Online Appendix: "Cognitive Ability and Perceived Disagreement in Learning"

Piotr Evdokimov and Umberto Garfagnini*

October 22, 2025

This online appendix contains more detailed information about the cognitive test used in the experiment and additional analysis of first- and second-order beliefs:

1. Section A contains the statement of the questions used in the cognitive test and an analysis of correct vs. heuristic responses compared to the version by Primi et al. (2016);
2. Section B analyzes the effect of cognitive ability on the updating behavior of first-order beliefs using Grether-style regressions;
3. Section C further investigates the type of second-order beliefs reported by subjects relative to their own first-order beliefs;
4. Section D develops and estimates a structural model of second-order beliefs to shed further light on subjects' mental models.

A Cognitive Test

*School of Economics, University of Surrey.

	Primi et al. (2016)	Our experiment
Question 1	A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? [Correct answer = 5 cents; Heuristic answer = 10 cents]	A mug and a widget cost $2.30 in total. The mug costs $2.00 more than the widget. How much does the widget cost? [Correct answer = 15 cents; Heuristic answer = 30 cents]
Question 2	If it takes 5 minutes for five machines to make five widgets, how long would it take for 100 machines to make 100 widgets? [Correct answer = 5 minutes; heuristic answer = 100 minutes]	If it takes 5 days for five bees to collect five teaspoons of honey, how long would it take for 100 bees to collect 100 teaspoons of honey? [Correct answer = 5 days; heuristic answer = 100 days]
Question 3	In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? [Correct answer = 47 days; heuristic answer = 24 days]	In a garden, there is a patch of weeds. Every day, the patch doubles in size. If it takes 24 days for the patch to cover the entire garden, how long would it take for the patch to cover half of the garden? [Correct answer = 23 days; heuristic answer = 12 days]
Question 4	If three elves can wrap three toys in an hour, how many elves are needed to wrap six toys in 2 hours? [Correct answer = 3 elves; heuristic answer = 6 elves]	If three dogs can eat three treats in a minute, how many dogs are needed to eat six treats in 2 minutes? [Correct answer = 3 dogs; heuristic answer = 6 dogs]
Question 5	Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class? [Correct answer = 29 students; heuristic answer= 30 students]	Adam was both the 15th fastest and the 15th slowest runner in a race. How many runners are there in the race? [Correct answer = 29 people; heuristic answer = 30 people]
Question 6	In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes? [Correct answer = 15 medals; heuristic answer = 20 medals]	In a family game, young family members are three times more likely to win the game than older family members. Suppose that the game has been played 60 times so far. How many of these have been won by older family members? [Correct answer = 15 times; heuristic answer = 20 times]
Question 7	Ellen and Kim are running around a track. They run equally fast but Ellen started later. When Ellen has run 5 laps, Kim has run 15 laps. When Ellen has run 30 laps, how many has Kim run? [Correct answer = 40 laps; heuristic answer = 90 times; **this question is not part of the long CRT.**]	Russell and Beth are running around a track. They run equally fast but Beth started earlier. When Beth has run 15 laps, Russell has run 5 laps. When Russell has run 10 laps, how many has Beth run? [Correct answer = 20 laps; heuristic answer = 30 times]

Table A.1: **Comparison of questions from the long version of the CRT in Primi et al. (2016) and our test.** We initially randomized the order of the questions and kept it the same for all participants. The question order is Q5, Q6, Q3, Q2, Q1, Q4, Q7. Q7 was initially proposed by Van Dooren et al. (2004).

	% Correct (Heuristic)	%Correct (Heuristic)
	Primi et al. (2016)	Our experiment
Question 1	39% (49%)	37% (41.7%)
Question 2	44% (48%)	51.6% (34.4%)
Question 3	54% (36%)	56.1% (27%)
Question 4	81% (15%)	59.5% (18.4%)
Question 5	37% (36%)	36.8% (20.6%)
Question 6	49% (37%)	35.3% (26.9%)
Question 7		43.5% (22.8%)

Table A.2: **Percentages of correct and heuristic responses in the CRT: our data (*Baseline*, *InformedBottom*, *InformedTop* and *InformedOwn*) vs. Primi et al. (2016).** The information for Question 7 in Primi et al. cannot be reported because it was not included in the final test questions and thus the proportion of correct versus heuristic answers is not mentioned in their paper.

B Cognitive Ability and First-order Belief Updating

To investigate subjects' updating rules in more detail, we follow the approach in Grether (1980). Recall from Bayes' rule and the binary signal structure that a subject's posterior odds are given by

$$\frac{\mu_t}{1-\mu_t} = \frac{\mu_{t-1}}{1-\mu_{t-1}} \underbrace{\frac{\text{Prob(Current ball|Orange urn)}}{\text{Prob(Current ball|Purple urn)}}}_{LR_t}, \tag{B.1}$$

where μ_t is the subject's first-order posterior belief in period t, μ_{t-1} is their prior belief,[1] and LR_t is the likelihood ratio following the observation of the period-t ball, with $LR_t \in \{LR^{Orange} = 2, LR^{Purple} = \frac{1}{2}\}$.[2]

A long literature has used equation (B.1) as the building block for studying updating behavior by estimating the following log-linearized empirical model (e.g., Grether, 1980; Holt and Smith, 2009; Mobius et al., 2014):

$$\ln\left(\frac{\mu_t}{1-\mu_t}\right) = \beta_0 + \beta_{Prior} \ln\left(\frac{\mu_{t-1}}{1-\mu_{t-1}}\right) + \beta_{LR}\ln(LR_t) + \epsilon. \tag{B.2}$$

As common in the literature, we recode 0 guesses as 0.01 and 1 guesses as 0.99 to ensure that equation (B.2) is well-defined.

One potential issue with the OLS estimation of Equation B.2 is that prior beliefs are formed

[1] The prior beliefs in each period are defined as the reported first-order beliefs from the previous period. In period 1, the prior beliefs are exogenously given and set at 0.5.

[2] Recall that in our experiment Prob(Orange ball|Orange urn) = 2/3 and Prob(Orange ball|Purple urn) = 1/3.

	1st-order beliefs	1st-order beliefs
	IV	IV
	(1)	(2)
β_{Prior}	0.9508	0.9217
	(0.0109)	(0.0376)
β_{LR}	0.2512	0.0720
	(0.0338)	(0.0472)
Test score		0.0127
		(0.0049)
β_{Prior} × Test score		0.0055
		(0.0066)
β_{LR} × Test score		0.0504
		(0.0142)
Constant	−0.0225	−0.0719
	(0.0068)	(0.0262)
Observations	20160	20160

Subject-clustered standard errors in parentheses

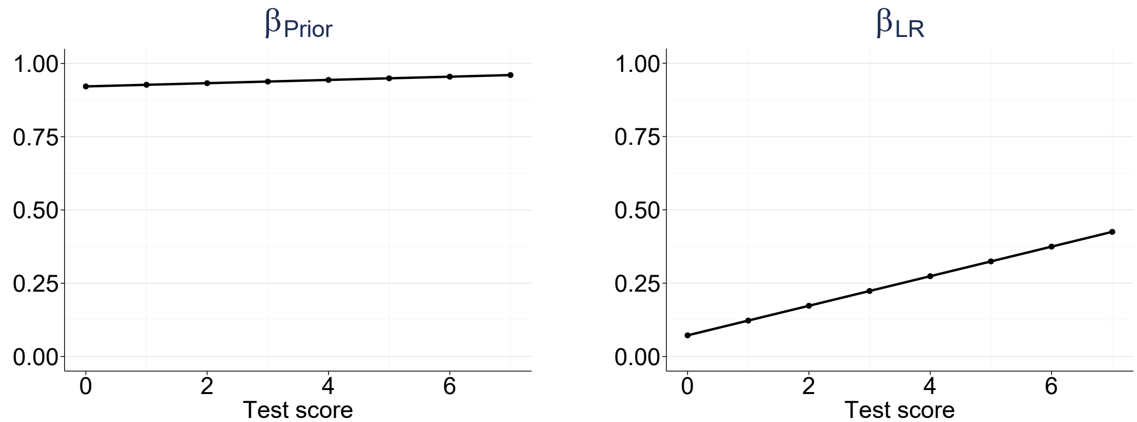
Table B.1: **IV estimation results for Equation (B.2).** The data used for the estimation involves the three main treatments: *Baseline*, *InformedBottom* and *InformedTop*.

as a one period lag of the dependent variable. This could create an endogeneity problem if the regressor is correlated with the error term. To circumvent this issue, we compute the Bayesian posterior beliefs for each observation and use $\ln\left(\frac{Bayes_{t-1}}{1-Bayes_{t-1}}\right)$ as an instrument for $\ln\left(\frac{\mu_{t-1}}{1-\mu_{t-1}}\right)$.

The IV estimates are reported in the first column of Table B.1. That both β_{Prior} and β_{LR} are significant (largest $P < 0.001$) suggests that subjects are responsive to both prior and new information. Moreover, both coefficients are significantly less than one (largest $P < 0.001$), which suggests that subjects are not as responsive as predicted by the Bayesian benchmark, where $\beta_{Prior} = \beta_{LR} = 1$. This observation conforms with well-known stylized facts about belief updating, namely base-rate neglect ($\beta_{Prior} < 1$) and underinference ($\beta_{LR} < 1$) (Benjamin, 2019, Section 4.3, Stylized Fact 9).

The second column of Table B.1 reports the estimation results of a model where the updating coefficients are interacted with cognitive ability, as measured by test score. The estimation results suggest that the degree of base-rate neglect is similar across cognitive levels ($P = 0.402$) whereas underinference decreases with cognitive ability ($P < 0.001$). Figure B.1 shows the effect of test score on both coefficients.

Observation 1. *Cognitive ability has no effect on the degree of base-rate neglect but higher test scorers exhibit lower underinference.*



(a) Prior. (b) Signal.

Figure B.1: **Average marginal effects by test score.** (a) β_{Prior} is similar across subjects with different cognitive ability. (b) β_{LR} increases with cognitive ability.

One possible interpretation for the reduced responsiveness of lower cognitive ability subjects to information in the updating task could be that those subjects paid less attention to the task at hand. Thus, their responses would be less reliable. One way to test this hypothesis is to look at subjects' response times when reporting first-order beliefs which the literature has interpreted as indicative of cognitive effort (Rubinstein, 2007, 2016; Enke et al., 2021). Contrary to this hypothesis, we find that subjects with low test scores in our three main treatments spent more time to form first-order beliefs compared to high test scorers ($P < 0.01$ in an OLS regression of the response time on test score with subject-clustered standard errors).

C Additional Analysis of Second-order Beliefs

In relation to one's own first-order beliefs, second-order beliefs could either be: *i*) more *conservative* than one's own in the sense of being closer to the uniform prior (see Edwards, 1968); *ii*) *polarized*, if second-order beliefs assign relatively more weight to the state thought to be less likely; *iii*) *overinferred*, if second-order beliefs assign relatively more weight to the state that the subject also believes to be more probable; or *iv*) simply *equal* to one's own first-order beliefs. We again exclude observations in which a subject reported first-order beliefs equal to 0.5 because second-order beliefs are hard to categorized when compared to uniform first-order beliefs. This exclusion involves 11.6% of observations from all treatments (including *InformedOwn*) and 17/918 subjects who always reported first-order beliefs equal to 0.5. However, we note that whenever a subject reports first-order beliefs equal to 0.5, their second-order beliefs are also equal to 0.5 in 64.2% of cases.

As an example, suppose that a subject holds first-order beliefs equal to 0.63 that the state is orange. Conservative second-order beliefs would belong to the interval $[0.5, 0.63)$; polarized second-order beliefs would belong to the interval $[0, 0.5)$ as the subject thinks that the matched partner believes the opposite state, i.e. purple, to be more likely; and finally, overinferred beliefs would belong to the interval $(0.63, 1]$.

	Conservative	Polarized	Overinferred	Equal
Baseline	28.9%	13.3%	14.3%	43.5%
InformedTop	25.6%	13.5%	15.1%	45.8%
InformedBottom	34.1%	17.9%	19.5%	28.5%
Overall	29.8%	14.9%	16.3%	39%
InformedOwn	35.6 %	18.2 %	28.7 %	17.5 %

Table C.1: **Distribution of second-order belief types.**

	Conservative	Polarized	Overinferred	Equal
Baseline	29.3%	10%	10.1%	50.6%
InformedTop	22.4%	7.4%	10.4%	59.8%
InformedBottom	37.1%	14.1%	12.1%	36.7%
Overall	29.4%	10.3%	10.8%	49.5%
InformedOwn	37.3%	14%	24.1%	24.6%

Table C.2: **Distribution of second-order belief types for high CRT scorers.**

	Conservative	Polarized	Overinferred	Equal
Baseline	28.4%	17.3%	19.3%	35%
InformedTop	29.5%	21%	20.8%	28.7%
InformedBottom	32.1%	20.6%	24.7%	22.6%
Overall	30.2%	19.5%	21.8%	28.5%
InformedOwn	35.2%	19.3%	29.9%	15.6%

Table C.3: **Distribution of second-order belief types for low CRT scorers.**

Table C.1 shows the distribution of second-order belief types for each treatment and overall. The modal behavior across the *Baseline*, *InformedBottom* and *InformedTop* treatments (fourth row) is to report equal beliefs but this accounts for only 39% of all relevant observations. In line with our discussion about the direction of perceived disagreement in the main text, most observations involve either reporting that the matched partner has more conservative beliefs than one's own or even polarized beliefs. Tables C.2 and C.3 further disaggregate data by cognitive type, above or below a test score of 4 (high vs. low CRT, respectively). The tables show a noticeable difference in reporting behavior with higher cognitive types reporting second-order beliefs equal to their own first-order beliefs about 50% of the time in most treatments, except for *InformedBottom*. By contrast, low cognitive types' modal behavior is to report more conservative second-order beliefs.

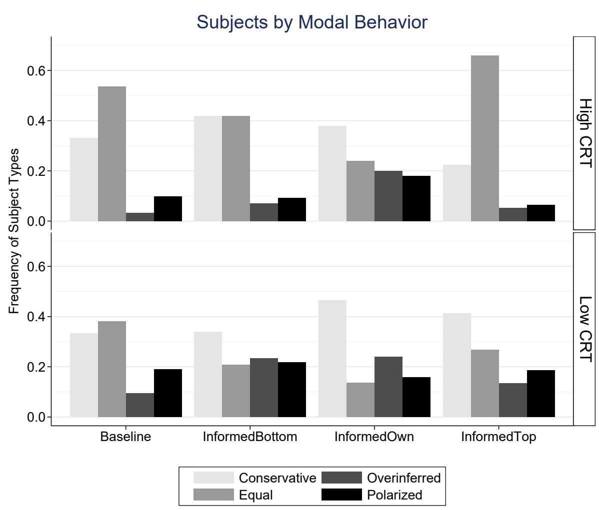
	Conservative	Overinferred	Polarized
	(1)	(2)	(3)
InformedTop	-0.1663	0.0192	-0.0221
	(0.1470)	(0.1685)	(0.2108)
InformedBottom	0.5324	0.6181	0.6117
	(0.1502)	(0.1696)	(0.1975)
InformedOwn	0.9649	1.2994	0.9334
	(0.1589)	(0.1772)	(0.2089)
Test score	-0.0974	-0.2173	-0.2042
	(0.0254)	(0.0281)	(0.0318)
Constant	-0.0181	-0.3186	-0.4297
	(0.1352)	(0.1558)	(0.1734)
Observations	24339	24339	24339

Subject-clustered standard errors in parentheses

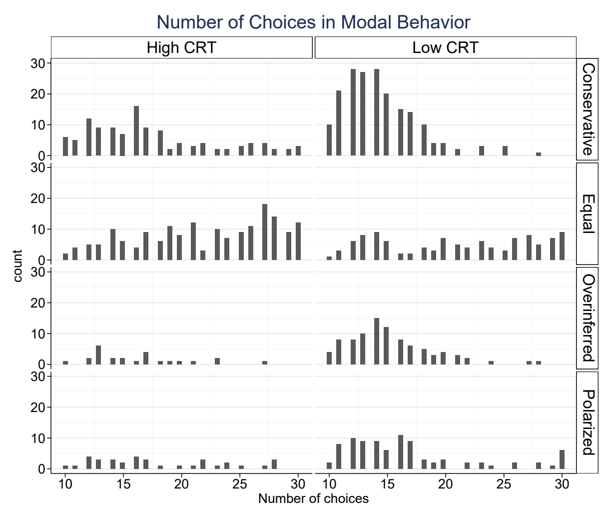
Table C.4: **Multinomial logit for the type of second-order beliefs with respect to one's own first-order beliefs.** The reference is the case of second-order beliefs equal to first-order beliefs.

To formally test for the effect of cognitive ability, Table C.4 reports the results of a multinomial logit regression for the type of second-order beliefs against treatment dummies and test score, with the case of equal first- and second-order beliefs as the reference case. The log odds of reporting conservative, polarized or overinferred beliefs versus equal beliefs all decrease with cognitive ability (largest $P < 0.001$). Furthermore, the log odds of reporting either polarized or overinferred beliefs vs equal beliefs decrease more than those of conservative beliefs as test score increases (largest $P < 0.001$ when comparing the coefficients on the test score variable across columns).

One possible interpretation of the results coming from the analysis of second-order beliefs is that the treatment effect between *InformedBottom* and the *Baseline* is simply due to subjects increasing the variance of their own choices as opposed to adjusting their mental model in a consistent way. To understand whether this is indeed the case. We can try to categorize subjects based on their modal behavior in terms of the types of second-order beliefs that they report relative to their own first-order beliefs: *conservative*, *equal*, *polarized* or *overinferred*. Suppose that we set the cutoff of at least 10 choices of a given type to categorize a subject by the type of beliefs they most frequently report. We can categorize 90.3% of all subjects (829 out of 918).[3] Figure C.1.a reports the frequencies of subject types by treatment for each cognitive type. There is a stark difference in behavior between cognitive types. Panel b) also shows how many choices are involved in the modal behavior of each subject which led to the categorization in Figure C.1.a. The figure shows that the categorization of several subjects is based on a persistent pattern of reporting second-order beliefs of a given type. So, any noise generated by lower test scorers is associated with the fact that there is a more uniform distribution of subjects across reporting types as opposed to each subject reporting more random second-order beliefs relative to their first-order beliefs.



(a) Modal types.



(b) Number of choices.

Figure C.1: **Modal types.** A subject is categorized as *Conservative*, *Equal*, *Polarized* or *Overinferred* if they report second-order beliefs of that type in at least 10 out of 30 periods. Panel b) reports the number of choices associated with modal behavior.

[3] 11% of subjects in the *Baseline* cannot be categorized compared to 12.6%, 10.1% and 5.3% in the *InformedBottom*, *InformedTop* and *InformedOwn* treatments, respectively.

We summarize the discussion so far in the following observation:

Observation 2. *Higher cognitive ability is associated with reporting less conservative, polarized or overinferred second-order beliefs relative to the equal case. The modal behavior of high cognitive ability subjects is to report second-order beliefs equal to their own first-order beliefs while less cognitively able subjects are more likely to report conservative or polarized second-order beliefs.*

Forming second-order beliefs requires an initial assumption about the behavior, or knowledge of others. In other words, a subject needs to hold a mental model about the updating rules that the matched partner might be using. Cognitive psychologists refer to the *"egocentric bias"* as the tendency for one's prediction of others' perspective to become skewed toward one's own viewpoint (see, Royzman, Cassidy and Baron, 2003; Samson et al., 2010; Frith, 2012). For example, Madarász (2012) develops a theoretical model of information projection in which agents exaggerate the extent to which their own private information is shared with others. Danz, Madarász and Wang (2018) experimentally investigate, and find support for, information projection in the laboratory.

An extension of the egocentric bias to our framework suggests that subjects could be projecting their own information processing abilities onto others. This would lead a subject to report second-order beliefs closer or even equal to one's own first-order beliefs which would account for a small or no perceived disagreement. To the extent that subjects do exhibit some form of egocentrism, Observation 2 suggests that egocentrism might be increasing with cognitive ability.

Turning to treatment effects, the multinomial logit analysis in Table C.4 shows that the log odds of reporting conservative, polarized or overinferred beliefs versus equal beliefs are all insignificantly different between the *Baseline* and *InformedTop* treatment (smallest $P = 0.258$). By contrast, all the log odds are higher in the *InformedBottom* treatment compared to the *Baseline* (largest $P < 0.01$). Thus, although the direction of perceived disagreement shows no treatment effects (third column of Table 3), the distribution of actual responses does vary between the *Baseline* and *InformedBottom* treatment.

Table C.5 performs an additional multinomial logit regression of belief types in which the treatment dummies are interacted with a dummy variable for observations associated with subjects with a low CRT score. The analysis shows that for high cognitive types a positive signal (*InformedTop*) about their partner reduces the log odds of reporting conservative versus equal beliefs compared to the *Baseline* ($P < 0.05$), whereas a negative signal (*InformedBottom*) has the opposite effect ($P < 0.05$). Also, a negative signal about the partner increases both the log odds of reporting polarized and overinferred beliefs (largest $P < 0.05$), while a positive signal reduces both log odds, neither effect is significant ($P = 0.166$ and $P = 0.592$, respectively).

	Conservative	Overinferred	Polarized
	(1)	(2)	(3)
InformedTop	-0.4362	-0.1354	-0.4706
	(0.1999)	(0.2529)	(0.3401)
InformedBottom	0.5518	0.5020	0.6584
	(0.2153)	(0.2555)	(0.3216)
InformedOwn	0.9627	1.5917	1.0590
	(0.2479)	(0.2848)	(0.3674)
Low CRT score	0.3351	1.0155	0.9117
	(0.1931)	(0.2344)	(0.2918)
Low CRT score × *InformedTop*	0.6764	0.4107	0.8668
	(0.3004)	(0.3501)	(0.4452)
Low CRT score × *InformedBottom*	0.0055	0.1807	-0.0420
	(0.2967)	(0.3450)	(0.4113)
Low CRT score × *InformedOwn*	0.0593	-0.3457	-0.1344
	(0.3230)	(0.3645)	(0.4488)
Constant	-0.5442	-1.6117	-1.6201
	(0.1322)	(0.1634)	(0.2246)
Observations	24339	24339	24339

Subject-clustered standard errors in parentheses

Table C.5: **Multinomial logit for type of second-order beliefs with respect to one's own first-order beliefs by cognitive type.** The reference is the case of second-order beliefs equal to first-order beliefs.

By contrast, for low cognitive types, a signal about the partner shifts the log odds of reporting different types of beliefs relative to equal beliefs only when the signal is negative (largest $P < 0.05$ when comparing *InformedBottom* to *Baseline*, and smallest $P = 0.168$ when comparing *InformedTop* to *Baseline*). If informed that the partner's test score is in the bottom half of the test score distribution, low cognitive types are more likely to report conservative, polarized and overinferred beliefs. Thus, information about the partner has a distinct effect on the mental model that subjects form about the updating behavior of others, which is mediated by their own cognitive ability.

Observation 3. *For high cognitive types, a positive signal about the partner leads to impute less conservative beliefs onto the partner relative to one's own beliefs, whereas a negative signal leads to imputing more conservative and polarized beliefs compared to the Baseline.*

For low cognitive types, only a negative signal matters in which case they are more likely to report conservative, polarized and overinferred versus equal beliefs.

Finally, the multinomial logit regression reported in Table C.4 also reports results for the *InformedOwn* treatment. The regression analysis confirms that information about own cognitive skills affects subjects' mental model about the belief updating behavior of the partner by increasing the likelihood of reporting conservative, overinferred and polarized versus equal beliefs (largest $P < 0.01$) compared to the *Baseline*.

D A Structural Model of Second-Order Beliefs

To shed light on the mental model behind the reported guesses, we formulate and estimate a structural model of second-order beliefs. A simple approach would involve estimating Grether-style regressions using subjects' second-order beliefs. However, this is potentially problematic as second-order beliefs are not probabilities but rather first moments of the distribution of second-order beliefs unless subjects hold a degenerate distribution over their partner's updating behavior.

Our starting point is the incentive structure generated by the binarized scoring rule (Hossain and Okui, 2013) with a quadratic loss function $L(a, \omega) = (a - \omega)^2$, where ω is the realization of the random variable of interest. A decision maker (DM, henceforth) chooses an action a to maximize their expected utility

$$\max_{a \in [0,1]} E_\eta[u(a)] \quad = \quad u(0) + \big(1 - E_\eta[(a - \mu_{-i})^2]\big)\, \Delta u, \tag{D.1}$$

where η denotes the DM's distribution of second-order beliefs, μ_{-i} are the partner's first-order beliefs, and $\Delta u \equiv u(R) - u(0)$ with $R > 0$ denoting the monetary bonus awarded by the scoring rule in our experiment. The right-hand side in equation (D.1) follows from the definition of the binarized scoring rule which makes the curvature of the DM's utility function irrelevant for the maximization problem. We can thus rewrite the DM's expected utility simply as $U(a;\eta) = -E_\eta[(a-\mu_{-i})^2]$.

The expectation in equation (D.1) is more complex than the one involving the elicitation of first-order beliefs because it involves a distribution over the beliefs of the partner after having observed the same information. Any attempt at a structural estimation needs to first model how the DM thinks about the belief updating process used by the partner after the observation of an informative signal. To formally investigate belief updating behavior, recall the discussion around equations (B.1) and (B.2). We first rewrite the flexible updating rule in (B.2) in levels (without the error term and excluding β_0 for simplicity). Then, we formulate the following model for the partner's period-t posterior belief

$$\mu_{-i,t}\left(\mu_{-i,t-1}, s_t; \beta^{-i}_{Prior}, \beta^{-i}_{LR}\right) = \frac{\mu_{-i,t-1}^{\beta^{-i}_{Prior}}}{\mu_{-i,t-1}^{\beta^{-i}_{Prior}} + (1-\mu_{-i,t-1})^{\beta^{-i}_{Prior}} LR(s_t)^{-\beta^{-i}_{LR}}}, \tag{D.2}$$

for some parameter pair $\left(\beta^{-i}_{Prior}, \beta^{-i}_{LR}\right)$. Given a pair of updating coefficients, the partner's belief can be constructed recursively starting from $\mu_0 = 0.5$ and the public signals contained in history $h_t = \{s_1, ...s_t\}$.

Formally, let $\boldsymbol{\beta}^{-i} = \left(\beta^{-i}_{Prior}, \beta^{-i}_{LR}\right)$ denote the vector of updating coefficients used by the DM's partner. We postulate that the DM holds a belief over the coefficient vector $\boldsymbol{\beta}^{-i}$ which is represented by a distribution function $G(\cdot|\mathcal{P})$ with density $g(\cdot|\mathcal{P})$, given a set of parameters $\mathcal{P}$. Under these assumptions, we can write the expected utility for an arbitrary action a at time $t \geq 1$ as

$$\begin{aligned} U(a;\mu_0,h_t,\mathcal{P}) &= -E[(a-\mu_{-i})^2|\mu_0,h_t,\mathcal{P}] \\ &= -\int_{\boldsymbol{\beta}^{-i}} \left[a - \mu_{-i,t}\left(\mu_0, h_t; \boldsymbol{\beta}^{-i}\right)\right]^2 g\left(\boldsymbol{\beta}^{-i}|\mathcal{P}\right) d\beta^{-i}_{Prior} d\beta^{-i}_{LR}. \end{aligned} \tag{D.3}$$

Specifically, we will assume that the parameters are drawn from a bivariate normal distribution with mean $\bar{\boldsymbol{\beta}}^{-i} = (\bar{\beta}^{-i}_{Prior}, \bar{\beta}^{-i}_{LR})$ and variance-covariance matrix

$$\Sigma = \begin{bmatrix} (\tau^{-i}_{Prior})^2 & \rho^{-i}\tau^{-i}_{Prior}\tau^{-i}_{LR} \\ \rho^{-i}\tau^{-i}_{Prior}\tau^{-i}_{LR} & (\tau^{-i}_{LR})^2 \end{bmatrix}. \tag{D.4}$$

The parameter vector $\mathcal{P} = \left(\bar{\beta}^{-i}_{Prior}, \bar{\beta}^{-i}_{LR}, \tau^{-i}_{Prior}, \tau^{-i}_{LR}, \rho^{-i}\right)$ controls the DM's belief distribution over the partner's updating behavior. If $\tau^{-i}_{Prior} = \tau^{-i}_{LR} = 0$, then the DM holds a degenerate belief about the partner's updating coefficients. More generally, the DM may be uncertain about the partner's updating behavior. The above specification is flexible enough to accommodate such uncertainty.

We further assume a logit choice rule to accommodate for possible choice imprecision $\lambda \geq 0$, that is, the DM chooses action a at time t with probability

$$p_t(a|\lambda, \mu_0, h_t, \mathcal{P}) = \frac{e^{\lambda U(a;\mu_0,h_t,\mathcal{P})}}{\sum_{a'} e^{\lambda U(a';\mu_0,h_t,\mathcal{P})}}. \tag{D.5}$$

This is a computationally intensive estimation, which requires the computation of the integral in (D.3). One possibility would be to use a maximum simulated likelihood approach through Monte Carlo integration. Under the bivariate normal assumption, however, a more efficient approach is to use Gauss-Hermite quadrature which is implemented using 20 nodes for each of the two dimensions.

The estimation imposes some constraints on the choice of parameters to guarantee that:

1. $\lambda \geq 0$, $\tau^{-i}_{Prior} \geq 0$ and $\tau^{-i}_{LR} \geq 0$;
2. $-1 \leq \rho^{-i} \leq 1$.

It is well-known that constrained MLE can lead to inference issues due to the possibility of binding constraints, even when performing bootstrapping (Andrews, 2000). To mitigate this problem, we apply a parameter transformation which maps the coefficients to be estimated to an unconstrained parameter space where all non-negativity and inequality constraints are handled directly by the transformation, and perform an unconstrained maximum-likelihood estimation.[4] Each estimation involves an initial stage to determine a good starting point in the unconstrained space followed by an interative search on the entire space. We employ a horse race among three different algorithm combinations run in parallel: 1) the global optimization algorithm DIRECT-L (Gablonsky and Kelley, 2001) followed by an iterative search with the popular gradient-free algorithm developed by Nelder and Mead (1965); 2) DIRECT-L followed by the gradient-based algorithm SLSQP based on a sequential quadratic programming algorithm developed by Kraft (1988); 3) the global optimization algorithm STOGO (Gudmundsson, 1998; Madsen, Zertchaninov and Zilinskas, 1998) followed by an iterative application of Nelder-Mead.

[4]Table D.4 contains details of the transformation used for each parameter.

The iterative search involves using the estimated parameters from the previous search as starting points for the next round, which leads to improved optimization outcomes. All algorithms, which are deterministic, are accessed through the NLopt library in R (Johnson et al., 2014). The estimated coefficients are subsequently mapped back into the original space. This approach resolves issues of inconsistency with the bootstrap (see Davison and Hinkley, 1997, Chapter 5).[5] The final set of estimated coefficients is selected from the combination of algorithms which achieved the lowest negative log-likelihood in each sample. All estimation procedures were conducted on the University of Surrey's high-performance computing cluster, utilizing parallel processing techniques.

For each treatment, we perform clustered bootstrapping at the subject-level with 2000 replications to compute bias-corrected percentile-based confidence intervals for each estimated parameter. Bootstrapping follows the same estimation procedure outlined above for each replication with the final bootstrapped parameters transformed back into the constrained space based on the best performing algorithm combination.[6]

	Baseline	*InformedTop*	*InformedBottom*	*InformedOwn*
	(1)	(2)	(3)	(4)
λ	1.5089	1.3845	0.7568	0.1035
	[0.9920, 2.0661]	[0.8569, 2.0574]	[0.3370, 1.2864]	[0.0000, 0.4686]
$\bar{\beta}^{-i}_{Prior}$	1.0233	1.0495	1.0124	1.2756
	[0.9632, 1.0881]	[0.9377, 1.5490]	[0.8684, 2.0346]	[-5.4828, 10.0000]
$\bar{\beta}^{-i}_{LR}$	1.0222	1.6688	0.9500	7.1576
	[0.6520, 1.9419]	[0.8309, 5.1146]	[0.2874, 9.9766]	[-9.9976, 10.0000]
τ^{-i}_{Prior}	0.0100	0.1932	0.0355	0.0100
	[0.0100, 0.0979]	[0.0100, 0.4511]	[0.0100, 1.5310]	[0.0100, 0.5576]
τ^{-i}_{LR}	0.0102	0.3825	0.0114	0.8371
	[0.0100, 0.2949]	[0.0100, 2.0959]	[0.0100, 0.7606]	[0.0100, 10.0000]
ρ^{-i}	-0.9999	-0.9999	-0.0015	0.9592
	[-0.9999, -0.9157]	[-0.9999, 0.9999]	[-0.9999, 0.9999]	[-0.9999, 0.9999]
-LL	34703.8280	25755.4799	31738.2284	34056.1351

Note: 95% bias-corrected, subject-clustered bootstrapped confidence intervals in brackets.

Table D.1: **Bootstrap results for second-order belief estimations by treatment.**

Table D.1 reports the results of the estimations for each of our main treatments as well as *InformedOwn*, while Figure D.1 shows the bootstrapped cumulative distribution functions for

[5] While the search is unconstrained, the transformation imposes upper bounds on some of the coefficients when mapping back to the original space to avoid possible parameter explosion in some of the bootstrap replications.

[6] Each bootstrap replication involves sampling clusters with replacement. By employing seeds, we made sure that the same sampled clusters where used when estimating coefficients with a different algorithm combination. This allowed us to pick the best performing algorithm for each bootstrap replication.

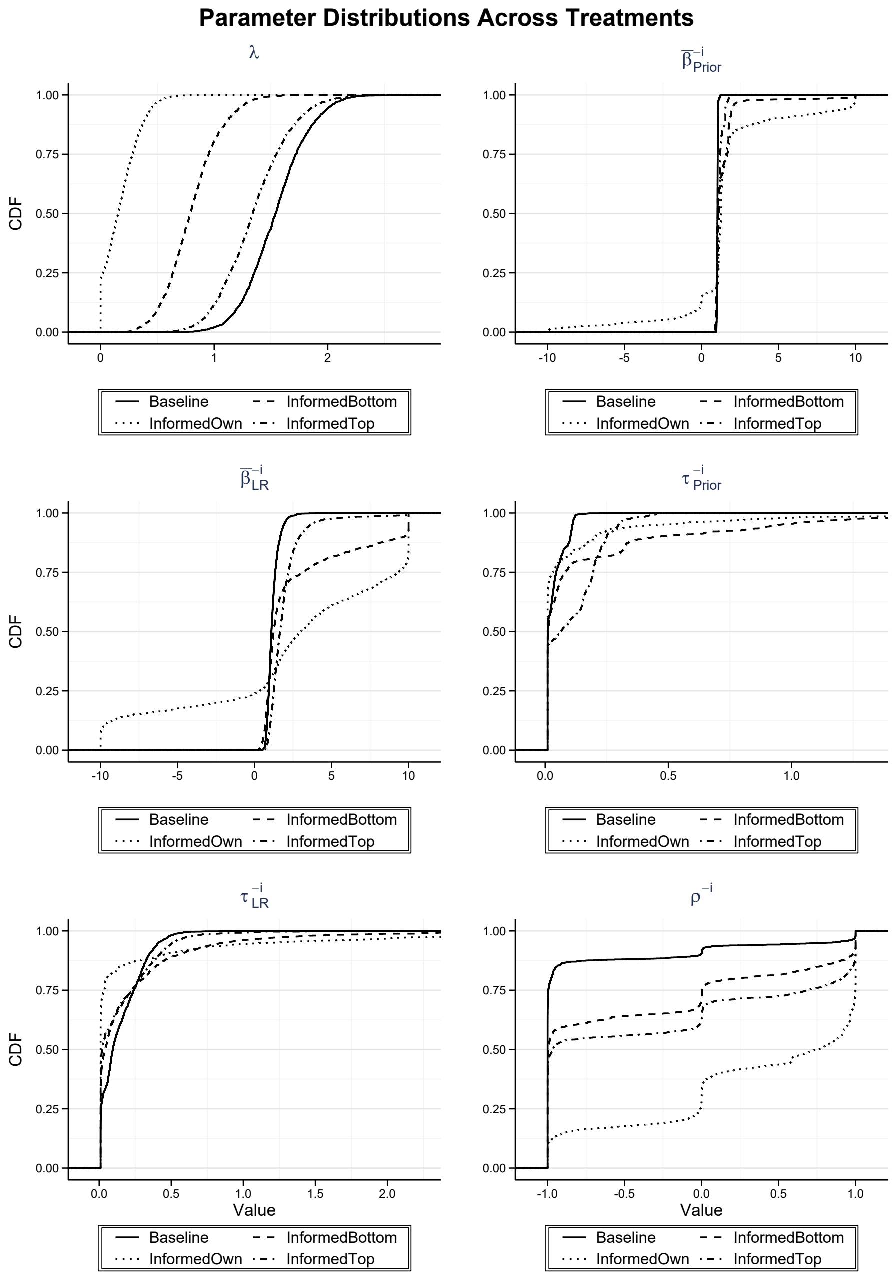


Figure D.1: **Bootstrapped distributions of estimated parameters.**

each estimated parameter. While there is variation across treatments, the estimated average updating coefficients are not quantitatively different across treatments (second and third row of Table D.1), except for *InformedOwn* which exhibits substantially greater variation in estimated parameters as shown by the confidence intervals. The greatest difference across treatments occurs in choice precision which is comparable in *Baseline* and *InformedTop* but progressively lower when moving to *InformedBottom* and, finally, *InformedOwn*.

To test for differences in choice precision between treatments, we randomly resample the precision parameter with replacement (1,000,000 replications) from the bootstrapped distribution of each treatment and compute the difference to generate a distribution which we can be used to construct a confidence interval to test for difference under an independence assumption.[7] Table D.2 reports the 95% confidence intervals for such differences. Choice precision is greater in *Baseline* than in both *InformedBottom* and *InformedOwn* with 95% confidence. By contrast, we do not find evidence that the parameters differ significantly at the 95% confidence level between *Baseline* and *InformedTop*.

Treatment comparison	Confidence interval
Baseline vs. *InformedBottom*	[0.0132, 1.4522]
Baseline vs. *InformedTop*	[-0.6272, 0.9771]
Baseline vs. *InformedOwn*	[0.7565, 1.9969]

Table D.2: **95% percentile-based confidence intervals for the difference in choice precision between treatments.**

A natural question is how these parameters compare to those used by subjects to update their own first-order beliefs. To do this, we estimate a model of first-order belief updating with choice imprecision.[8] Figure D.2 shows the bootstrapped distributions for the choice precision parameter estimated both from first- and second-order beliefs. For each treatment, none of the 95% confidence intervals computed for the choice precision parameter for first- and second-order beliefs overlap. Furthermore, each confidence interval for first-order beliefs lies to the right of the right-end point of the corresponding confidence interval for second-order beliefs. This shows a higher choice precision for first-order beliefs with 95% confidence. This could be an indication of subjects facing higher cognitive uncertainty in the formation of second-order beliefs (Enke et al., 2021).

Figures D.3 and D.4 further compare the bootstrapped distributions of updating coefficients from first-order beliefs (β_{Prior} and β_{LR}) with the average updating coefficients attributed to

[7]This is a reasonable assumption given our between-subjects design.

[8]For first-order beliefs, we specify a logit choice rule with the expected utility for an arbitrary action a given by

$$U(a;\mu_{t-1},s,\beta_{Prior},\beta_{LR}) = -\frac{\mu_{t-1}^{\beta_{Prior}}}{\mu_{t-1}^{\beta_{Prior}}+(1-\mu_0)^{\beta_{Prior}}LR(s)^{-\beta_{LR}}}(a-1)^2 - \left[1-\frac{\mu_{t-1}^{\beta_{Prior}}}{\mu_{t-1}^{\beta_{Prior}}+(1-\mu_0)^{\beta_{Prior}}LR(s)^{-\beta_{LR}}}\right]a^2 \quad \text{(D.6)}$$

where μ_{t-1} is the DM's reported first-order belief in the previous period.

one's partner ($\bar{\beta}^{-i}_{Prior}$ and $\bar{\beta}^{-i}_{LR}$). The figures seem to suggest that subjects ascribe to others an average responsiveness to information which is higher than their own. This appears at odds with the finding from the aggregate analysis that subjects report second-order beliefs that are more conservative than their own first-order beliefs. However, it is important to point out that choice precision is substantially lower for second-order beliefs. As choice precision diminishes, reported second-order beliefs get closer to the prior of 0.5.

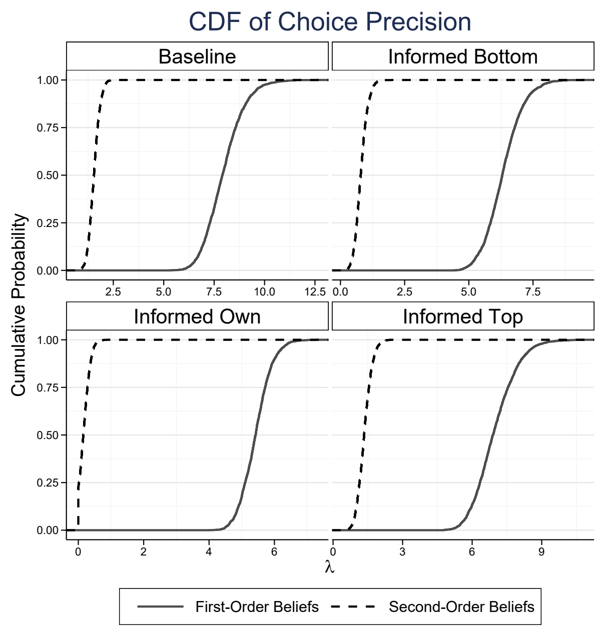


Figure D.2: **Choice precision (λ): first- vs. second-order beliefs.** Based on bootstrapped distributions.

The structural model is also flexible enough to allow for the estimation of parameters for each cognitive type. Formally, we modify the estimation to control for high test scores by specifying $\theta = \theta_0 + \theta_1 * High\_CRT$, for any $\theta \in \left\{\lambda, \bar{\beta}^{-i}_{Prior}, \bar{\beta}^{-i}_{LR}, \tau^{-i}_{Prior}, \tau^{-i}_{LR}, \rho^{-i}\right\}$, where $High\_CRT$ is a dummy variable equal to one for subjects with a test score of at least 4 out of 7.[9] Table D.3 reports the results of the estimations and Figures D.5, D.6 and D.7 show the bootstrapped distributions of the estimated coefficients, where we define $\theta_L = \theta_0$ and $\theta_H = \theta_0 + \theta_1$. Overall, the results show substantial variation in the structural parameters with confidence intervals that tend to be wider for lower cognitive types and, generally, for the *InformedOwn* treatment. However,

[9]See Table D.5 for details about the parameter transformations used in the estimation. We use a dummy variable for high test scores instead of controlling directly for test score because it simplifies the specification of the constraints on some of the structural parameters.

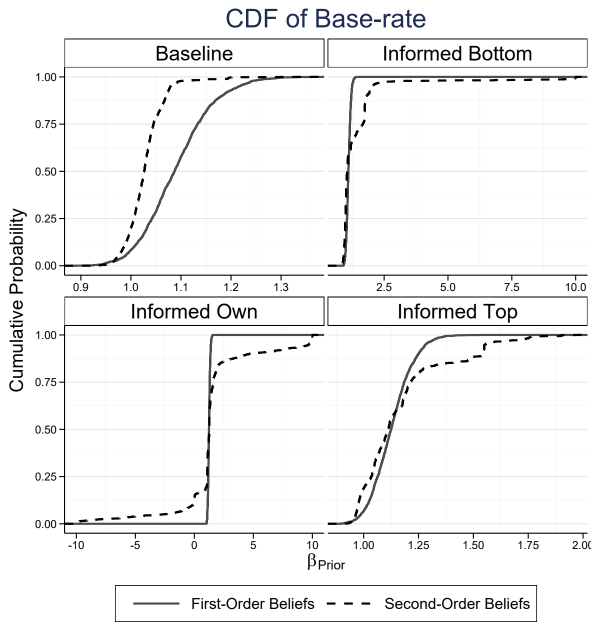


Figure D.3: **Base-rate (β_{Prior} and $\bar{\beta}^{-i}_{Prior}$): first- vs. second-order beliefs.** Based on bootstrapped distributions.

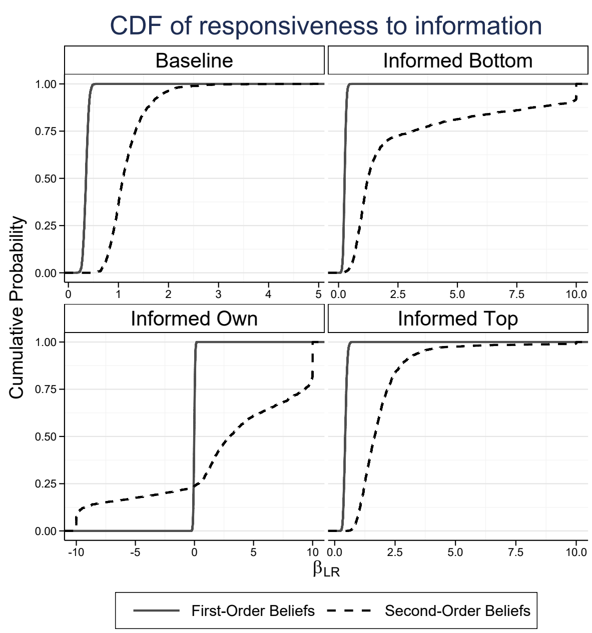


Figure D.4: **Responsiveness to information (β_{LR} and $\bar{\beta}^{-i}_{LR}$): first- vs. second-order beliefs.** Based on bootstrapped distributions.

choice precision is significantly larger for higher test scorers in all main treatments, except for *InformedOwn* ($P < 0.05$, third row of Table D.3).

	Baseline (1)	*InformedTop* (2)	*InformedBottom* (3)	*InformedOwn* (4)
λ_L	0.3787	0.5650	0.3130	0.0000
	[0.0000, 1.1611]	[0.0000, 1.2526]	[0.0000, 1.3066]	[0.0000, 0.3643]
λ_H	2.9914	2.2099	1.6302	0.8872
	[2.0667, 4.2746]	[1.3326, 3.4576]	[0.7840, 2.7481]	[0.0338, 2.2544]
λ: *Difference*	2.6127	1.6448	1.3172	0.8872
	[1.2850, 4.0142]	[0.4664, 3.0791]	[0.0554, 2.5533]	[-0.1172, 2.4141]
$\bar{\beta}^{-i}_{Prior,L}$	1.0087	1.2273	1.5069	0.8688
	[-0.2896, 2.1194]	[-0.5632, 2.9423]	[0.0068, 3.5912]	[-4.9685, 8.9873]
$\bar{\beta}^{-i}_{Prior,H}$	1.0343	1.0686	0.9921	1.7457
	[0.9667, 1.1454]	[0.9757, 1.2582]	[0.8907, 1.7359]	[-0.5453, 20.0000]
$\bar{\beta}^{-i}_{Prior}$: *Difference*	0.0256	-0.1587	-0.5148	0.8769
	[-1.1536, 1.1955]	[-1.8411, 1.6594]	[-2.3786, 1.3545]	[-2.0637, 10.0000]
$\bar{\beta}^{-i}_{LR,L}$	1.1243	2.6741	4.0648	1.6581
	[-2.1536, 6.9278]	[0.2278, 8.6685]	[-0.4278, 10.0000]	[-7.8516, 10.0000]
$\bar{\beta}^{-i}_{LR,H}$	1.1737	1.1160	0.9737	1.9677
	[0.6800, 2.2035]	[0.6306, 2.1078]	[0.3150, 8.1349]	[-10.0646, 20.0000]
$\bar{\beta}^{-i}_{LR}$: *Difference*	0.0494	-1.5581	-3.0911	0.3096
	[-5.8457, 3.0802]	[-7.5115, 0.9744]	[-9.1371, 2.2082]	[-10.0000, 10.0000]
$\tau^{-i}_{Prior,L}$	0.0100	0.3046	1.2213	0.1452
	[0.0100, 0.0100]	[0.0100, 8.0236]	[0.0100, 10.0000]	[0.0100, 10.0000]
$\tau^{-i}_{Prior,H}$	0.0875	0.1040	0.0106	0.6878
	[0.0100, 0.1625]	[0.0100, 0.2742]	[0.0100, 0.6481]	[0.0100, 9.9885]
τ^{-i}_{Prior}: *Difference*	0.0775	-0.2007	-1.2107	0.5426
	[-1.0199, 0.1815]	[-7.1938, 0.1535]	[-9.9900, 0.0674]	[-0.1014, 9.9842]
$\tau^{-i}_{LR,L}$	0.0832	0.0760	0.0558	2.2419
	[0.0100, 9.8670]	[0.0100, 9.9600]	[0.0100, 9.9999]	[0.0100, 10.0000]
$\tau^{-i}_{LR,H}$	0.2910	0.2751	0.0214	0.7861
	[0.0100, 0.6363]	[0.0100, 0.8261]	[0.0100, 1.0679]	[0.0100, 10.0000]
τ^{-i}_{LR}: *Difference*	0.2078	0.1991	-0.0344	-1.4558
	[-2.6777, 0.6549]	[-0.6080, 0.8300]	[-9.9898, 0.6668]	[-9.9900, 0.1445]
ρ^{-i}_L	-0.9999	-0.9999	0.6871	0.9999
	[-0.9999, 0.9999]	[-0.9999, 0.9851]	[-0.9999, 0.9999]	[-0.9583, 0.9999]
ρ^{-i}_H	-0.9999	-0.9999	-0.8350	0.9999
	[-0.9999, -0.9999]	[-0.9999, 0.9999]	[-0.9999, 0.9999]	[-0.9903, 0.9999]
ρ^{-i}: *Difference*	-0.0000	0.0000	-1.5221	-0.0000
	[-1.9998, 1.8924]	[-1.9998, 1.9998]	[-1.9998, 1.7276]	[-1.9998, 1.9998]
-LL	34426.4953	25653.3483	31658.2637	34027.5576

Note: 95% bias-corrected, subject-clustered bootstrapped confidence intervals in brackets.

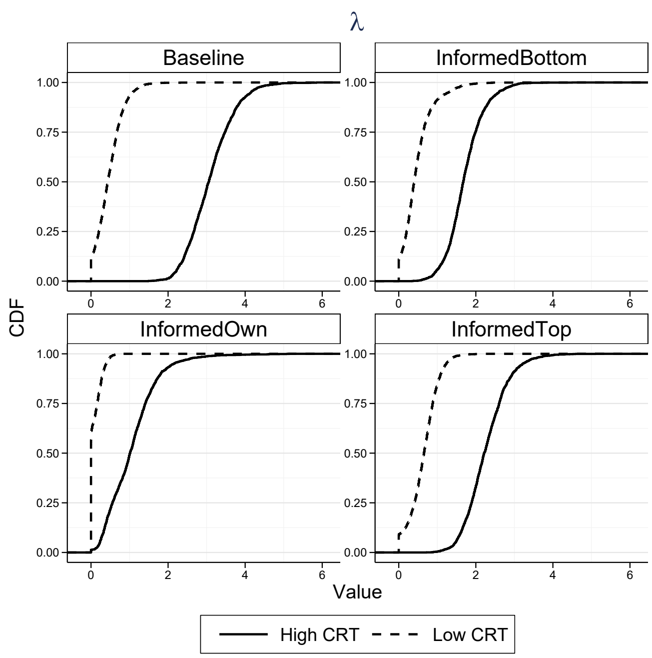
Table D.3: **Bootstrap results for second-order belief estimations with cognitive types.**

Parameter	**Transformation**	**Details**
λ	Positive (softplus)	$\lambda = \log(1 + e^{\phi})$
$\bar{\beta}^{-i}_{Prior}$	Bounded (scaled erf)	$\bar{\beta}^{-i}_{Prior} = s_{\max} \cdot \text{erf}\left(\frac{\phi_1}{\sqrt{2}}\right)$
$\bar{\beta}^{-i}_{LR}$	Bounded (scaled erf)	$\bar{\beta}^{-i}_{LR} = s_{\max} \cdot \text{erf}\left(\frac{\phi_2}{\sqrt{2}}\right)$
τ^{-i}_{Prior}	Positive (log-scaled sigmoid)	$\tau^{-i}_{Prior} = \exp\left[\log(s_{\min}) + (\log(s_{\max}) - \log(s_{\min})) \cdot \sigma(\phi_3)\right]$
τ^{-i}_{LR}	Positive (log-scaled sigmoid)	$\tau^{-i}_{LR} = \exp\left(\log(s_{\min}) + (\log(s_{\max}) - \log(s_{\min})) \cdot \sigma(\phi_4)\right)$
ρ^{-i}	Correlation (scaled tanh)	$\rho^{-i} = \rho_{\text{scale}} \cdot \tanh(\phi_5)$

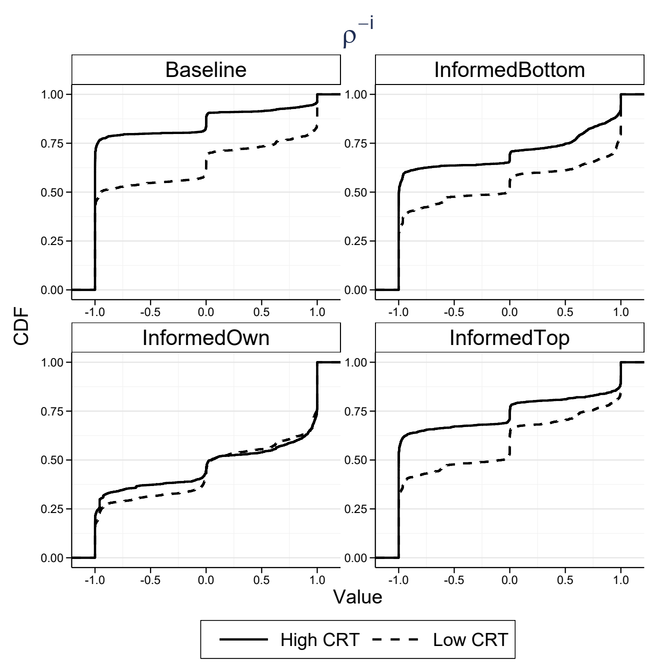
Table D.4: Smooth reparameterization from unconstrained parameters $\boldsymbol{\phi} \in \mathbb{R}^6$ to constrained parameters $\theta = \left(\lambda, \bar{\beta}^{-i}_{Prior}, \bar{\beta}^{-i}_{LR}, \tau^{-i}_{Prior}, \tau^{-i}_{LR}, \rho^{-i}\right)$. The scales are set to $s_{\min} = 0.01$, $s_{\max} = 10.0$, and $\rho_{\text{scale}} = 0.9999$. $\sigma(x)$ denotes the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$.

Parameter	**Transformation**	**Details**
λ_0	Positive (softplus)	$\lambda_0 = \log(1 + e^{\phi_0})$
λ_1	Sum-constrained	$\lambda_1 = \log(1 + e^{\phi_1}) - \lambda_0$
$\bar{\beta}^{-i}_{Prior,0}$	Bounded (scaled erf)	$\bar{\beta}^{-i}_{Prior,0} = s_{\max} \cdot \text{erf}\left(\frac{\phi_2}{\sqrt{2}}\right)$
$\bar{\beta}^{-i}_{Prior,1}$	Bounded (scaled erf)	$\bar{\beta}^{-i}_{Prior,1} = s_{\max} \cdot \text{erf}\left(\frac{\phi_3}{\sqrt{2}}\right)$
$\bar{\beta}^{-i}_{LR,0}$	Bounded (scaled erf)	$\bar{\beta}^{-i}_{LR,0} = s_{\max} \cdot \text{erf}\left(\frac{\phi_4}{\sqrt{2}}\right)$
$\bar{\beta}^{-i}_{LR,1}$	Bounded (scaled erf)	$\bar{\beta}^{-i}_{LR,1} = s_{\max} \cdot \text{erf}\left(\frac{\phi_5}{\sqrt{2}}\right)$
$\tau^{-i}_{Prior,0}$	Positive (log-scaled sigmoid)	$\tau^{-i}_{Prior,0} = \exp\left[\log(s_{\min}) + (\log(s_{\max}) - \log(s_{\min})) \cdot \sigma(\phi_6)\right]$
$\tau^{-i}_{Prior,1}$	Sum-constrained	$\tau^{-i}_{Prior,1} = \exp[\log(s_{\min}) + (\log(s_{\max}) - \log(s_{\min})) \cdot \sigma(\phi_7)] - \tau^{-i}_{Prior,0}$
$\tau^{-i}_{LR,0}$	Positive (log-scaled sigmoid)	$\tau^{-i}_{LR,0} = \exp\left(\log(s_{\min}) + (\log(s_{\max}) - \log(s_{\min})) \cdot \sigma(\phi_8)\right)$
$\tau^{-i}_{LR,1}$	Sum-constrained	$\tau^{-i}_{LR,1} = \exp[\log(s_{\min}) + (\log(s_{\max}) - \log(s_{\min})) \cdot \sigma(\phi_9)] - \tau^{-i}_{LR,0}$
ρ^{-i}_0	Correlation (scaled tanh)	$\rho^{-i}_0 = \rho_{\text{scale}} \cdot \tanh(\phi_{10})$
ρ^{-i}_1	Sum-constrained correlation	$\rho^{-i}_1 = \rho_{\text{scale}} \cdot \tanh(\phi_{11}) - \rho_0$

Table D.5: Smooth reparameterization with cognitive types from unconstrained parameters $\boldsymbol{\phi} \in \mathbb{R}^{12}$ to constrained parameters $\theta = \left(\lambda_0, \lambda_1, \bar{\beta}^{-i}_{Prior,0}, \bar{\beta}^{-i}_{Prior,1}, \bar{\beta}^{-i}_{LR,0}, \bar{\beta}^{-i}_{LR,1}, \tau^{-i}_{Prior,0}, \tau^{-i}_{Prior,1}, \tau^{-i}_{LR,0}, \tau^{-i}_{LR,1}, \rho^{-i}_0, \rho^{-i}_1\right)$. The scales are set to $s_{\min} = 0.01$, $s_{\max} = 10.0$, and $\rho_{\text{scale}} = 0.9999$. $\sigma(x)$ denotes the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$.

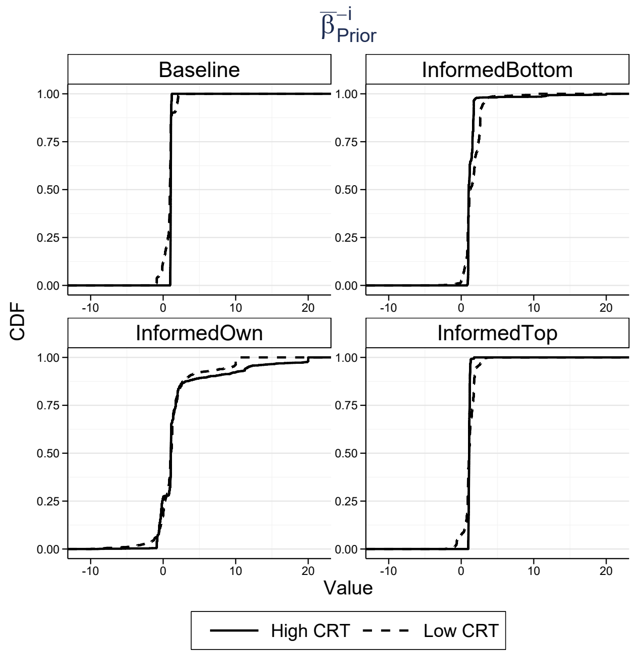


(a) λ.

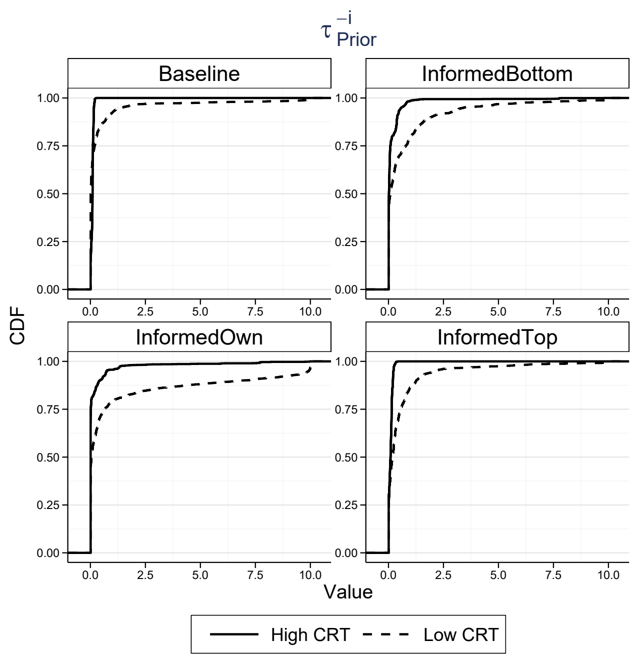


(b) ρ^{-i}.

Figure D.5: **Distributions of estimated parameters for second-order beliefs: λ, ρ^{-i}.** Based on bootstrapped distributions.

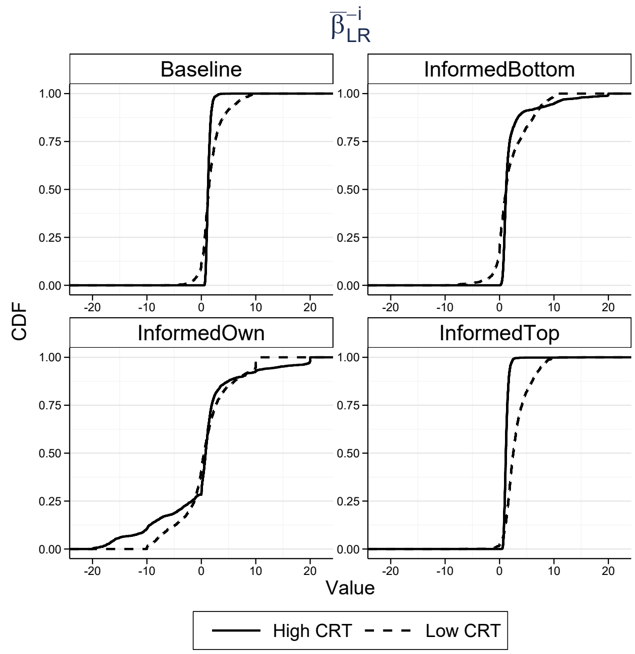


(a) $\bar{\beta}^{-i}_{Prior}$.

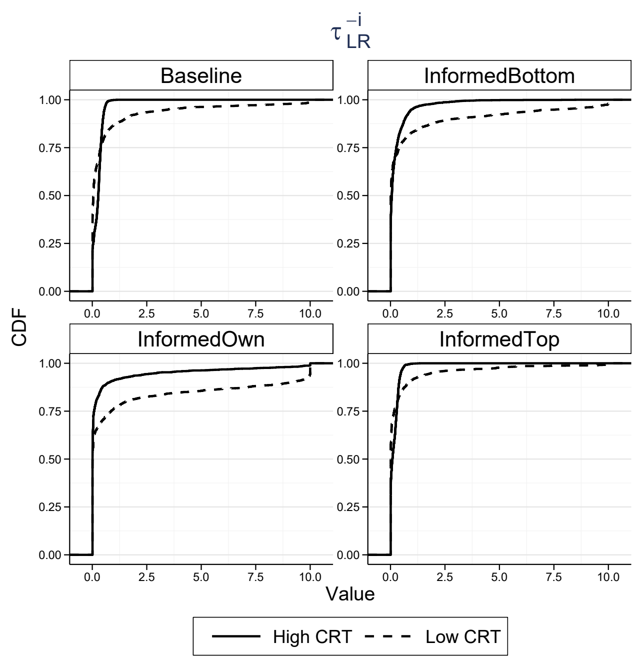


(b) τ^{-i}_{Prior}.

Figure D.6: **Distributions of estimated parameters for second-order beliefs: $\bar{\beta}^{-i}_{Prior}$, τ^{-i}_{Prior}.** Based on bootstrapped distributions.



(a) $\bar{\beta}_{LR}^{-i}$.



(b) τ_{LR}^{-i}.

Figure D.7: **Distributions of estimated parameters for second-order beliefs: $\bar{\beta}_{LR}^{-i}$, τ_{LR}^{-i}.** Based on bootstrapped distributions.

References

Andrews, Donald WK. 2000. "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space." *Econometrica*, 399–405.

Benjamin, Daniel J. 2019. "Errors in probabilistic reasoning and judgment biases." *Handbook of Behavioral Economics: Applications and Foundations 1*, 2: 69–186.

Danz, David, Kristóf Madarász, and Stephanie Wang. 2018. "The biases of others: Projection equilibrium in an agency setting." unpublished.

Davison, Anthony Christopher, and David Victor Hinkley. 1997. *Bootstrap methods and their application.* Cambridge university press.

Edwards, Ward. 1968. "Conservatism in human information processing." In *Formal Representation of Human Judgement.* , ed. B Kleinmuntz, 17–52. New York.

Enke, Benjamin, Uri Gneezy, Brian Hall, David Martin, Vadim Nelidov, Theo Offerman, and Jeroen Van De Ven. 2021. "Cognitive biases: Mistakes or missing stakes?" *Review of Economics and Statistics*, 1–45.

Frith, Chris D. 2012. "The role of metacognition in human social interactions." *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599): 2213–2223.

Gablonsky, Joerg M, and Carl T Kelley. 2001. "A locally-biased form of the DIRECT algorithm." *Journal of Global optimization*, 21: 27–37.

Grether, David M. 1980. "Bayes rule as a descriptive model: The representativeness heuristic." *The Quarterly Journal of Economics*, 95: 537–557.

Gudmundsson, Steinn. 1998. *Parallel global optimization.* Institut for Matematisk Modellering, Danmarks Tekniske Universitet.

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

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

Johnson, Steven G, et al. 2014. "The NLopt nonlinear-optimization package." *URL http://ab-initio. mit. edu/nlopt.*

Kraft, Dieter. 1988. "A software package for sequential quadratic programming." *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt.*

Madarász, Kristóf. 2012. "Information projection: Model and applications." *The Review of Economic Studies*, 79: 961–985.

Madsen, Kaj, Serguei Zertchaninov, and Antanas Zilinskas. 1998. "Global Optimization using Branch-and-Bound."

Mobius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat. 2014. "Managing self-confidence: Theory and experimental evidence." National Bureau of Economic Research.

Nelder, John A, and Roger Mead. 1965. "A simplex method for function minimization." *The computer journal*, 7(4): 308–313.

Primi, Caterina, Kinga Morsanyi, Francesca Chiesi, Maria Anna Donati, and Jayne Hamilton. 2016. "The development and testing of a new version of the cognitive reflection test applying item response theory (IRT)." *Journal of Behavioral Decision Making*, 29: 453–469.

Royzman, Edward B, Kimberly Wright Cassidy, and Jonathan Baron. 2003. ""I know, you know": Epistemic egocentrism in children and adults." *Review of General Psychology*, 7(1): 38–65.

Rubinstein, Ariel. 2007. "Instinctive and cognitive reasoning: A study of response times." *The Economic Journal*, 117: 1243–1259.

Rubinstein, Ariel. 2016. "A typology of players: Between instinctive and contemplative." *The Quarterly Journal of Economics*, 131(2): 859–890.

Samson, Dana, Ian A Apperly, Jason J Braithwaite, Benjamin J Andrews, and Sarah E Bodley Scott. 2010. "Seeing it their way: evidence for rapid and involuntary computation of what other people see." *Journal of Experimental Psychology: Human Perception and Performance*, 36: 1255.

Van Dooren, Wim, Dirk De Bock, An Hessels, Dirk Janssens, and Lieven Verschaffel. 2004. "Remedying secondary school students' illusion of linearity: A teaching experiment aiming at conceptual change." *Learning and instruction*, 14(5): 485–501.