratic scoring rule for this prupose, but adapting a precise rule to elicit coarsely does not let the experiment leverage their willingness to give up precision to improve incentives. As an alternative, we have outlined menu procedures which achive this goal. Our menu procedures are simple, robust, and flexible.

In this paper, we have focused on eliciting probabilistic beliefs, and this work raises questions about how coarse elicitation can be achieved when other types of beliefs are required. Our menu procedures can be used for eliciting quantiles. For instance, it is possible to elicit a coarse median belief about some quantity X with questions of the form: *"would you rather have* \$10 *with a* 50% *or* \$10 *if quantity X is below Y"*. Changing the 50% to another probability adjusts this to eliciting other quantiles. However, it is not clear how such simple procedures could be used for eliciting means or other moments. Furthermore, we have not attempted to extend our charactarization of proper scoring rules for coarse beliefs to proper scoring rules for these other types of beliefs.

References

Azrieli, Y., Chambers, C. P., and Healy, P. J. (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, 126(4):1472–1503.

Babcock, L., Recalde, M. P., Vesterlund, L., and Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, 107(3):714–47.

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

Danz, D., Vesterlund, L., and Wilson, A. J. (2020). Belief elicitation: Limiting truth telling with information on incentives. Technical report, National Bureau of Economic Research.

Dianat, A., Echenique, F., and Yariv, L. (2018). Statistical discrimination and affirmative action in the lab.

Dustan, A., Koutout, K., and Leo, G. (2021). Reduction in belief elicitation. *Working paper*.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378.

Hossain, T. and Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, 80(3):984–1001.

Schlag, K. H., Tremewan, J., and Van der Weele, J. J. (2015). A penny for your thoughts: A survey of methods for eliciting beliefs. *Experimental Economics*, 18(3):457–490.

Schotter, A. and Trevino, I. (2014). Belief elicitation in the laboratory. *Annu. Rev. Econ.*, 6(1):103–128.

Vagias, W. M. (2006). Likert-type scale response anchors. *Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University*.

VI. APPENDIX

Proof of Proposition 2

For simplicity denote the lottery a if e as S_e, a if e′ as $S_{e'}$ and a with probability p as O_p.

By state-wise monotonicity (5) and because the binary lottery procedure is a choice-from-sets-experiment, a participant will choose their most preferred lottery in each menu (see Azrieli et al., 2018). Under replacement (2), there is some $\tilde{p}$ (the participant's belef) such that $\left(a \circ e, b \circ e'\right) \sim (a \circ \tilde{p}, b \circ (1-\tilde{p}))$.

Let p_i be the largest upper interval threshold smaller than $\tilde{p}$. For each menu with objective probability $p < p_i$, $S_e \sim O_{\tilde{p}} \succ O_p$ by monotonicity (4). By transitivity (1), $S_e \succ O_{\tilde{p}}$. For each menu with objective probability $p > p_i$, by monotonicity (4) and transitivity (1), $O_p \succ S_e$. If $p = p_i$ for some menu, then the participant is not on the interior a set in the belief categorization. Thus, a participant chooses S_e in all menus with $p < p_i$ and O_p in all menus with $p > p_i$. A participant with belief in the interior of the i_{th} interval chooses S_e in the first $i-1$ menus and O_p in the remaining. Thus, the choices of participants on the interior of each set is different.

Proof of Proposition 3

For simplicity denote the lottery a if e as S_e, a if e′ as $S_{e'}$ and a with probability p as O_p.

By state-wise monotonicity (5) and because the ternary lottery procedure is a choice-from-sets-experiment, a participant will choose their most preferred lottery in each menu (see Azrieli et al., 2018). Under replacement (2), there is some $\tilde{p}$ (the participant's belef) such that $\left(a \circ e, b \circ e'\right) \sim (a \circ \tilde{p}, b \circ (1-\tilde{p}))$.

Suppose $\tilde{p} > 0.5$**:** By complementarity (3), $S_e \sim O_{\tilde{p}}$ and $S_{e'} \sim O_{1-\tilde{p}}$. By monotonicity (4) and since $\tilde{p} > 0.5$, $O_{\tilde{p}} \succ O_{1-\tilde{p}}$. By transitiity (1), $S_e \succ S_{e'}$. Let p_i be the largest upper interval threshold smaller than $\tilde{p}$. For each menu with objective probability $p < p_i$, by monotonicity (4) $S_e \sim O_{\tilde{p}} \succ O_p$ and by transitivity (1), $S_e \succ O_{\tilde{p}}$. For each menu with objective probability $p > p_i$, by monotonicity (4) transitivity (1), $O_p \succ S_e$. If $p = p_i$ for some menu, then the participant is not on the interior a set in the belief categorization. Thus, this participant chooses S_e in all menus with $p < p_i$ and O_p in all menus with $p > p_i$. In a participant with belief in the interval with lower bound p_i chooses S_e in the first i menus and O_p in the remaining.

Suppose $\tilde{p} < 0.5$**:** then $1-\tilde{p} > 0.5$. By symmetry (3), $S_e \sim O_{\tilde{p}}$ and $S_{e'} \sim O_{1-\tilde{p}}$. By monotonicity (4) and since $1-\tilde{p} > 0.5$, $O_{1-\tilde{p}} \succ O_{\tilde{p}}$. By transitivity (1), $S_{e'} \succ S_e$. Let p_i be the largest upper interval threshold smaller than $1-\tilde{p}$. Repeating the analysis above, this participant chooses $S_{e'}$ in all menus with $p < p_i$ and O_p in all menus with $p > p_i$. In summary a subject with belief $\tilde{p}$ in in the interval with upper bound $1-p_i$ chooses $S_{e'}$ in the first i menus and O_p in the remaining.

Suppose $\tilde{p} = 0.5$**:** By symmetry and transitivity, $S_e \sim S_{e'} \sim O_{0.5}$. If there if the smallest of the upper interval thresholds is 0.5, then the participant is not on the interior of an interval. Otherwise, $O_p \succ S_e$ and $O_p \succ S_{e'}$ for all objective menu probabilities p. Thus, the participant chooses O_p in each menu.

Thus, participants with beliefs on the interior of each interval choose a different set of lotteries.