Cognitive Ability and Perceived Disagreement in Learning

Piotr Evdokimov* and Umberto Garfagnini†

February 1, 2023

Abstract

Do agents believe to be agreeing more with others in the long-run? This paper designs an experiment to study how cognitive abilities affect actual and perceived disagreement in a standard sequential belief updating task with public signals. We document a persistent gap in the perception of disagreement as a function of cognitive ability. Higher cognitive ability is associated with less perceived disagreement, although the average subject underestimates the extent of actual disagreement regardless of cognitive ability. Learning about the state of the world has little effect on the evolution of perceived disagreement when controlling for cognitive ability. Providing subjects with information about their partner's cognitive ability affects perceived disagreement only when the partner is less cognitively able.

JEL Classification: C90, D83, D89

Keywords: Cognitive ability, disagreement, learning

*Humboldt University, Berlin.

†School of Economics, University of Surrey.

‡Financial support from the University of Surrey is acknowledged. The first round of data collection took place in 2019 and was subject to the SAFE (Self Assessment for Ethics) procedure at the University of Surrey which did not require the need for an ethics application. The second round of data collection was subject to an ethics and governance review and received a favorable ethical opinion (FASS 067 21-22 EGA - FEO and FASS 067 21-22 EGA - Amend 1 FEO). Piotr Evdokimov gratefully acknowledges funding through CRC (Project A01).

"Biased assumptions about what others know lead in turn to erroneous expectations about how someone will behave or feel, and can thereby permeate our moral judgments, our assessments of other's behavior, our feelings and attitudes towards others, and virtually all social interaction." (Birch and Bloom, 2004)

1 Introduction

When two Bayesian agents share a common prior and observe the same sequential information about a state of the world, no disagreement between their first-order beliefs can ever arise. If, in addition, there is mutual knowledge of Bayesianism, then *perceived* disagreement, i.e. the difference between one's own first-order beliefs and beliefs about the partner's first-order beliefs (i.e., second-order beliefs), would also be zero after every period. Thus, more information about the state has no effect on disagreement, whether actual or perceived.

In practice, disagreement could arise because either the two agents update differently, in spite of observing the same information, or they believe that their partner's updating behavior differs from their own, or both. Failure to apply Bayes' rule is a well-known issue in the economics and psychology literature (e.g., Edwards, 1968; Grether, 1980). As recently reviewed by Benjamin (2019), the literature points to a number of information processing biases in how experimental subjects update their first-order beliefs.[1] The literature also points to heterogeneity in information processing (e.g., Holt and Smith, 2009). While heterogeneity in belief updating rules can easily lead to actual disagreement, agents might perceive lower, higher or even no disagreement depending on their mental model about their partner's updating rule. In the extreme cases, subjects could perceive disagreement when there is none or perceive no disagreement when actual disagreement is positive. In these cases, more information about the state of the world could affect the dynamics of disagreement even when agents believe their partner's updating rule differs from their own. This would be the case if they expected their partner to update in their same direction over time, albeit, at a possibly different rate.

We exploit variation in cognitive skills among agents to investigate the evolution of actual and perceived disagreement in a sequential learning task. Cognitive abilities are known to generate heterogeneity in belief updating behavior (more on this below). Less known is the relationship between cognitive abilities and mental models about the updating behavior of others, which is the focus of this paper.

[1]For instance, many subjects suffer from base-rate neglect, where prior information about the state of the world is to some extent ignored, and underinference, where new signals about the state of the world (e.g., draws from an urn) are not given as much weight as they should receive.

To illustrate why perceived disagreement might be economically relevant, let us expand the introductory example. Suppose that our two agents must decide whether or not to undertake an investment opportunity which is efficient if and only if the fundamentals of the economy, i.e. the common state of the world, are good and both agents invest. After observing the same information about the fundamentals, each agent's decision to invest relies on believing with high enough probability that (1) the fundamentals are good, (2) the other agent believes with high enough probability that the fundamentals are good, and so on. Even if both agents process information the same way, beliefs about the updating rule used by the other agent will determine whether an agent invests or not. Furthermore, erroneous beliefs about updating rules of others can result in miscoordination despite the fact that the same information was observed, both agents process information in a Bayesian fashion, *and* coordination is efficient. Thus, an understanding of the link between cognitive abilities and mental models about the updating behavior of others can shed light on possible drivers of coordination failure and provide insights for policy design.

Our experimental design is as follows. First, we measure cognitive ability using the Cognitive Reflection Test, or CRT (Frederick, 2005), which scores a person's success in resisting an intuitive response that produces an incorrect but effortless answer in order to derive the more effortful correct answer.[2] Then, subjects play a simple dynamic betting game in groups of two. Before any decision is made, an unknown state of the world is drawn at random and held fixed for the duration of the game. It is common knowledge that the state is the same for both subjects in the group. As the game proceeds, the group members receive public signals about this state over 30 consecutive periods. After observing each signal, each subject is incentivized to report their beliefs about the realized state, as well as–on the next decision screen–their beliefs about the beliefs of their partner. The experiment has three treatments. In the **Baseline**, subjects receive no information about their own or their partner's result on the test until the end of the experiment. In the informed treatments, before subjects make any decisions in the learning task, we inform each subject about whether the matched partner's test score is above or below the median test score of a subset of participants who previously completed the test. We refer to the former case as the **Informed-Top** treatment and the latter case as the **Informed-Bottom** treatment.

We interpret the absolute value of the difference between a subject's elicited first- and second-order beliefs as perceived disagreement. If this variable takes on a positive value, subjects

[2]While Frederick (2005) conceptualizes the CRT as measuring the capacity for mental reflection (i.e., overriding an intuitive response), others have argued that it provides a more general measure of cognitive ability (Blacksmith et al., 2019).

anticipate disagreement between theirs and their matched partner's beliefs about the state. If it is small, the subject anticipates less disagreement, either because the subject believes that the partner processes information similarly or for other reasons. Our first question is somewhat exploratory:

QUESTION 1. *How does one's own cognitive ability affect perceived disagreement?*

Previous studies have found that higher CRT scores are associated with more Bayesian posterior beliefs (Oechssler et al., 2009; Hoppe and Kusterer, 2011; Charness et al., 2012).[3] To the extent that more cognitively able subjects are able to anticipate that the average subject is less Bayesian than they themselves are, we expect such subjects to report a positive perceived disagreement.[4] Whether perceived disagreement will be smaller or greater for subjects with low cognitive skills is difficult to predict: on the one hand, it is possible that low ability subjects will update close to the prior, as shown by previous studies, and assume that others do as well. Such subjects might also reason hierarchically and act as if others are less Bayesian than they are. We therefore make no prediction about the effect of cognitive ability on the size of perceived disagreement.

Our second research question has to do with treatment effects:

QUESTION 2. *How is perceived disagreement affected by information about the cognitive ability of one's partner?*

Prior experiments have shown that subjects' decisions in games respond to information about the cognitive ability of their partners. For instance, Agranov et al. (2012) find that when informed that they are playing against a level-zero player (i.e., a computer player that chooses randomly),

[3] In Oechssler et al. (2009), subjects were asked the following question:

> Imagine there are two urns—urn A and urn B. Urn A contains 3 blue balls and 7 red balls. Urn B contains 7 blue balls and 3 red balls. Balls are now randomly drawn from one of these urns where the drawn ball is always placed back into the same urn. Twelve such random draws yielded 8 red balls and 4 blue balls. What do you estimate the probability that the balls were drawn from urn A?

The correct posterior belief is 0.97; Oechssler et al. (2009) find that subjects with high CRT scores report a belief of 0.69 on average, while those with low CRT scores report a belief of 0.59. Similar findings are reported in Hoppe and Kusterer (2011). Charness et al. (2012) measure cognitive ability using Raven's progressive matrices, and find that subjects with greater cognitive ability make less updating mistakes. Raven's is thought to be a test of fluid intelligence (IQ), which is highly correlated with performance on the CRT (Primi et al., 2016).

[4] The common view of sophisticated behavior in economics is that an agent is able to best respond to the distribution of other agents' actions (Costa-Gomes et al., 2001; Costa-Gomes and Crawford, 2006). Burks et al. (2009) show a positive link between higher cognitive skills and the ability to forecast the behavior of others. Similarly, Brañas-Garza et al. (2012) find that the Cognitive Reflection Test (CRT) (Frederick, 2005) can predict behavior in experimental p-beauty contests. Carpenter et al. (2013) also report evidence of a strong relationship between cognitive ability and sophistication in the lab.

subjects decrease their level-k in beauty contest games. Likewise, they increase their strategic sophistication when playing against graduate students in economics, whose supposed ability is higher. We look for similar effects on the size of perceived disagreement. We predict that perceived disagreement will increase in the Informed-Bottom treatment, which sends a negative signal about the partner, and decrease in the Informed-Top treatment, which sends the opposite signal.

The effect of the Informed-Top treatment hinges on perceived disagreement being present in absence of information about the partner's cognitive ability. If a subject believes that the partner uses a similar updating rule to her own and thus reports identical first- and second-order beliefs, the treatment will have no effect. If a subject shades her second-order beliefs toward the prior in absence of information about the partner's cognitive skills, a signal that the partner is a high ability type might reduce the extent of perceived disagreement. We highlight this prediction below:

PREDICTION. *Relative to the baseline treatment, perceived disagreement decreases in the Informed-Top treatment and increases in the Informed-Bottom treatment.*

Our results are as follows. In line with previous literature, high cognitive ability subjects hold more accurate first-order beliefs. Thus, actual disagreement increases as the cognitive gap between subjects widens. In regard to perceived disagreement, though, we find that subjects with higher cognitive ability report lower perceived disagreement between their first-order beliefs and those of their matched partner compared to less cognitively able subjects. Overall, the average subject assumes that the matched partner assigns a lower probability to the state thought to be more likely. For example, a subject that assigns probability greater than 0.5 (resp., less than 0.5) to one of the two states tends to report second-order beliefs that are lower (resp., higher) than their own. Surprisingly, more cognitively able subjects are less inclined to report a perceived disagreement even when they hold no information about the cognitive ability of their partner.

Our Prediction is also partially in line with the data. Controlling for cognitive ability, subjects in the Informed-Bottom treatment report higher perceived disagreement relative to the baseline. This is consistent with subjects believing that a partner who performed poorly in the test is also a worse updater which justifies anticipating a wider disagreement between their beliefs. By contrast, we find no effect of the Informed-Top treatment. One possible interpretation for the lack of a treatment effect in the latter case is that subjects might generally assume that their cognitive skills are similar to their partners. Together with the documented tendency for overconfidence in one's own skills (Moore and Healy, 2008), this could explain why perceived disagreement does not vary between the baseline and the Informed-Top treatment.

Our sequential design further allows to investigate the evolution of actual and perceived disagreement over time. Intuitively, one would expect disagreement to decrease as more informative public signals about the same state of the world are observed. This would still be reasonable even in the presence of heterogeneity in belief updating behavior as long as subjects believe their partners to respond at least in part to new information. In previous work, Evdokimov and Garfagnini (2021) investigate the evolution of higher-order beliefs and find a persistent gap between first- and second-order beliefs in a between-subjects design. However, it is again reasonable to suspect that cognitive ability might play a role in the evolution of disagreement.

We find that subjects are less likely to perceive disagreement over time in all but the Informed-Bottom treatment, where it remains persistently higher than in all other treatments. This would be the case, for instance, if subjects thought a partner who performs poorly in the test is so bad at processing information that even a large number of signals is unlikely to move the partner's beliefs closer to one's own. Being informed that your partner is in the top half of the distribution has instead no long-term effects on perceived disagreement relative to the baseline. This is still consistent with our previous explanation as subjects receive no additional feedback about their partner's cognitive ability during the experiment. Furthermore, learning also does not help to close the gap between actual and perceived disagreement. If anything, the gap widens over time in the Informed-Top treatment.

Our results provide insights into the design of policies aimed at reducing coordination failures in markets. For example, we document a strong persistence in the perception of disagreement which is unaffected by the amount of information observed. Thus, policies that target more public information disclosure may have limited effects if agents believe that other market participants are unable to correctly process the available information, whether or not this is actually the case. This tendency is less pronounced for more cognitively able agents but heterogeneity in information processing means that these agents will underestimate the extent of true disagreement even more because of their stronger egocentric tendencies. By constrast, policies aimed at mitigating mistakes that agents make in processing information, as well as theory-of-mind, can lead to improved coordination outcomes. For example, Fe et al. (2022) find that theory-of-mind develops at a young age in children and can be improved through increases in school spending.

This paper contributes to the literature on belief disagreement. Acemoglu et al. (2006, 2016) show theoretically that actual disagreement may arise and even persist over time no matter how much public information agents observe provided that they hold different prior beliefs and face subjective uncertainty about the informativeness of signals. Baliga et al. (2013) further argue that belief polarization following the observation of public information cannot arise if agents share the same beliefs about the informativeness of signals, even if they hold different priors about the state of the world. Cripps et al. (2008) further show that an unknown state of the world will become approximate common knowledge when agents observe sequential public, or even private iid, signals. Thus, they will commonly learn the true value of the state.

Finally, our paper contributes to the experimental literature on belief disagreement. Andreoni and Mylovanov (2012) show theoretically and experimentally that actual disagreement can persist even when subjects observe sufficient information about the state of the world. Crucial for their model and experiment though is that different subjects observe a private signal about the state of the world *before* observing any public information. The private signal generates heterogeneity in initial opinions about the state. By contrast, our subjects only observe public information. Nonetheless, we show that disagreement can persist over time and sometimes get worse with additional public information even when subjects share a common prior due to the heterogeneous way in which information is processed. Our results further investigate the link between cognitive abilities and perceived disagreement. Future research could explore how these results relate to coordination in games and generalize to other settings.

2 Experimental Design

The experimental design consists of two parts. In the first part, each participant individually completed an extended version of the Cognitive Reflection Test (CRT from now on, Frederick, 2005). While Frederick (2005) argues that the CRT measures the capacity for mental reflection (i.e., overriding an intuitive response) specifically, others have argued that it provides a more general measure of cognitive ability (Blacksmith et al., 2019). We adopted the six-questions version developed by Primi et al. (2016), with an additional seventh question also taken from their study. This long version retains the balance between correct and heuristic, but incorrect, answers typical of the original CRT. Given the possibility of previous exposure to the standard CRT, we modified the wording of each question.[5] The original questions and the version used in our experiment are reported in Table 1.

We paid subjects \$0.10 for each correct answer. Subjects were given up to 10 minutes to complete the test. The time limit was set to ensure that subjects had ample time to complete the test without feeling time-pressured.

In the second part of the experiment, subjects were matched into teams of two and told that there are two urns, *Orange* and *Purple*, each containing 3 balls. The *Orange* urn contains 2 orange balls and 1 purple ball, while the *Purple* urn contains 1 orange ball and 2 purple balls.

[5] Other researchers have also opted for variations on the original questions for similar reasons (e.g., Snowberg and Yariv, 2021).

	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 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.

The computer selected one of the two urns with equal probability for each two-player team. None of the subjects were told which urn was selected for their team.

In the **Baseline** treatment, the computer drew a new ball with replacement from the selected urn for 30 consecutive periods. The color of each drawn ball was shown to all subjects in the same team; that is, the signals were public. The informed treatments were identical to the baseline treatment with the exception that each subject was given some information about her partner's test score before any balls were drawn. Specifically, each subject was told whether her partner's test score was in the top (**Informed-Top** treatment) or bottom (**Informed-Bottom** treatment) half of test scores of 50 randomly chosen subjects that previously took the same test.[6] None of the subjects was informed about their own test score until the end of the experiment.[7]

[6]Given the online nature of the experiment (more on this below) with participants arriving randomly over time, we first collected test score data from about 100 subjects in the baseline treatment in 2019. We then selected the scores of 50 randomly chosen subjects and computed the median test score which was used to generate the partners' types for the informed treatments. A subject was told that her partner's test score was in the *Top* half of the distribution if it was at or above the median and *Bottom* otherwise. Before the second round of data collection in 2022, we randomly redrew the scores of 50 randomly chosen subjects who completed the experiment in the first round of data collection. In both instances, the median test score of the randomly selected samples was 4 out of 7.

[7]We did not provide subjects with information about their own test score to avoid drawing attention to it as our treatment involves manipulating information about the partner's test score. Also, information about own test score could affect the formation of first-order beliefs if subjects were to doubt their own updating behavior as

Each subject made two choices in each period. First, they bet on the event that the secret urn is *Orange*. Then, they bet on the guess made by their partner. Each choice was presented on a separate page, and we gave each subject up to two minutes (one minute in the second round of data collection) to submit each choice.[8] The experiment was framed neutrally in that it avoided all references to guesses or beliefs, instead explaining the task as a betting problem. The experiment had 30 periods to allow sufficient time for subjects to refine their beliefs about the state of the world as well as their partner's beliefs about the state.

Each subject was paid on the basis of one randomly chosen decision out of sixty. Paying for only one randomly chosen decision breaks any intertemporal hedging across and within periods, turning each choice into a standalone decision problem.[9] To avoid the influence of risk aversion on subjects' elicited beliefs, we employed the binarized scoring rule of Hossain and Okui (2013) with a bonus of $3 or nothing. This elicitation procedure is incentive compatible irrespective of attitudes toward risk and relatively simple to implement. The rule is applied twice to elicit a subject's first- and second-order beliefs in every period.[10]

Subjects received no feedback about the actions of their matched partner during the experiment.[11] Providing subjects with information about their partner's guesses of the state could affect their own belief updating model. For example, subjects could start doubting their own interpretation of previous signals and simply mimic past behavior of their partner. Feedback could also affect a subject's mental model about others' updating processes whose study is a central goal of this paper.

At the end of the experiment, each subject received information about which urn was selected, which decision was randomly chosen for payment, as well as the corresponding target (state for

well as about how their own test score compares with their partner's. For example, Lambrecht et al. (2021) find that disclosure of information about both subjects' test scores can have different effects on behavior in repeated games depending on the type of game being played.

[8]A hard time constraint is commonly used in interactive online experiments to minimize the chance that a subject is stuck on a waiting page without the ability to submit their assignment due to their partner either experiencing any technical issue or simply dropping out. After observing that less than 1% of decisions in the first round of data collection took more than 60 seconds, we reduced the time limit to one minute per decision in the second round of data collection.

[9]Azrieli et al. (2018) show theoretically that selecting one task at random is the only incentive compatible way to pay subjects under a mild monotonicity assumption on subjects' preferences.

[10]Danz et al. (2020) recently found a pull-to-the-center effect on beliefs elicited with the binarized scoring rule, which we were unaware of when we ran the first round of data collection in 2019. The second round of data collection also employed the same elicitation procedure for consistency. While this effect could in principle affect our elicitations, it is unlikely to affect the elicitation of first- and second-order beliefs differently for the same subject. Furthermore, it is also unclear whether other well-known elicitation procedures might be subject to the same effect.

[11]Lack of feedback is common in experiments measuring subjects' beliefs about the beliefs of others (see e.g., Stahl and Wilson, 1995; Costa-Gomes et al., 2001; Costa-Gomes and Crawford, 2006; Costa-Gomes and Weizsäcker, 2008).

the first decision, partner's choice for the second decision), which number was drawn to implement the binarized scoring rule, whether or not the subject obtained the additional bonus, and the subject's own test score.

The instructions and screenshots can be found in the online appendix.[12]

3 Results

We collected data from U.S. workers on Amazon Mechanical Turk (MTurk) that completed at least 100 tasks with an overall approval rate of at least 95%. Data were collected between June and September 2019 and also in May 2022 using the software oTree (Chen et al., 2016). 672 unique subjects were matched and successfully completed the experiment, with 254 subjects in the baseline, 188 in Informed-Top, and 230 in Informed-Bottom.[13] The median hourly wage was \$10.91, and the median subject spent approximately 19 minutes to complete the experiment.[14]

	Primi et al. (2016)	Our experiment
	% Correct (Heuristic)	% Correct (Heuristic)
Question 1	39% (49%)	39.9% (44.9%)
Question 2	44% (48%)	58.9% (28.9%)
Question 3	54% (36%)	58.5% (26.6%)
Question 4	81% (15%)	71.0% (15.0%)
Question 5	37% (36%)	37.5% (23.2%)
Question 6	49% (37%)	39.0% (31.4%)
Question 7		52.7% (24.4%)

Table 2: **Percentages of correct and heuristic responses in the CRT: our data vs. Primi et al. (2016).**

Table 2 reports the proportion of correct versus heuristic (but incorrect) answers in the CRT

[12] `https://drive.google.com/file/d/117vY47Z7q27VeKRvSkEUi5AMl22buUT1/view?usp=sharingg`.

[13] Part of the data collected for the baseline treatment (204 subjects in 2019) was analyzed in Evdokimov and Garfagnini (2021). We did not look at the effect of cognitive ability in that paper. The data collected for the Informed-Top and Informed-Bottom treatments are analyzed here for the first time. Amazon MTurk allows requesters (i.e., experimenters) to assign a qualification to each subject that accepts one of the posted assignments. This qualification prevents an MTurk worker from participating in the experiment more than once. The same qualification was used in both rounds of data collection.

[14] For the second round of data collection in May 2022, we also introduced a completion bonus of \$1 which, however, did not affect incentives in the elicitation tasks. Actual payments averaged \$3.09 with a median payment of \$3.55. Hara et al. (2018) estimate an average hourly wage per worker on Amazon MTurk of between \$3.13 and \$3.48. Thus, the average hourly compensation in our experiment, \$10.96, is about 3 times the average hourly wage as well as above the minimum federal wage of \$7.25. Furthermore, the median hourly wage per worker estimated by Hara et al. is between \$1.77 and \$2.11, compared to a median hourly wage per worker of \$10.91 from participating in our experiment.

from the study by Primi et al. (2016) and our data.[15] As can be seen, our results are similar to theirs despite the fact that the questions were worded differently. The median score in the overall sample is four (four in both Baseline and Informed-Top, and three in Informed-Bottom).

Figure 1 shows the average distance between subjects' beliefs and the corresponding Bayesian posterior beliefs as a function of test score. First-order beliefs of subjects with higher cognitive ability deviate less from the Bayesian benchmark ($P < 0.001$, first column of Table 3). Figure 2 further maps the evolution of normalized first-order beliefs of subjects below (Panel a) and above (Panel b) the median test score. If μ is the elicited belief, the normalized belief is defined as μ when the state is *Orange* and $1 - \mu$ when the state is *Purple*. Thus, higher values correspond to more accurate beliefs. The figure shows a stark difference in belief updating behavior as a function of test score.

In the second column of Table 3, we regress subjects' normalized first-order beliefs on CRT scores and the treatment variables. In line with prior studies (Oechssler et al., 2009; Hoppe and Kusterer, 2011; Charness et al., 2012), we find that subjects with higher CRT scores form more accurate first-order beliefs ($P < 0.001$).[16] Neither the Informed-Bottom nor the Informed-Top treatment had an effect on normalized first-order beliefs while controlling for test score (smallest $P = 0.425$). These results serve as a sanity check, as there is no reason to expect that updating of first-order beliefs is affected by the treatment variables.

We have thus established that cognitive ability has a significant effect on the way in which subjects update beliefs. The next section investigates whether cognitive ability also affects the mental model that subjects hold about the belief updating behavior of others and what effect this has on their perception of disagreement between their beliefs and their partner's beliefs.

3.1 Cognitive abilities and perceived disagreement

In this section, we investigate the effect of cognitive abilities on the extent of disagreement perceived by a subject between their beliefs about the state, i.e. their first-order beliefs, and their partner's first-order beliefs. Recall that in our design subjects belonging to the same team observe the same public information about the state of the world over which they form beliefs. While Bayesian subjects with knowledge of each others' Bayesianism would show no actual or perceived disagreement, this may not be the case in practice. Formally, we use the following

[15] We cannot report the information from Question 7 in Primi et al. because it was not included in the final test questions and thus the proportion of correct versus heuristic answers is not reported in their paper.

[16] Here, as in the rest of our statistical analysis, the standard errors are clustered by session. We had 58 sessions in the experiment. Appendix A provides a more in-depth analysis of first-order beliefs by estimating Grether-style regressions. An IV analysis finds that the degree of base-rate neglect is independent of test score while higher test scores are associated with less under-inference.

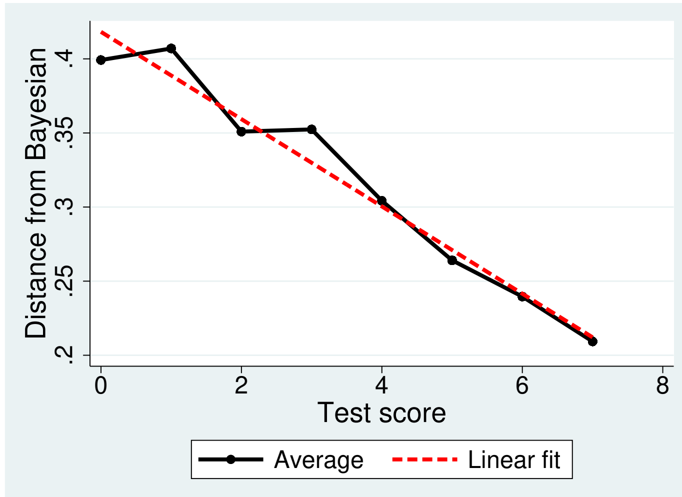
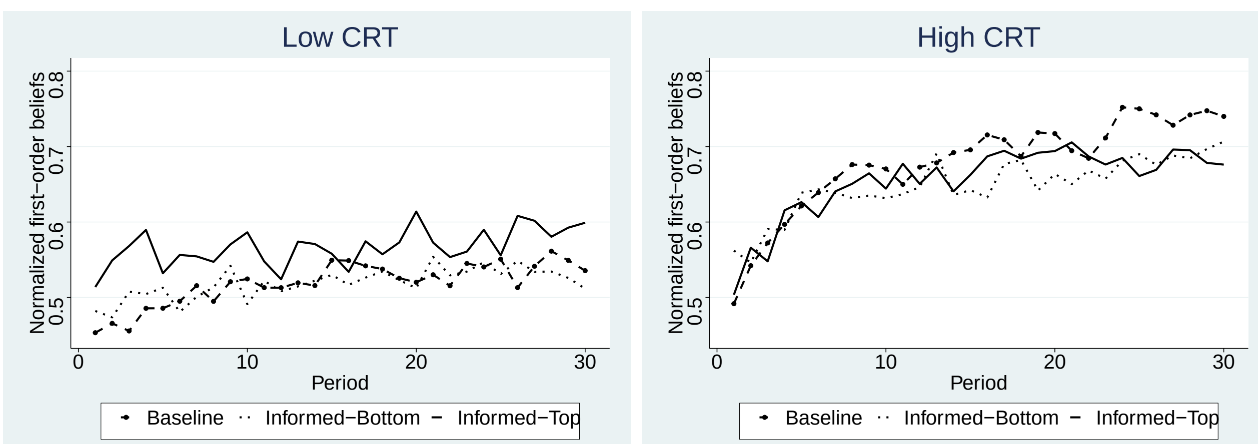


Figure 1: **Distance from Bayesian beliefs by test score.** Dashed line represents a linear fit.



(a) Below median.

(b) Above median.

Figure 2: **Normalized first-order beliefs by cognitive type.** When controlling for test score, subjects update their first-order beliefs in a similar manner across treatments.

	(1) Distance from Bayesian	(2) Normalized 1st-order beliefs
	All subjects	All subjects
Informed-Top	-0.0073	0.0069
	(0.0219)	(0.0249)
Informed-Bottom	0.0114	-0.0137
	(0.0222)	(0.0250)
Test Score	-0.0290****	0.0288****
	(0.0031)	(0.0030)
Constant	0.4148****	0.4971****
	(0.0221)	(0.0227)
Observations	20160	20160

Session-clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table 3: **First-order beliefs by treatment.**

measures:

- **Perceived disagreement.** We define perceived disagreement as $PD_{it} = |First_{it} - Second_{it}|$, where $First_{it}$ and $Second_{it}$ are the first- and second-order beliefs of subject i in period t. Intuitively, a subject will show minimal perceived disagreement if she believes that her matched partner uses the same updating rule as she herself does.[17]

- **Perceived disagreement dummy.** We use the binary measure, PD_Dum_{it}, that is equal to one if $PD_{it} = 0$ and zero otherwise. This measure keeps track of how often subjects ascribed to their matched partners beliefs that are in disagreement with their own first-order beliefs. Beliefs show a positive perceived disagreement for 56.65% of all observations.

- **Direction of Perceived disagreement.** We define this variable as follows:

$$DirPD_{it} = \begin{cases} First_{it} - Second_{it}, & \text{if } First_{it} > 0.5, \\ Second_{it} - First_{it}, & \text{if } First_{it} < 0.5. \end{cases} \tag{1}$$

[17]It if of course possible that a subject might assign positive probability to several updating rules used by the matched partner. As our design only elicits the first moment of the distribution of second-order beliefs, we cannot test whether two different subjects who report the same mean perceive uncertainty about their partner's updating behavior differently. Nevertheless, both subjects still share the same expectation about the first-order beliefs of the matched partner. Manski and Neri (2013) elicit both the mean and the distribution of subjects' second-order beliefs in a Hide-and-Seek game and find a general consistency between the two, that is, the mean of the distribution of second-order beliefs is consistent with its point estimate.

Intuitively, if subject i in period t, holds first-order beliefs above 0.5, a positive value of the $DirPD_{it}$ variable hints at a tendency to believe that another subject, who has observed the same information about the state of the world holds a belief that puts less weight on the same state of the world than one's own belief. A similar argument holds if the subject's first-order beliefs are below 0.5. When a subject's first-order beliefs are equal to 0.5, second-order beliefs above or below 0.5 are difficult to interpret. For this reason, the definition of the direction of perceived disagreement excludes observations where $First_{it} = 0.5$ from the analysis (13.42% of total observations).[18]

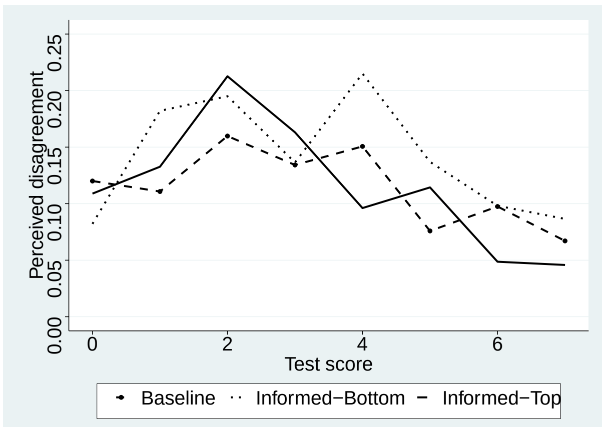
Figure 3 (a) shows the average perceived disagreement for different levels of cognitive ability. Perceived disagreement seems to be negatively related to cognitive ability. In the first column of Table 4, we regress perceived disagreement against the treatment variables while controlling for test score in the CRT. We find that subjects with higher cognitive ability do show less perceived disagreement ($P < 0.001$). Similarly, the probability of reporting a positive perceived disagreement is lower for higher CRT scores ($P < 0.001$, second column of Table 4 and Figure 3 b).

Looking at its direction (third column of Table 4), perceived disagreement is systematically biased toward imputing on others beliefs that assign a lower likelihood to the state which is thought to be more likely. For instance, suppose that a subject assigns probability greater than 0.5 to the state being orange (resp., less than 0.5), then on average the subject tends to report second-order beliefs which are closer to 0 (resp., 1) than one's own beliefs. This tendency is stronger for subjects of lower cognitive ability ($P < 0.05$). This result is reminiscent of the assumption in level-k models that decision makers ascribe lower levels of sophistication to others than to themselves (see also, Huck and Weizsäcker, 2002; Kübler and Weizsäcker, 2004)

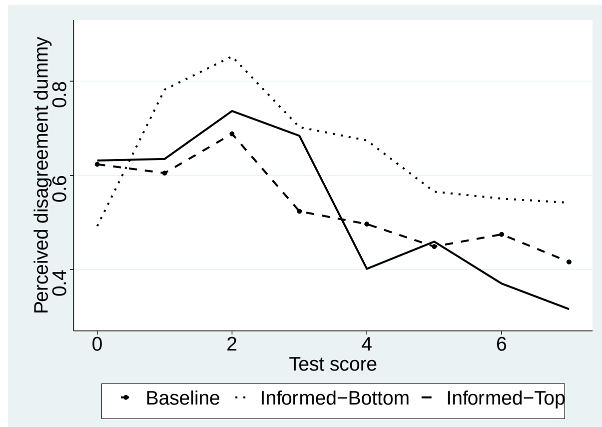
Result 1. *Subjects with lower cognitive ability report greater perceived disagreement between their first-order beliefs and those of their matched partner compared to more cognitive able subjects and, on average, tend to assume that their partner assigns a lower probability to the state that they believe to be most probable.*

One natural question to ask is how perceived disagreement compares with actual disagreement, i.e., the distance between matched partners' first-order beliefs. Given the relationship between cognitive skills and belief formation, it is reasonable to expect that actual disagreement might be related to the distance between matched partners' cognitive skills. Figure 3 (c) plots average actual disagreement as a function of the test score gap, defined as the absolute value of

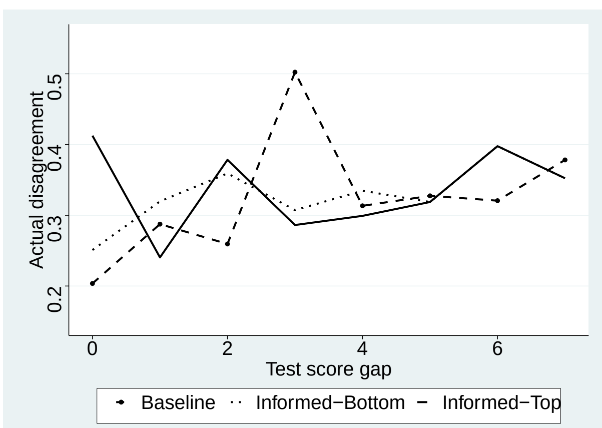
[18] The probability of reporting first-order beliefs equal to 0.5 is unaffected by test score or treatment (smallest $P = 0.660$ in a logit regression controlling for test score and treatment dummies).



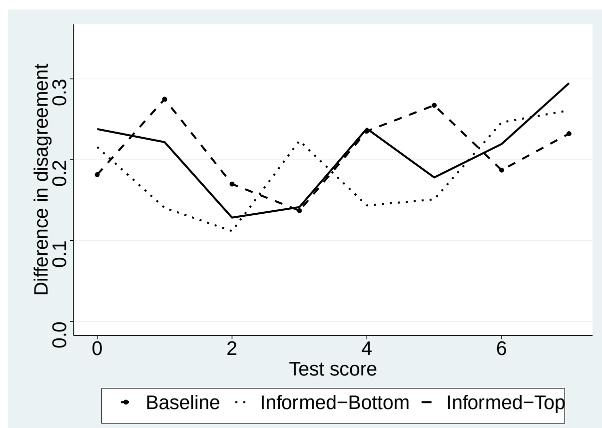
(a) Perceived disagreement



(b) Perceived disagreement dummy



(c) Actual disagreement



(d) Difference between actual and perceived disagreement

Figure 3: **Cognitive ability and disagreement.** Actual disagreement is higher than perceived disagreement, regardless of test score, and subjects with higher cognitive ability report a lower perceived disagreement.

	(1) Perceived disagreement	(2) Perceived disagreement dummy	(3) Direction of perceived disagreement	(4) Actual disagreement
	OLS	Logit	OLS	OLS
Informed-Top	-0.0001	-0.0715	-0.0052	0.0057
	(0.0115)	(0.1478)	(0.0121)	(0.0230)
Informed-Bottom	0.0273**	0.4381***	0.0186	0.0104
	(0.0123)	(0.1638)	(0.0142)	(0.0187)
Test score	-0.0094****	-0.1471****	-0.0056**	-0.0016
	(0.0025)	(0.0310)	(0.0026)	(0.0025)
Test score gap				0.0117***
				(0.0034)
Constant	0.1447****	0.6768****	0.1146****	0.2915****
	(0.0135)	(0.1856)	(0.0131)	(0.0180)
Observations	20160	20160	17455	20160

Session-clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table 4: **Disagreement.** Lower cognitive ability is associated with higher perceived disagreement and subjects are more likely to believe that their partner assigns a lower probability to the state thought to be more probable. Being informed that your partner is in the bottom half of the distribution increases perceived disagreement.

the difference in test scores of the two matched partners. The figure suggests a positive relationship, which we find to be significant when we regress actual disagreement on the test score gap ($P < 0.01$, fourth column of Table 4), while the coefficient on the test score is not significant ($P = 0.52$) as one would have expected. These results suggest that the actual disagreement at the level of a match is determined not by individual cognitive ability, but by the cognitive ability difference in the match, which our subjects are not informed about.

Figure 3 (d) shows that actual disagreement is higher than perceived disagreement regardless of cognitive ability, on average. This is confirmed by a regression of the difference between actual and perceived disagreement against the treatment dummies, the test score and the test score gap (largest $P < 0.001$). Furthermore, an increase in cognitive ability widens the gap between actual and perceived disagreement ($P < 0.05$).

Result 2. *Actual disagreement is larger than perceived disagreement regardless of cognitive ability. Also, higher cognitive types underestimate disagreement more than lower cognitive types.*

A deeper look at 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.

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 $(0.63, 1]$.

Table 5 shows the distribution of second-order belief types for each treatment and overall. The modal behavior across all treatments (fourth row) is to report equal beliefs but this accounts for only 39% of all relevant observations. In line with our previous discussion of the direction of perceived disagreement, most observations involve either reporting that the matched partner has more conservative beliefs than one's own or even polarized beliefs. Tables 6 and 7 further disaggregate data by cognitive type, above or below the median test score, 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 Informed-Bottom. By contrast, low cognitive types' modal behavior is to report more conservative and/or polarized second-order beliefs.

To formally test for the effect of cognitive ability, Table 8 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 $P < 0.01$). Furthermore, the effect of cognitive ability is stronger for polarized and overinferred compared to conservative beliefs (largest $P < 0.001$).

	Conservative	Polarized	Overinferred	Equal
Baseline	28.9%	13.3%	14.3%	43.5%
Informed-Top	25.6%	13.5%	15.1%	45.8%
Informed-Bottom	34.1%	17.9%	19.5%	28.5%
Overall	29.8%	14.9%	16.3%	39.0%

Table 5: Distribution of second-order belief types.

	Conservative	Polarized	Overinferred	Equal
Baseline	29.3%	10.0%	10.1%	50.6%
Informed-Top	22.4%	7.4%	10.4%	59.8%
Informed-Bottom	37.1%	14.1%	12.1%	36.7%
Overall	29.4%	10.3%	10.8%	49.5%

Table 6: Distribution of second-order belief types for above median test scorers.

	Conservative	Polarized	Overinferred	Equal
Baseline	28.4%	17.3%	19.3%	35.0%
Informed-Top	29.5%	21.0%	20.8%	28.7%
Informed-Bottom	32.1%	20.6%	24.7%	22.6%
Overall	30.2%	19.5%	21.8%	28.5%

Table 7: Distribution of second-order belief types for below median test scorers.

We summarize the discussion so far in the following result:[19]

Result 3. *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.*

[19]Table 11 in the Appendix replicates the analysis at the subject level by categorizing subjects based on their modal behavior and shows similar results to the aggregate analysis.

	(1) Conservative	(2) Overinferred	(3) Polarized
Informed-Top	-0.1666	0.0218	-0.0195
	(0.1534)	(0.1715)	(0.1868)
Informed-Bottom	0.5342****	0.6114***	0.6042***
	(0.1612)	(0.2011)	(0.2004)
Test Score	-0.0943***	-0.2363****	-0.2245****
	(0.0331)	(0.0318)	(0.0431)
Constant	-0.0271	-0.2561	-0.3625
	(0.1852)	(0.1956)	(0.2260)
Observations	17455	17455	17455

Session-clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table 8: **Multinomial logit for 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.

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 (Royzman et al., 2003).[20] 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 et al. (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 project 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, Result 3 suggests that egocentrism might be increasing with cognitive ability.

The next section discusses treatment effects.

3.2 Treatment effects and learning

Table 4 shows the treatment effects of being informed of your partner's cognitive ability on perceived disagreement, its direction and actual disagreement. We control for own cognitive ability, as measured by the test score.

[20] See also Samson et al. (2010), and Frith (2012) for a recent review.

Being informed that your partner is in the top half of the distribution has no significant effect on any of the dependent variables (smallest $P = 0.629$, third row of Table 4). One possible explanation is that subjects generally assume that both they and their partners are in the top half of the distribution even if, in the Informed-Top treatment, their partners could be more sophisticated than they themselves are. While we do not investigate overconfidence directly, this hypothesis finds some support in the psychology literature, where the majority of subjects rate themselves above the median (Moore and Healy, 2008). Overconfidence could also be affected by the nature of the CRT whereby subjects who reported heuristic, albeit incorrect, answers in the cognitive test might incorrectly believe to have scored above the median.[21]

By contrast, being informed that your partner is in the bottom half of the distribution increases both the size and the frequency of perceived disagreement ($P < 0.05$ with both measures, first and second column of Table 4). While there is no effect on the direction of disagreement between the baseline and either treatment (smallest $P = 0.195$), there is a marginally significant and positive effect between the two treatments ($P = 0.053$). I.e., subjects reasonably act as if their partner is more conservative than they are when processing information if they are informed that their partner's CRT score is low as opposed to high. However, these results hide some heterogeneity in the types of second-order beliefs formed across treatments. The multinomial logit analysis in Table 8 shows that the log odds of reporting conservative, polarized or overinferred beliefs versus equal beliefs are all insignificantly different between the baseline and Informed-Top treatment (smallest $P = 0.277$). By contrast, all the log odds are higher in the Informed-Bottom treatment compared to the baseline (largest $P < 0.01$). Thus, although the direction of disagreement shows no treatment effects, the distribution of actual responses does vary with an additional 10% of observations in the Informed-Bottom treatment being either conservative or polarized. These responses though are counterbalanced by an increase in overinferred beliefs by 5 percentage points.

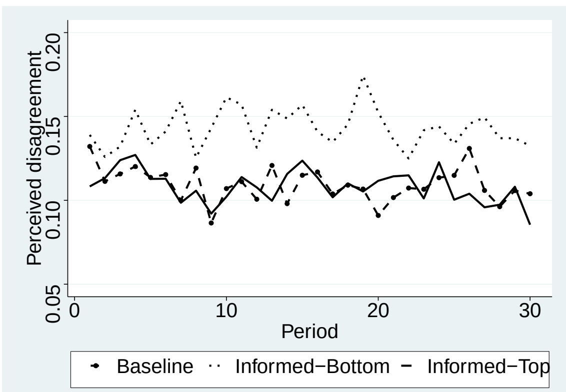
As one would expect, information about your partner's test score has no significant effect on actual disagreement (smallest $P = 0.578$, fourth column of Table 4). This is because information does not affect the formation of first-order beliefs as already pointed out at the beginning of Section 3.

We summarize these results as follows:

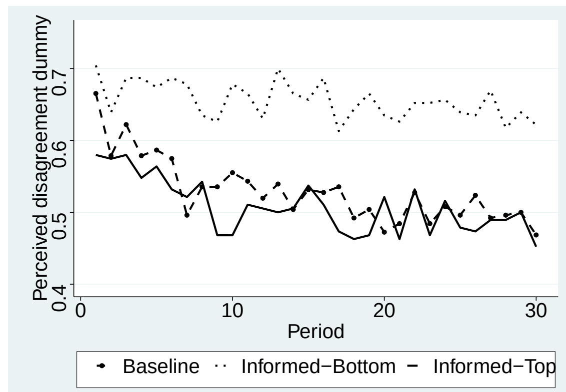
RESULT 4. *Providing subjects with information about their partner's test score increases the extent of perceived disagreement in the Informed-Bottom treatment while it does not affect it in the Informed-Top treatment compared to the baseline.*

[21]Including heuristic answers as if they were correct, the median test score would increase from 4 to 6.

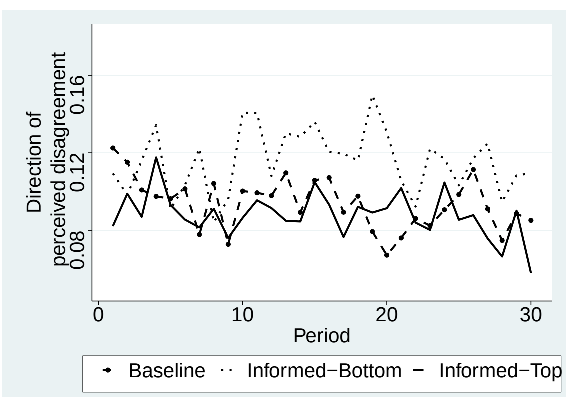
Given our sequential design, we can also investigate the evolution of perceived, and actual, disagreement over time. While the theory literature has shown that disagreement can persist in the long-run, this result rests on heterogeneity in prior beliefs or subjects holding some private information. By contrast, in our setting subjects hold a uniform prior and all information is public. Under these conditions and despite the presence of some well-known heterogeneity in belief updating behavior across agents (e.g., Holt and Smith, 2009), it is more reasonable to conjecture that an increase in the amount of information available to agents, namely, the number of public signals, might have a mitigating effect on the extent of both actual and perceived disagreement.



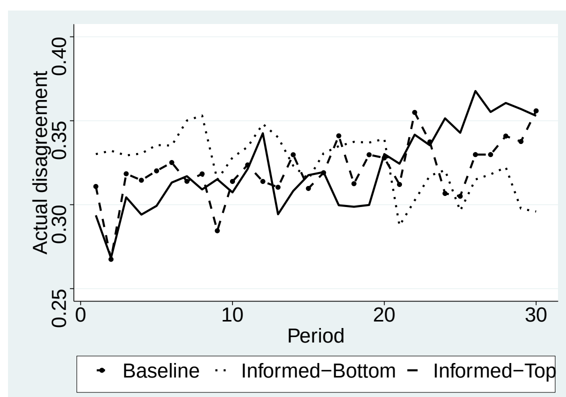
(a) Perceived disagreement



(b) Perceived disagreement dummy



(c) Direction of perceived disagreement



(d) Actual disagreement

Figure 4: **Treatments effects and learning.**

Figure 4 shows the evolution of the measures of disagreement over time. Consistent with the reported treatment effects, the dotted line (behavior in the Informed-Bottom treatment) is above the others in most periods for the first three panels. The figure also suggests that perceived disagreement becomes less likely over time, although the size of actual disagreement displays an opposite trend in most treatments. We test these hypotheses in Table 9 in which we add a time trend and an interaction between the time trend and the Informed-Bottom variable to the regressions in Table 4. We aggregate data for the Baseline and Informed-Top treatments, as we find no significant treatment effect of being informed that your partner is in the top half of the distribution.

The regression results show that the time trend is negative and significant for the frequency of perceived disagreement ($P < 0.001$ for the baseline and Informed-Top treatment, $p < 0.05$ for the Informed-Bottom treatment in a logit regression) although not statistically significant for the size of perceived disagreement (smallest $P = 0.323$). Thus, despite the fact that the size of perceived disagreement stays roughly constant over time for all treatments, subjects are more likely to report second-order beliefs equal to their own first-order beliefs (i.e., no perceived disagreement) as more informative signals about the state are observed, regardless of the treatment. Further disaggregating the data (see Figure 6 in the appendix) shows that the treatment effect is entirely driven by higher cognitive types.

Looking at actual disagreement, we find a positive time trend in the baseline and Informed-Top treatments ($P < 0.05$) and a negative but insignificant trend in the Informed-Bottom treatment ($P = 0.102$). Furthermore, the difference between actual and perceived disagreement increases over time in the baseline and Informed-Top treatment ($P < 0.05$). Thus, although the average subject perceives little change in disagreement over time, the opposite is the case.

RESULT 5. *The frequency of perceived disagreement decreases over time although perceived disagreement remains higher in the Informed-Bottom treatment compared to the other treatments. At the same time, actual disagreement increases over time in all but the Informed-Bottom treatment where it stays roughly constant.*

As our design features belief elicitation over only 30 consecutive periods, it would be reasonable to conjecture that a larger number of periods (i.e., informative signals) might eventually lead to lower disagreement. Evdokimov and Garfagnini (2021) define higher-order learning as increasing accuracy of higher-order beliefs and find that higher-order learning fails in a treatment in which subjects observe 300 public signals about the state of the world. That paper finds that the failure to form correct beliefs about the beliefs of others is linked to heterogeneity in updating behavior and base-rate neglect, which is likely to affect the dynamics of perceived disagreement as well. Similarly, Esponda et al. (2022) find persistence of suboptimal behavior, i.e. deviation from the Bayesian benchmark, even after 200 rounds of feedback in a standard belief updating task, also linking their findings to base-rate neglect. Thus, it is possible that additional information might have a limited effect on the dynamics of perceived disagreement.

	(1) Perceived disagreement	(2) Perceived disagreement dummy	(3) Direction of perceived disagreement	(4) Actual disagreement
	OLS	Logit	OLS	OLS
t	-0.0004	-0.0148****	-0.0006	0.0016**
	(0.0004)	(0.0035)	(0.0004)	(0.0008)
Informed-Bottom	0.0224*	0.3503**	0.0117	0.0512***
	(0.0116)	(0.1464)	(0.0136)	(0.0158)
t × Informed-Bottom	0.0003	0.0076	0.0006	-0.0028***
	(0.0005)	(0.0049)	(0.0006)	(0.0009)
Test score	-0.0094****	-0.1477****	-0.0056**	-0.0016
	(0.0025)	(0.0312)	(0.0026)	(0.0025)
Test score gap				0.0117***
				(0.0036)
Constant	0.1501****	0.8779****	0.1211****	0.2688****
	(0.0129)	(0.1585)	(0.0122)	(0.0159)
Observations	20160	20160	17455	20160

Session-clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table 9: **Regressions for treatment effects and learning.**

4 Conclusion

This paper is concerned with the role of cognitive ability on the evolution of perceived disagreement as individuals observe sequential public information about a state of the world, such as economic fundamentals. We find that higher cognitive ability is associated with lower perceived disagreement although actual disagreement stays persistently high over time. Providing subjects with information about the cognitive skills of others has an effect on mental models only when a subject is informed that the matched partner is of low cognitive ability. In this case, subjects correctly perceive higher disagreement although they still underestimate its real extent.

Our experiments show that perceived disagreement is unlikely to disappear over time even when subjects start from the same prior and only observe public information. This reinforces the findings of Andreoni and Mylovanov (2012) that also document persistence of disagreement. In their design, however, subjects initially observe a private signal followed by a sequence of public signals. They find that providing information about others' actions does not resolve disagreement as subjects mistrust the decision-making skills of others and thus the information conveyed through their actions.[22] Our design only involves public information, yet disagreement

[22]This phenomenon is similar to the overweighting of one's own private information documented in social

persists even with double the number of public signals employed by Andreoni and Mylovanov.

The fact that perceived disagreement is much lower than what it really is and negatively affected by cognitive ability raises several interesting questions. For instance, why do more cognitively able subjects seem to systematically underestimate true disagreement to a greater extent?

From a policy perspective, our study shows that policies aimed at increasing the amount of public information available to agents may fail to improve decision making in strategic environments which involve coordination motives. Our documented persistence in the perception of disagreement implies that agents believe, on average, that others hold worse beliefs than they do even as more information is observed. This may result in coordination failures. More importantly, perception of disagreement underestimates actual disagreement. So, even a reduction in perceived disagreement could still result in miscoordination. The silver lining, though, is that policies aimed at mitigating mistakes that agents make in processing information, by reducing actual disagreement via more accurate beliefs, may lead to improved coordination outcomes at least among more cognitively able agents as those exhibit stronger egocentric tendencies and are thus more prone to believe that others form similar beliefs to theirs.[23] Understanding the link between information processing, perceived disagreement and coordination represents an interesting avenue for future research.

References

Acemoglu, D., V. Chernozhukov, and M. Yildiz (2006): "Learning and disagreement in an uncertain world," .

——— (2016): "Fragility of asymptotic agreement under Bayesian learning," *Theoretical Economics*, 11, 187–225.

Agranov, M., E. Potamites, A. Schotter, and C. Tergiman (2012): "Beliefs and endogenous cognitive levels: An experimental study," *Games and Economic Behavior*, 75, 449–463.

Andreoni, J. and T. Mylovanov (2012): "Diverging opinions," *American Economic Journal: Microeconomics*, 4, 209–32.

learning experiments (e.g. Weizsäcker, 2010). One possible interpretation is that subjects believe that others are less rational than they are and thus more prone to make mistakes when using their own private information. However, Drehmann et al. (2005) have shown that even when subjects observe others' private information directly, as opposed to just actions, overweighting of one's own private information still occurs. Thus, private information may add an additional channel for disagreement to persist which is not present in our design.

[23]This observation already finds some support in recent work by Gill and Prowse (2016) and Proto et al. (2019).

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

Baliga, S., E. Hanany, and P. Klibanoff (2013): "Polarization and ambiguity," *American Economic Review*, 103, 3071–83.

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

Birch, S. A. and P. Bloom (2004): "Understanding children's and adults' limitations in mental state reasoning," *Trends in Cognitive Sciences*, 8, 255–260.

Blacksmith, N., Y. Yang, T. S. Behrend, and G. A. Ruark (2019): "Assessing the validity of inferences from scores on the cognitive reflection test," *Journal of Behavioral Decision Making*, 32, 599–612.

Brañas-Garza, P., T. Garcia-Munoz, and R. H. González (2012): "Cognitive effort in the beauty contest game," *Journal of Economic Behavior & Organization*, 83, 254–260.

Burks, S. V., J. P. Carpenter, L. Goette, and A. Rustichini (2009): "Cognitive skills affect economic preferences, strategic behavior, and job attachment," *Proceedings of the National Academy of Sciences*, 106, 7745–7750.

Carpenter, J., M. Graham, and J. Wolf (2013): "Cognitive ability and strategic sophistication," *Games and Economic Behavior*, 80, 115–130.

Charness, G., A. Rustichini, and J. Van de Ven (2012): "Self-confidence and strategic behavior," Tech. rep., unpublished version.

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.

Costa-Gomes, M., V. P. Crawford, and B. Broseta (2001): "Cognition and behavior in normal-form games: An experimental study," *Econometrica*, 69, 1193–1235.

Costa-Gomes, M. A. and V. P. Crawford (2006): "Cognition and behavior in two-person guessing games: An experimental study," *American Economic Review*, 96, 1737–1768.

Costa-Gomes, M. A. and G. Weizsäcker (2008): "Stated beliefs and play in normal-form games," *The Review of Economic Studies*, 75, 729–762.

Cripps, M. W., J. C. Ely, G. J. Mailath, and L. Samuelson (2008): "Common learning," *Econometrica*, 76, 909–933.

Danz, D., K. Madarász, and S. Wang (2018): "The biases of others: Projection equilibrium in an agency setting," .

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

Drehmann, M., J. Oechssler, and A. Roider (2005): "Herding and contrarian behavior in financial markets: An internet experiment," *American Economic Review*, 95, 1403–1426.

Edwards, W. (1968): "Conservatism in human information processing," in *Formal Representation of Human Judgement*, ed. by B. Kleinmuntz, New York, 17–52.

Enke, B., U. Gneezy, B. J. Hall, D. Martin, V. Nelidov, T. Offerman, and J. v. de Ven (2020): "Cognitive Biases: Mistakes or Missing Stakes?" .

Esponda, I., E. Vespa, and S. Yuksel (2022): "Mental models and learning: The case of base-rate neglect," Tech. rep., mimeo.

Evdokimov, P. and U. Garfagnini (2021): "Higher-order learning," *Experimental economics*, forthcoming.

Fe, E., D. Gill, and V. Prowse (2022): "Cognitive skills, strategic sophistication, and life outcomes," *Journal of Political Economy*, 130, 2643–2704.

Frederick, S. (2005): "Cognitive reflection and decision making," *Journal of Economic Perspectives*, 19, 25–42.

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

Gill, D. and V. Prowse (2016): "Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis," *Journal of Political Economy*, 124, 1619–1676.

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

Hara, K., A. Adams, K. Milland, S. Savage, C. Callison-Burch, and J. P. Bigham (2018): "A data-driven analysis of workers' earnings on amazon mechanical turk," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 449.

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

Hoppe, E. I. and D. J. Kusterer (2011): "Behavioral biases and cognitive reflection," *Economics Letters*, 110, 97–100.

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

Huck, S. and G. Weizsäcker (2002): "Do players correctly estimate what others do? Evidence of conservatism in beliefs," *Journal of Economic Behavior & Organization*, 47, 71–85.

Kübler, D. and G. Weizsäcker (2004): "Limited depth of reasoning and failure of cascade formation in the laboratory," *The Review of Economic Studies*, 71, 425–441.

Lambrecht, M., E. Proto, A. Rustichini, and A. Sofianos (2021): "Intelligence disclosure and cooperation in repeated interactions," .

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

Manski, C. F. and C. Neri (2013): "First-and second-order subjective expectations in strategic decision-making: Experimental evidence," *Games and Economic Behavior*, 81, 232–254.

Moore, D. A. and P. J. Healy (2008): "The trouble with overconfidence." *Psychological review*, 115, 502.

Oechssler, J., A. Roider, and P. W. Schmitz (2009): "Cognitive abilities and behavioral biases," *Journal of Economic Behavior & Organization*, 72, 147–152.

Primi, C., K. Morsanyi, F. Chiesi, M. A. Donati, and J. 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.

Proto, E., A. Rustichini, and A. Sofianos (2019): "Intelligence, Personality, and Gains from Cooperation in Repeated Interactions," *Journal of Political Economy*, 127, 1351–1390.

Royzman, E. B., K. W. Cassidy, and J. Baron (2003): ""I know, you know": Epistemic egocentrism in children and adults," *Review of General Psychology*, 7, 38–65.

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

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

Samson, D., I. A. Apperly, J. J. Braithwaite, B. J. Andrews, and S. 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.

Snowberg, E. and L. Yariv (2021): "Testing the waters: Behavior across participant pools," *American Economic Review*, 111, 687–719.

Stahl, D. O. and P. W. Wilson (1995): "On players' models of other players: Theory and experimental evidence," *Games and Economic Behavior*, 10, 218–254.

Weizsäcker, G. (2010): "Do we follow others when we should? A simple test of rational expectations," *American Economic Review*, 100, 2340–60.

A 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

$$\frac{\mu_n}{1-\mu_n} = \frac{\mu_{n-1}}{1-\mu_{n-1}} \underbrace{\frac{\text{Prob(Current ball}|\text{Urn = Orange)}}{\text{Prob(Current ball}|\text{Urn = Purple)}}}_{LR_n}, \tag{2}$$

where μ_n is the subject's first-order posterior beliefs, μ_{n-1} is her prior beliefs,[24] and LR_n is the likelihood ratio following the observation of the current ball, with $LR_n \in \{LR^{Orange} = 2, LR^{Purple} = \frac{1}{2}\}$.

We estimate the following model:

$$\ln\left(\frac{\mu_n}{1-\mu_n}\right) = \beta_0 + \beta_{Prior} \ln\left(\frac{\mu_{n-1}}{1-\mu_{n-1}}\right) + \beta_{LR} \ln(LR_n) + \epsilon. \tag{3}$$

As common in the literature, we recode 0 guesses as 0.01 and 1 guesses as 0.99 to ensure that equation (3) is well-defined.

[24]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.

	(1) IV	(2) IV
β_{Prior}	0.9508****	0.9217****
	(0.0109)	(0.0376)
β_{LR}	0.2512****	0.0720
	(0.0338)	(0.0472)
β_{Prior} × Test score		0.0055
		(0.0066)
Test score		0.0127***
		(0.0049)
β_{LR} × Test score		0.0504****
		(0.0142)
Constant	-0.022****	-0.0719***
	(0.0068)	(0.0262)
Observations	20160	20160

Session-clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

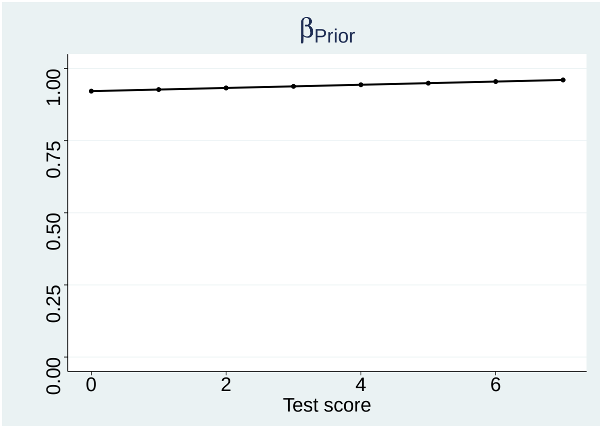
Table 10: **IV estimation results for Equation 3.** Column (1) shows that subjects display underinference and base rate neglect. Column (2) shows that subjects of higher cognitive ability show less underinference.

One potential issue with the OLS estimation of Equation 3 is that prior beliefs are formed 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 overcome it, we compute the Bayesian posterior beliefs for each observation and use $\ln\left(\frac{Bayes_{n-1}}{1-Bayes_{n-1}}\right)$ as an instrument for $\ln\left(\frac{\mu_{n-1}}{1-\mu_{n-1}}\right)$.

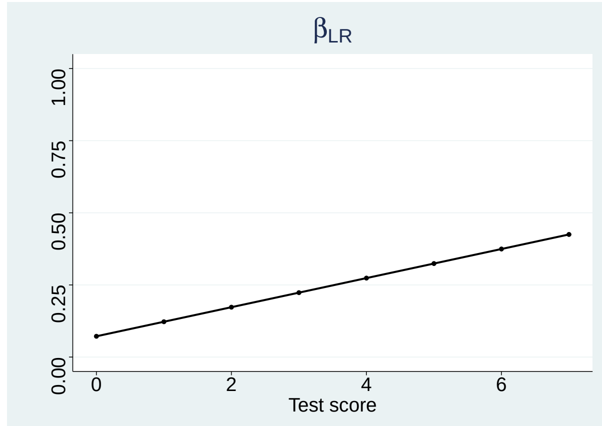
The IV estimates are reported in the first column of Table 10. That both β_{Prior} and β_{LR} are significant (largest $P < 0.001$) suggests that subjects are responsive to both prior and new information. Moreover, both coefficients are significantly less than one (largest $P < 0.001$), which suggests 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 10 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 subjects with higher cognitive ability show similar levels of base rate neglect ($P = 0.402$) but less underinference than those with lower cognitive ability ($P < 0.001$). Figure 5 shows the effect of test score on both coefficients.

One interpretation for the smaller 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



(a) Prior.

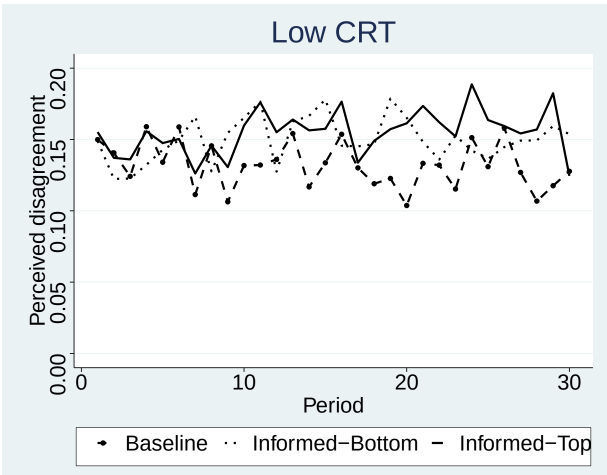


(b) Signal.

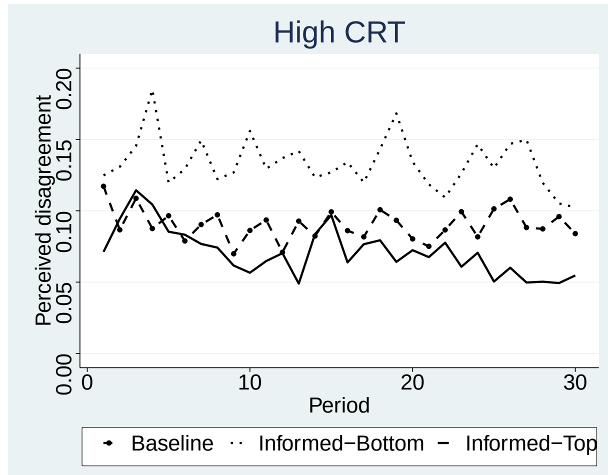
Figure 5: **Average marginal effects by test score.** (a) β^{Prior} is similar across subjects with different cognitive ability. (b) β_{LR} increases with cognitive ability.

subjects' response times when reporting first-order beliefs which the literature has interpreted as indicative of cognitive effort (Rubinstein, 2007, 2016; Enke et al., 2020). Contrary to this hypothesis, we find that subjects with low test scores spent more time to form first-order beliefs compared to high types ($P = 0.01$ in an OLS regression of the response time on test score with session-clustered standard errors).

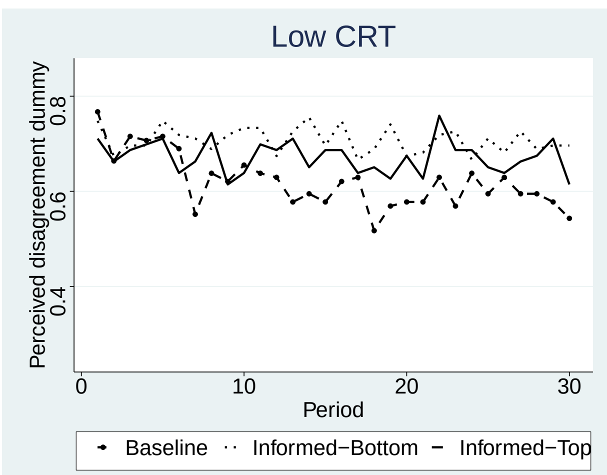
B Omitted Figures and Tables



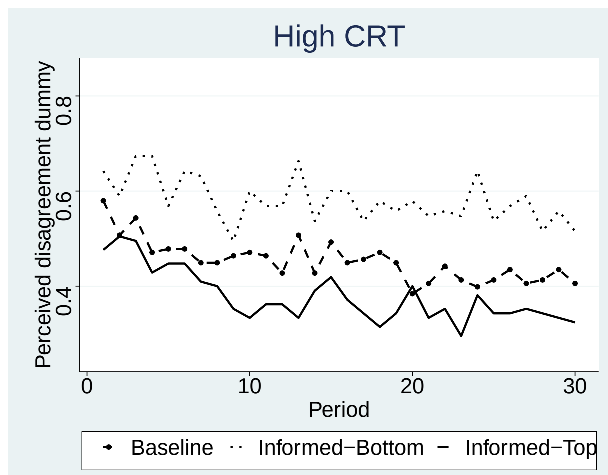
(a) Below median score.



(b) Above median score.



(c) Below median score.



(d) Above median score.

Figure 6: **Perceived disagreement and perceived disagreement dummy over time by cognitive type.**

	(1) Conservative	(2) Overinferred	(3) Polarized
Informed-Top	-0.1870	0.2034	-0.3242
	(0.2321)	(0.3947)	(0.3204)
Informed-Bottom	0.4894**	1.1617****	0.3936
	(0.2266)	(0.3503)	(0.2932)
Test Score	-0.1419****	-0.3069****	-0.2541****
	(0.0408)	(0.0657)	(0.0551)
Constant	0.2750	-0.8384**	-0.1736
	(0.2185)	(0.3341)	(0.2684)
Observations	622	622	622

Session-clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table 11: **Multinomial logit for type of subject.** The reference is the case of a subject whose modal behavior is to report second-order beliefs equal to first-order beliefs. A subject is categorized as *Conservative*, *Equal*, *Overinferred* or *Polarized* based on their modal behavior. Recall that we only consider observations in which a subject reported first-order beliefs not equal to 0.5. In order to avoid a subject from being categorized based only on a few number of observations, we require the modal behavior to be based on at least 8 (30/4) observations. Given these criteria, we are able to categorize 93% of subjects in our dataset (622 out of 672). Increasing the required number of observations to 10 would still categorize 88.7% of subjects with little differences for the estimation.