Social Exchange of Motivated Beliefs

Ryan Oprea Sevgi Yuksel

March, 2021

Abstract

We use laboratory experiments to study whether biases in beliefs grow more severe when people socially exchange these beliefs with one another. We elicit subjects' (naturally biased) beliefs about their relative performance in an IQ test and allow them to update these beliefs in real time. Part of the way through the task we give each subject access to the beliefs of a counterpart who performed similarly on the test and allow them both to observe the evolution of one another's beliefs. We find that subjects respond to one another's beliefs in a highly asymmetric way, causing a severe amplification of subjects' initial bias. We find no such patterns in response to objective public signals or in control treatments without social exchange or scope for motivated beliefs. We also provide evidence that the pattern is difficult to reconcile with Bayesianism and standard versions of confirmation bias. Overall, our results suggest that bias amplification is likely driven by "motivated assignment of accuracy" to others' beliefs: subjects selectively attribute higher informational value to social signals that reinforce their motivation.

Oprea: Economics Department, University of California, Santa Barbara, Santa Barbara, CA, 93106, roprea@gmail.com; Yuksel: Economics Department, University of California, Santa Barbara, Santa Barbara, CA, 93106, sevgi.yuksel@ucsb.edu.

1 Introduction

Accumulating evidence suggests that decision makers suffer from "motivated beliefs," believing propositions in part because they would prefer them to be true. Motivated beliefs (and the "motivated reasoning" that fuels it) can generate serious biases, and it is therefore important for social scientists and policy makers to understand what institutional settings make these biases more and less severe. One common intuition is that motivated biases can become especially severe when groups of people with similar motivation are given the opportunity to repeatedly socially exchange their beliefs with one another. For instance, many fear that proliferating online forums for exchanging motivated beliefs create "echo chamber" effects in which biases amplify over time, leading to polarization and extremism in political beliefs. Likewise, observers have perennially raised concerns that over-optimistic traders in the information rich setting of financial markets reinforce one another's optimism over time, leading to manias, bubbles and financial instability. Other examples may include group-think and tunnel vision in firms, mob behavior, inter-group conflict and the proliferation of conspiracy theories.

Does social exchange of motivated beliefs make beliefs more biased? The idea rests on a behavioral asymmetry that, to our knowledge, has not yet been examined or documented in the literature. In order for social exchange to worsen beliefs, relatively *less biased* parties must systematically update their beliefs *more* in response to social information (i.e. towards the beliefs of relatively *more biased* parties), than relatively *more biased* parties do (i.e. towards the beliefs of relatively *less biased* parties).

In this paper we report an experiment designed to look for evidence of this asymmetry and to take some first steps towards understanding its source. In our experiment, we pair subjects based on their score on an IQ test such that subjects within a pair either *both* have scores above the median or both have scores below the median. We then allow these subjects to exchange beliefs concerning a proposition we expect them to be motivated to believe is true: that they both scored above the median on the test. Subjects use a slider to reveal their probabilistic belief that they are in the high IQ group (using an incentive compatible mechanism) and they are given 45 seconds to adjust these beliefs. After 45 seconds we show subjects each other's sliders and allow them to see movements in one another's beliefs in real time for the rest of the task. Later, we give both subjects the same noisy (but objective) public signal on which group (high or low IQ) they are in and, again, observe how their beliefs co-evolve afterwards. We also report a control treatment in which subjects do not observe one another's beliefs and another in which subjects exchange beliefs that are unmotivated (i.e. that concern the outcome of a meaningless random variable).

This design reproduces what we take to be the key ingredients of the settings motivating our question. First, beliefs in our design concern a proposition that subjects are likely to have motivated beliefs over (their relative intelligence), capturing a crucial component of settings like political echo chambers or euphoric financial markets. Second, by matching subjects with similar IQ scores together, we ensure that subjects exchange beliefs about the same proposition and, further, have (weakly) aligned motivation over this proposition – another crucial characteristic of the main applied settings of interest. Finally, a key innovation in our design is that we allow subjects to exchange beliefs in a *real time* environment, adjusting their beliefs whenever and as often as they like. We believe this is important because, ex ante, it is unclear how the relative speed of adjustment of beliefs is endogenously determined and what role (if any) this timing might play in producing patterns of asymmetric adjustment. By having subjects adjust beliefs in real time, we allow subjects to adjust without the constraints present in a more conventional (but artificial) design, plausibly improving the ecological validity of our hypothesis test.

We find strong support for the asymmetry hypothesis. In line with the prior literature, our subjects hold biased, overconfident initial beliefs. These beliefs grow *significantly worse* (more biased) with social exchange. Average beliefs adjust systematically upwards (i.e. in an optimistic direction) over the course of social exchange, particularly the beliefs of subjects in the low IQ group for whom such movements necessarily decrease accuracy.[1]This effect is driven by the fact that subjects systematically adjust their beliefs upwards when their counterparts reveal relatively "optimistic" beliefs (that they are in the "high IQ" group), while the reverse does not occur: subjects do not *decrease* their beliefs when their counterparts reveal relatively "pessimistic" beliefs (that they are *not* in the "high IQ" group). By contrast, there is no systematic change in beliefs over time in control treatments without social exchange or without scope for motivated beliefs.

Though our main aim is simply to test the hypothesis that this asymmetry systematically *arises* in a relatively ecologically realistic setting, our data also provides some first clues as to the mechanism behind this asymmetry. Perhaps most diagnostic is that the effect is difficult to rationalize as a Bayesian phenomenon in our data even if we assume subjects are naive and fail to account for the bias in other's beliefs. Interestingly, we also find that this asymmetric adjustment is unlikely to be due to standard forms of confirmation bias, an inferential bias often associated with motivated beliefs. Confirmation bias would predict that subjects who initially hold pessimistic views (i.e. have priors that they are more likely to be in the low IQ group) will overweight social

[1]We also show in the Online Appendix that such changes imply a significant decrease in expected payoffs.

signals that reinforce their prior that they are in the low IQ group. We do not find evidence for this. In our data, the bias-amplifying effects of social exchange are due to subjects putting too much weight on beliefs they are motivated to agree with, *not* due to overweighting beliefs that match their priors.

One potentially important clue as to the mechanism driving this asymmetry is that our subjects do *not* overweight optimistic signals when these signals have a clear and unambiguous accuracy. When we provide subjects with unambiguous information (public signals with known, objective accuracy) they respond in a Bayesian way, putting equal (and proper) weight on both optimistic and pessimistic signals. This suggests that the asymmetry we find in social exchange may arise because subjects do not know, ex ante, how accurate their counterparts' beliefs are and this creates scope for motivated inference (motivated neglect of some information and overweighting of other information). In this account of the data, subjects respond to uncertainty about the accuracy of their counterparts by systematically assigning greater informational value (responding more strongly) to more optimistic (and more biased) counterparts that they would prefer to agree with than to pessimistic (less biased) counterparts they would rather not agree with. Further experiments designed prospectively to examine the relationship between signal ambiguity and social exchange of beliefs in more depth seems like an important next step in this agenda.

In addition to documenting the bias-amplifying effects of social exchange, we also investigate a policy instrument that is often discussed for correcting such bias: the provision of reliable, public information. After observing biases amplify as subjects exchange beliefs, we provide both subjects with the same noisy (but informative) public signal on whether they are in the high or low IQ group. We find that subjects respond quite strongly to these signals (including signals that contradict their motivation to believe they are in the high IQ group). Although we find evidence that social exchange has persistent effect on those most influenced by social exchange, we find that public signals cause aggregate beliefs to move closer to the beliefs of subjects in our control treatment who never experienced social exchange at all. That is, public signals are effective at correcting biases generated by social exchange, at least in the aggregate.

Our paper contributes to several literatures. First, there is a growing body of work studying how belief updating can diverge from Bayesian benchmarks.[2] This literature has documented asymmetric updating – biased over-weighting of good relative to bad news – in response to objective information, especially in settings in which there are motivated beliefs. However findings on this

[2]See, for example, Kahneman & Tversky (1972); Camerer (1998); Holt & Smith (2009); Grether (1980); Grether (1992) for important contributions.

point are mixed: while some papers (Eil & Rao (2011), Mobius et al. (2011), Grossman & Owens (2012)) find evidence of asymmetry, others (Ertac (2011); Gotthard-Real (2017); Buser et al. (2018); Barron (2020); Coutts (2019); Schwardmann & Van der Weele (2019)) either don't find an asymmetry or find asymmetry that also extends to non-Motivation settings.[3] Our paper brings new insights to this literature. To our knowledge, ours is the first paper to document, within a unified experimental design, contrasting evidence of (i) asymmetric updating of beliefs as a result of social exchange of motivated beliefs but (ii) no such asymmetry in response to objective, exogenously provided information (indeed we observe this systematic contrast within-subject and in response to events that occur mere seconds from one another). Our results suggest perceptions about the accuracy of an information source can play an important role in whether there is a differential response to good vs. bad news. An implication of our results is that even when subjects are provided objective information with known accuracy, asymmetric response might arise if subjects have difficulty understanding and internalizing this accuracy.

Our results suggesting that biased (non-Bayesian) updating may occur because of uncertainty about the accuracy of information in social exchange, connects with growing evidence in psychology and political science suggesting that subjects make motivated inferences about the trustworthiness of different information sources when updating beliefs on issues that they have preferences over. While experimental economists typically study settings in which information sources have objective (and known) accuracy, experiments in these adjacent literatures often study information processing in more natural settings where signal distributions conditional on each state can be more difficult to assess, potentially enabling the same "motivated assignment of accuracy" we document here.[4] Many voters evaluate political information to be of higher quality when it is reinforcing of their initial ideological position.[5] This tendency to assign higher accuracy to information sources selectively is

[3]Relatedly, Kuhnen (2015) finds that in responding to financial information, people react more to low outcomes in a loss domain relative to a gain positive domain leading to overly pessimistic beliefs.

[4]Belief updating when there is ambiguity (either with respect to priors or signals) is only starting to be studied experimentally in economics. Three recent papers focus on different aspects of updating behavior with ambiguity in abstract settings where subjects have no attachment to their prior: Liang (2019), Shishkin & Ortoleva (2019) and Larry & Halevy (2019). De Filippis et al. (2017) report results of a social learning experiment where updating behavior is consistent with a generalization of the Maximum Likelihood Updating rule when subjects are assumed to have ambiguous beliefs over the rationality of others. Asparouhova et al. (2015) study portfolio choice and resulting asset pricing when subjects switch from perceiving the environment as risky to ambiguous. To our knowledge, however, belief updating with ambiguity has not been directly studied yet in a context where subjects may hold motivated beliefs.

[5]Lord et al. (1979) is a seminal paper on biased information processing. See Tappin et al. (2020) Kahan (2015*a*,*b*), Taber & Lodge (2006) for more recent reviews of this literature.

observed in other contexts as well.[6] In many of these experiments it is difficult to nail down the behavioral mechanism driving selective accuracy assignment because, under many designs, such assignment is consistent with both motivated reasoning and Bayesianism (i.e. it *can* be Bayesian to assess an information source to be of higher quality, ex post, if it generates signals that reinforce one's prior beliefs). Thaler (2019) provides a recent contribution to the economics literature in this area using an experimental design that cleanly separates motivated reasoning from Bayesian updating. In a context where signals by construction have *no* informational value, subjects assess signals that better align with their motivated beliefs to be more trustworthy. In this paper, we highlight how "motivated assignment of accuracy" can be particularly hazardous when agents are able to socially exchange beliefs.

Relatedly, overconfidence, especially in beliefs about relative performance has also been an important focus of research. In addition to laboratory experiments[7] there is emerging evidence of overconfidence in field settings.[8] Recently, Zimmermann (2020) and Huffman et al. (2019) have studied the connection between persistent overconfidence and distortions in memory through selective recall when there is repeated feedback. Our contribution to this literature is to demonstrate how social exchange of beliefs can enable selective interpretation of information to reinforce and amplify overconfidence.

The growing literature on motivated beliefs and reasoning emphasizes that people can directly attach value to beliefs (beyond their instrumental value) because they fulfill important psychological and functional needs of the individual. Several different channels have been proposed in the literature on how beliefs could be motivated. Beliefs could directly enter the utility function if individuals value holding optimistic beliefs about themselves and about their future outcomes (Köszegi (2006) and Brunnermeier & Parker (2005)). Optimistic beliefs can also play a strategic

[6] Kahan et al. (2011) and Koehler (1993) discuss this issue in the context of how evidence is used in the evaluation of scientific theories. Jain & Maheswaran (2000) provides evidence in the context of a product choice experiment that when faced with ambiguous information subjects become skeptical of its value when it runs counter to the preference: Subjects who were exposed to evidence discrediting their optimality of their product choice scrutinized the evidence more and were more likely to provide counter arguments against the evidence. More recently, Peysakhovich & Karmarkar (2016) study how different factual statements are interpreted (coded by subjects as "favorable" vs. "unfavorable") in a design where this information may impact subjects' own financial prospects. They find that objectively favorable information is much more strongly integrated into financial evaluations than unfavorable information. Chaiken & Maheswaran (1994) and Russo et al. (1996) also discuss how information processing can be distorted further when there is ambiguity about the credibility of the source.

[7] See Charness & Dave (2017) for a review.

[8] See Barber & Odean (2001); Malmendier & Tate (2005); DellaVigna & Malmendier (2006); Oster et al. (2013); Scheier & Carver (1987); Hoffman & Burks (2020); Park & Santos-Pinto (2010).

role in helping present-biased agents overcome self-control problems (Bénabou & Tirole (2002)), or act as a social signal (Burks et al. (2013), Charness et al. (2018), Ewers & Zimmermann (2015), Schwardmann & Van der Weele (2019)).[9] Such considerations can lead individuals to process information in a way that is incompatible with Bayesianism, guaranteeing persistence of such optimistic beliefs. Our findings provide a new perspective on how biased beliefs can survive or even become more extreme through social interactions. In this way, our results also connect to the literature studying polarization of beliefs within groups. As demonstrated by Roux & Sobel (2015), polarization of beliefs within a group can be a simple consequence of information aggregation and hence need not be driven by behavioral decision making. However, such a model necessarily predicts groups to make better decisions (in our setting, form more accurate beliefs) relative to individuals. In contrast, Glaeser & Sunstein (2009) argues that social exchange of information can suffer from two distortions in aggregating beliefs: individual level information tends to be biased in favor of an initially dominant opinion, and individuals can fail to account for such bias when learning from others. In the presence of such distortions, groups exchanging information can become more extreme while also making inferior decisions relative to individuals. Our experiments highlights another related distortion that can take place in social interactions. While subjects in our experiment cannot misinterpret public signals with known accuracy, ambiguity with respect to the informational value in others' beliefs allow these beliefs to be processed in a motivated way: when social information increases optimism, it is highly valued; when it pushes beliefs in the opposite direction, it is disregarded.

The remainder of the paper is organized as follows. In Section 2 we describe our experimental design. In Section 3 we report our main experimental findings. In Section 4 we study mechanisms and compare results to several models in the literature. Finally, in Section 5 we conclude.

2 Experimental Design

We designed the experiment to measure the evolution of beliefs in a setting in which (i) subjects are prone to bias and (ii) can repeatedly exchange beliefs with other subjects who are likely to suffer the same bias. In Section 2.1 we describe the main task used in the experiment. In Section 2.2 we describe how this task is used in our sessions and how it varies across treatments. In Section 2.3 we explain how this design allows us to achieve the key inferential goals of our paper. Finally,

[9] See Bénabou & Tirole (2016) for a more extensive review of the literature. Bénabou (2013) presents a model of groupthink where agents with anticipatory preferences form beliefs in response to public signals and take actions that are payoff relevant for the group.

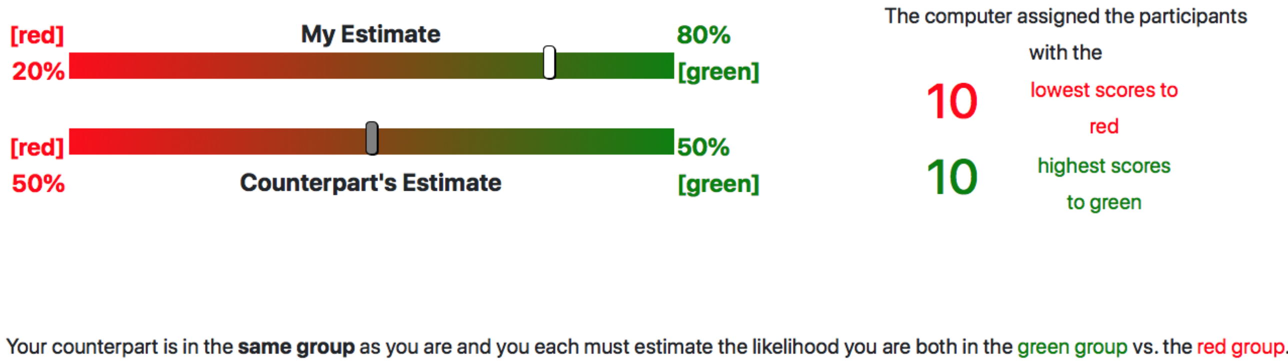

Figure 1: Screenshot from the Elicitation Task software. *Notes: Panel shows the screen for the main elicitation task in the Exchange-Motivation treatment.*

in Section 2.4, we describe details on the implementation of the experiment.

2.1 The Elicitation Task

In our main Elicitation Task, we assign subjects to one of two *groups* (either randomly or based on the results of an IQ test) and then allow subjects, in real time, to report their probabilistic beliefs about which group they have been assigned to for 180 seconds. We inform subjects of the manner in which groups were formed and allow them to continuously adjust their beliefs (at each second) using a slider. In Figure 1 we show a screenshot from the software. We paid subjects using a scoring rule applied to one second, randomly and uniformly selected from the 180 seconds. Thus, subjects have an incentive to report their beliefs truthfully at each moment of the task.

The Elicitation Task is divided into three phases whose exact character depends on treatment.

1. **Phase 1: (Initial beliefs, Seconds 1-44)** From seconds 1-44, the subject can adjust the slider continuously and observes only her own current slider position (belief).[10]

2. **Phase 2: (Interim beliefs, Seconds 45-89):** In second 45, subjects in Exchange treatments are paired with another subject who have been assigned to the same group and are shown this counterpart's slider. Figure 1 provides an example of a screenshot from such a treatment. In addition to seeing their own slider ("My Estimate") they see below it a second slider ("Counterpart's Estimate") showing their counterpart's current belief. This lower slider moves in real time throughout the remainder of the task as the counterpart adjusts her

[10] In this phase, the bottom slider in Figure 1, for example, would not be visible.

beliefs. In No Exchange treatments, the subject does not see any other subject's slider, i.e. phase 2 identical to phase 1.

3. **Phase 3: (Public Signal, Seconds 90-180)** In second 90 subjects are shown a *public signal with known accuracy* in all treatments. Specifically they are shown a test result indicating which group they are in and are aware that the test is accurate with 75% chance. In the Exchange treatment, subjects continue to see their counterpart's beliefs (observe the counterpart's slider) throughout this phase. Moreover, in the Exchange treatments, subjects are aware that they receive the exact same signal as their counterpart.

In addition to varying whether subjects exchange beliefs in phases 2 and 3 (Exchange vs. No Exchange), we also varied the basis on which subjects are assigned to a group. In the Motivation treatments, subjects are assigned based on their *relative* performance on an IQ test deployed at the beginning of the session. Specifically, subjects in all sessions are assigned 10 Raven's matrices and incentivized to complete them correctly (as described below). Subjects are assigned to either a high IQ group (framed as the green group) or low IQ group (framed as the red group) based on their performance relative to the median score. All sessions consisted of 20 subjects meaning that the subjects with the top (bottom) ten scores are in the high (low) IQ group. The right side of the panel in Figure 1 provides an example of how subjects were reminded of the group assignment rule in the Motivation treatments. In the No Motivation treatment, assignment to groups is random.

2.2 Session and Treatment Design

All sessions consisted of five parts.

1. **Raven's Matrices:** Subjects were shown ten Raven's Progressive Matrices (including a range from relatively easy to relatively difficult matrices) and were given 75 seconds to complete each one. If a Raven's matrix was selected for payment subjects received $10 if they answered it correctly. In Motivation treatments, scores on these matrices were used to determine group assignment.[11]

2. **Practice #1:** Subjects participated in an Elicitation task with No Exchange and with random group assignment (with a probability 0.6 of assignment to one group and 0.4 to the

[11] Subjects were not informed about how their performance on this part might impact later parts of the experiment.

other). The purpose of this task (and the task in part 3) is to familiarize subjects with the software and the task.[12]

3. **Practice #2:** Identical to part 2, except that we set a probability of 0.7 of assignment to one group and 0.3 of assignment to the other.

4. **Main Elicitation:** Subjects participated in an Elicitation task that varied with treatment as described below.

5. **Survey:** We asked subjects a number of survey questions including additional cognitive tests, beliefs about others, gender and major.

We ran the experiment using a between-subjects design consisting of three treatments. The treatments differed only in part 4.[13] In the Exchange-Motivation (E-M) treatment (our main condition), in part 4, subjects were assigned to groups based on their scores in part 1 (Motivation) and observed one another's beliefs (slider positions) in phases 2 and 3 of the task (Exchange).[14]

Table 1: Summary of Treatments

	Exchange-Motivation (**E-M**) 5 sessions, 100 subjects	No Exchange-Motivation (**NE-M**) 3 sessions, 60 subjects	Exchange-No Motivation (**E-R**) 3 sessions, 60 subjects
Group Assignment	Based on IQ score	Based on IQ score	Random
Group Composition	10 in Green, 10 in Red	10 in Green, 10 in Red	14 in Green, 16 in Red
Phase 1	No social interaction	No social interaction	No social interaction
Phases 2 & 3	Beliefs public in pairs	No social interaction	Beliefs public in pairs

Treatments differ only in the main elicitation task in Part 4. These differences are described above.

We also ran two control treatments to aid in interpreting results from the Exchange-Motivation treatment. In the No Exchange-Motivation (NE-M) treatment, we assigned subjects to their group in part 4 based on part 1 scores (Motivation) but we did not allow subjects to observe one another's beliefs in phases 2 and 3 (No Exchange). In the Exchange-No Motivation (E-NM) treatment, in part 4, we assigned subjects randomly to their group (No Motivation) but allowed them to observe

[12]Groups were referred to differently between parts 2-4 to make it clear to subjects that the group assignments are independent across these tasks. In part 2, groups were coded to be either Turquoise or Purple; in part 3 they were Orange or Blue; and in part 4 (main elicitation task) they were Green or Red. We also varied the direction of the majority group on the slider between the parts 2 and 3.

[13]Each part was introduced as a surprise. Instructions relating to all phases of a part (for example, phases 1-3 of the main elicitation task) were read out loud before starting that part.

[14]Instructions for Exchange-Motivation treatment are included in the Online Appendix F.

one another's beliefs in phases 2 and 3 (Exchange).[15] In this treatment, for each subject, the probability of being in the Green (rather than the Red) group was 0.7.[16] Table 1 summarizes the differences between the treatments.

2.3 Understanding the Design

We designed the experiment to serve several inferential goals.

First, because our aim is to understand how pre-existing biases are amplified or moderated by social exchange of beliefs, it was important that our main treatment induce some degree of initial bias in beliefs. For our Motivation treatments we therefore assign subjects to groups based on their score on an IQ test, which has been shown to successfully generate biased, motivated beliefs (overconfidence) in a number of previous studies.[17] Throughout the paper, we refer to beliefs as being "biased" whenever subjective assessments on the likelihood of being in the high IQ group differ in aggregate from the objective likelihood.

Second, because our interest is in environments (e.g. echo chambers, financial markets) in which people exchange beliefs with others who share *similar motivation* it was important that we allowed for communication among subjects whose motivation, at least on average, are pointed in the same direction. For this reason, we paired each subject in our Motivation treatments with another subject who is in the same IQ group. We believe it is of great intellectual interest to consider other ways of pairing subjects (e.g. subjects in *opposite* groups or randomly) but these would not shed much light on our main question.[18] Nonetheless, the belief exchange methodology we introduce in our design can be extended to other matching protocols to ask a host of questions about the consequences of the social exchange of beliefs.

Third, in order to match the environments we are interested in, we wanted to allow subjects the opportunity to repeatedly exchange beliefs with one another and to mutually adjust these beliefs in potentially complex ways. One way to do this might have been to have subjects repeatedly

[15] Note that we required subjects in the No Motivation treatment to complete Raven's Matrices in part 1 even though we do not need this data. We did this in order to keep sessions identical in all respects except for part 4 (the main elicitation task).

[16] In parameterizing, we deviated from the uniform prior to make the updating in response to the public signal nontrivial. This value also allowed us to match the median probability subjects assigned (at the end of phase 1) to being in the high IQ group in the Motivation treatments.

[17] See Burks et al. (2013) and citations within.

[18] Note that regardless of the type of pairing used, observing another person's belief can always have informational value.

and simultaneously submit beliefs at a preset grid of dates. Instead we opted to allow subjects to change beliefs asynchronously in continuous time. We chose this course because the environments we are interested in also unfold in continuous time, with decision makers revising their beliefs at endogenous times. Ex ante, this endogenous timing might matter for our main question because the speed at which relatively optimistic or pessimistic members of a social exchange change their beliefs could potentially have a role in the eventual emergence of asymmetric adjustment of beliefs. Ex post, based on our data, we believe it is likely we would have found similar asymmetry in a more conventional lockstep design, and studying such designs in the future will be valuable for better understanding the belief formation process.[19]

Fourth, it was important that the design allow us to attach causality to the effect of introducing social exchange on subsequent beliefs. Two features of the design enable this. Within subject, we introduced phase 1 to our Elicitation tasks, giving subjects substantial time (45 seconds) to set and adjust initial beliefs before they are ever exposed to the beliefs of their counterparts (in the Exchange treatments). Doing this gives us a clear pre- and post- comparison on the rate and direction of change *within subject*. Between subject, we introduced the No Exchange-Motivation control treatment which allows us to observe a counterfactual on what subjects would have done (on average) after second 45 in phases 2 and 3 of the task. Identification of the effect of social exchange is achieved by comparing behavior in these phases between Exchange-Motivation and No Exchange-Motivation treatments.

Fifth, in order to deepen our understanding of the effects of social exchange it is important to study how beliefs formed in these settings respond to the introduction of external information. One reason for this is purely diagnostic; observing updating behavior in response to a public signal provides us with a benchmark on how subjects respond to information of known quality.[20] Another reason is policy related: a common remedy suggested in response to biased and extremist beliefs in general and echo chambers in particular is the provision of objective public information. Phase 3 of our Elicitation Task allows us to study whether biases in subjects' beliefs (potentially reinforced

[19]A potential direction for future research is to compare one-way communication with two-way communication. We focus on the latter here in part because we wanted to create scope for the sorts of feedback loops observers often hypothesize exist in echo chambers. As it turns out there isn't much evidence of this in our data, but this in itself is a useful piece of information that we wouldn't have gathered had we used a one-way protocol.

[20]For a more direct comparison to the effects of social exchange, a cleaner test would be to study control treatments in which the public signal happens at the 45th second (rather than the 90th). Because NE-M behavior is quite stable between the 45th and 90th seconds, we suspect our results would be robust to such a variation. Nonetheless, this may be valuable to study directly in follow-up work.

through social exchange) are corrected in response to objective (though noisy) public signals.[21]

Finally, it is important to understand the degree to which the effects of social exchange are generated by the possibility that there is private information embedded in subjects' stated beliefs. To rule out alternative channels for these effects (e.g. sheer imitation of suboptimal heuristics, generating Bayesian updating errors), we introduced a second control treatment – Exchange-No Motivation – which is designed to be identical to our main treatment but in which there is no scope for motivated reasoning or private information (i.e. the treatment features objective and common priors for all subjects).

2.4 Implementation Details

We ran the experiment in the EBEL laboratory at the University of California, Santa Barbara in September and October of 2019. We recruited 220 subjects from across the curriculum to participate in 11 sessions using the ORSEE recruiting software (Greiner (2015)). The experiment was conducted using software programmed by the authors in oTree. Exactly 20 subjects participated in each session and sessions lasted for 45-55 minutes.[22]

We randomly selected one of the five parts for payment at the end of the experiment and provided no other feedback to subjects beforehand. If part 1 or part 5 was selected for payment, we paid subjects $10 for submitting a correct answer on a randomly selected question from that part.[23] If parts 2-4 were selected for payment, we paid subjects $0 or $10 based on their belief in a randomly selected second (out of 180) using the Binarized Scoring Rule.[24] Including a $10 showup fee, subjects earn $17 on average in the experiment.

3 Main Results

We focus our analysis on Part 4 of the experiment and especially on the contrast between our main Exchange-Motivation (E-M) treatment and the No Exchange-Motivation (NE-M) treatment.

[21] Note also that the No Exchange-Motivation treatment provides us with a benchmark on how subjects respond to these public signals in the absence of any social exchange.

[22] Our dataset includes 5 sessions of E-M and 3 sessions of NE-M and E-NM.

[23] There were a few belief questions in part 5 that were incentivized in the same way as the belief questions in parts 2-4.

[24] The advantage of this over other traditional mechanisms is that it does not rely on risk neutrality to be incentive-compatible. We use the implementation outlined in Wilson & Vespa (2018); the method was first developed in Hossain & Okui (2013).

Occasionally we will contextualize further with reference to the Exchange-No Motivation (E-NM) treatment (analysis of this treatment can be found in Section 4.3 and analysis of behavior in the practice periods can be found in Online Appendix A). We report our results by describing behavior from each phase of the task, in turn.

Before doing so, we first make several points about the terminology we will use in the analysis. When we refer to *beliefs* we will be referring to the beliefs (slider positions) subjects submit about the likelihood that they are assigned to the high IQ group (e.g. a belief of 0.9 means that the subject believes she is 90% likely to be in the high IQ group).[25] We will also repeatedly separately analyze the belief of the relatively more *pessimistic* and relatively more *optimistic* members of a part 4 matched pair. We will code a subject as pessimistic if she holds a strictly lower belief than her counterpart at the end of phase 1 (in the 44th second, prior to observing her counterpart's belief) and optimistic if she holds a weakly higher belief. Importantly, our software "matched" subjects in the No Exchange treatment just as it did in the Exchange treatments, but didn't show subjects in the former case their counterparts' beliefs in phases 2 and 3. Because of this, we can also code subjects in the No Exchange treatment as optimistic or pessimistic, providing us a natural counterfactual for outcomes in the Exchange treatment.

3.1 Phase 1: Initial Beliefs

We begin by describing behavior in phase 1 (seconds 1-44), during which subjects cannot observe or exchange beliefs in any treatment. Figure 2 plots the evolution of beliefs across the first two phases (divided by a vertical gray line). Solid lines show the average belief from the Exchange-Motivation treatment and dashed lines from the No Exchange-Motivation treatment.

In conducting this analysis, it is important that we break up the data in two ways.

First, because our main interest is in the *asymmetry* of the response to social information, it is important that we separately track the beliefs of subjects whose initial beliefs (prior to social exchange) are *more* consonant with motivated beliefs (i.e. are initially more "optimistic") and those whose initial beliefs are *less* consonant (i.e. are initially more "pessimistic"). Doing this will allow us to directly examine the key asymmetry the experiment was designed to identify. In Figure 2, the vertically lower series (in red) in each panel shows average beliefs for pessimistic subjects (categorized at the 44th second, the end of phase 1) while the higher series (in blue) plots beliefs for optimistic subjects.

[25] In E-NM, reported beliefs will refer to subjects' beliefs about the likelihood of being in the majority group.

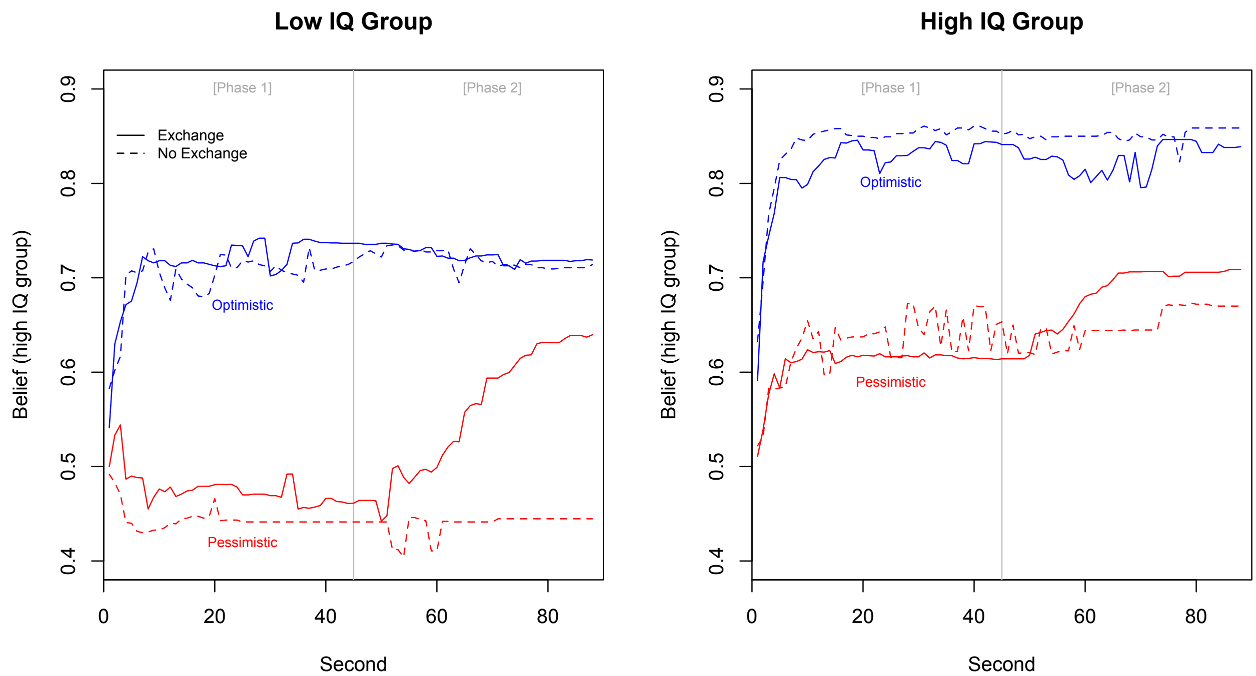

Figure 2: Evolution of average beliefs in Phases 1 and 2 for the E-M and NE-M treatments. *Notes: The left hand panel shows subjects assigned to the low IQ group and the right hand panel shows subjects assigned to the high IQ group. The upper series (in blue) plots data for relatively optimistic counterparts and the lower series (in red) relatively pessimistic counterparts.*

Second, because we are interested in the evolution of *biased* beliefs it is important that we separately examine subjects for whom optimistic beliefs are in fact biased (inaccurate) from those for whom it is not.[26] To do this, Figure 2 breaks the data into two panels: the left panel shows subjects assigned to the low IQ group and the right panel subjects assigned to the high IQ group (recall that subjects are not told which group they are assigned to). The subjects in the low IQ group (left hand panel) are those for whom optimistic beliefs are clearly biased and are therefore most relevant to our question.

Focusing first on phase 1, we highlight several key points. First, behavior quickly stabilizes and holds relatively constant throughout the phase. In the last 10 second of the phase, 85% of subjects do not change their beliefs at all and the average difference between highest and lowest belief given by a subject during this interval is only 2 percentage points.

Second, subjects who performed better on the Raven's task (high IQ group panel) correctly hold higher beliefs than subjects who did not (low IQ group panel): combining optimistic and

[26]Here and throughout, we refer to beliefs as "biased" whenever average elicited beliefs on the likelihood of being in the high IQ group differ in aggregate from the objective likelihood.

pessimistic subjects, high IQ group subjects hold beliefs that are 14 percentage points higher than low IQ group subjects ($p < 0.01$). Nonetheless, subjects in the low IQ group believe, on average, incorrectly, that the likelihood that they are in the high IQ group is greater than 50% ($p < 0.01$). All tests reported in the text (unless stated otherwise) are based on probit (for binary variables) or linear (for continuous variables) regressions clustered either at subject level (phase 1) or the pair level (phases 2 and 3). These regressions are reported in Online appendix B, C and D.[27] Combining optimistic and pessimistic subjects, the mean (median) belief of subjects in the low IQ group at the 44th second is 0.59 (0.65), while the mean (median) belief of subjects in the high IQ group at this point is 0.74 (0.75).

Third, there is no evidence of pre-trend in these series, indicating that subjects in Exchange-Motivation do not adjust their beliefs strategically in anticipation of mutual revelation of beliefs in phase 2. Likewise, comparing dashed and solid lines in the Figure, there is no meaningful difference between average initial beliefs in the Exchange versus the No Exchange treatment. Beliefs at the 44the second are not statistically different between the two treatments.[28] The equivalence of beliefs between the two treatments by end of phase 1 suggest that subjects are not (at least) preemptively adopting more confident beliefs in order to send positive signals to others about their own skill.[29]

Result 1 Subjects' beliefs stabilize in phase 1 of the experiment. On average, subjects in the low IQ group hold upward biased initial beliefs about their assignment to the high IQ group. Beliefs in phase 1 are not statistically different between the Exchange and No Exchange treatment.

[27]When we run regressions on subset of the data—either dividing by whether subjects are optimistic or pessimistic relative to their counterpart, or by whether they are in the high or low group—we estimate standard errors using bootstrapping.

[28]This is unsurprising since the two treatments are identical in phase 1, but is important to highlight in order to contextualize results in phase 2, below.

[29]It is possible that such social signaling incentives kick in in phase 2, even if subjects fail to anticipate them in phase 1. Such a hypothesis could be tested directly in future work by comparing one-way communication with two-way communication. A treatment with one-way communication would allow for information transmission while shutting down social signaling concerns (for at least one side.) However, post-experiment survey responses are consistent, instead, with our main interpretation that subjects learn from one anothers' beliefs: 71% of subjects in E-M state that their counterparts beliefs had some influence on their own beliefs; 46% of subjects state they were influenced "a moderate amount" or "quite a lot".

3.2 Phase 2: Social Exchange

The right hand side of each panel in Figure 2 shows the evolution of beliefs in phase 2 (from the 44th to the 90th second). The dashed lines (NE-M) in each panel, unsurprisingly, show evidence of no convergence between optimistic and pessimistic beliefs in the absence of social exchange. By contrast, the solid lines (E-M) come significantly closer to one another over the course of phase 2. This convergence in beliefs between relatively optimistic and pessimistic counterparts within a pair is substantial: the difference in beliefs in the Exchange treatment are half as large at the end of phase 2 (89th second) as at the end of phase 1 (44th second).[30] However convergence is not complete: on average the absolute difference between subjects' beliefs is 14 percentage points at the end of phase 2 (statistically different from zero with $p < 0.01$).[31]

Result 2 Social exchange causes subjects' beliefs to partially converge but subjects show evidence of persistent disagreement.

Our main finding is that this convergence in beliefs is highly asymmetric in two ways: (1) the adjustment is systematically *upwards*, driven by relatively pessimistic subjects moving towards their relatively more optimistic counterparts (notably, in a direction consistent with motivated beliefs) while there is no systematic downward adjustment by optimistic subjects; (2) this upward adjustment in beliefs is strongest for those in the low IQ group (i.e. for those subjects for whom such movements necessarily decrease accuracy). In the aggregate (i.e. aggregation across IQ group and optimism/pessimism), beliefs become significantly more optimistic ($p < 0.01$) as a result of social exchange.

These patterns are evident in Figure 2: The strongest upward adjustment in beliefs is observed for pessimistic subjects in the low IQ group *for whom upward adjustment is a biased response that necessarily reduces accuracy.* The increase in beliefs for these subjects from the end of phase 1 to the end of phase 2 is 17 percentage points! There is also a significant, but less pronounced (9 percentage point) upward adjustment by pessimistic subjects in the high IQ group. By contrast, there is no significant adjustment downwards for optimistic subjects in either group.[32]

[30]Within a pair, the average absolute value difference between beliefs decrease by 14 percentage points ($p < 0.01$) from the 44th second to the 89th second in E-M.

[31]This contrasts with standard game theory "agree to disagree" results (Aumann (1976)) on how rational agents cannot persistently disagree in the absence of any informational differences. In our design, subjects' beliefs at the end of phase 1 should be a sufficient statistic for any private information they hold up to that point, so non-convergence of beliefs is indicative of disagreement within a pair about the accuracy of initial beliefs.

[32]See Online Appendix C for regression analysis establishing these patterns statistically. Our main results focus on

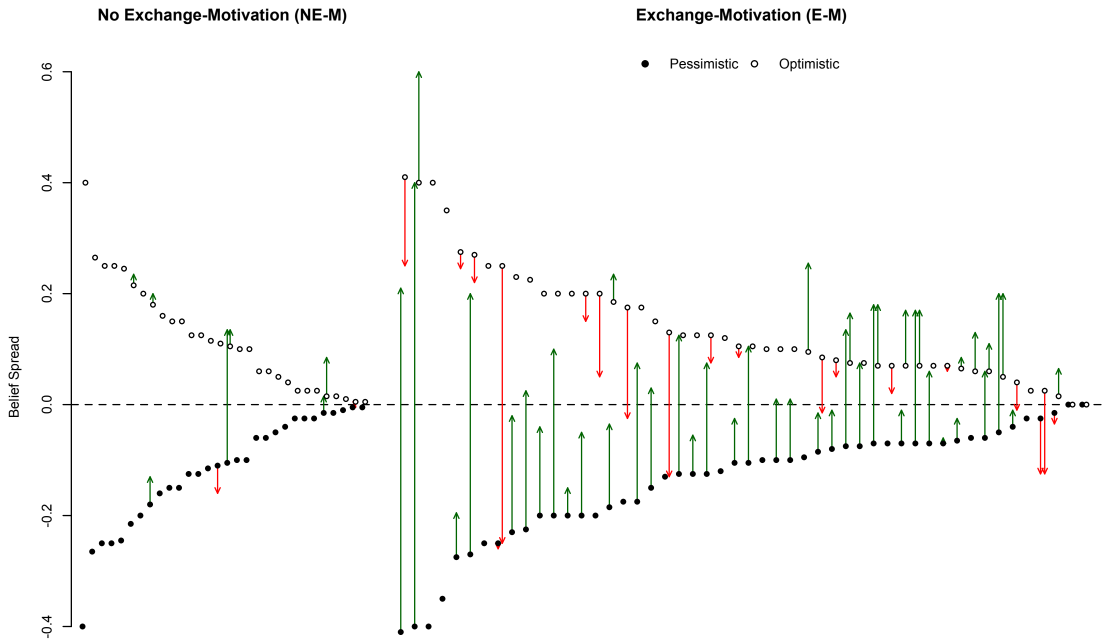

Figure 3: Initial belief differences and Phase 2 belief adjustments for all matched pairs in the E-M and NE-M treatments. *Notes: Dots show beliefs at the 44th second for the optimistic (hollow dots) and pessimistic (solid dots) normalized symmetrically around zero. Arrows show adjustments in beliefs between the end of phase 1 and the end of phase 2. Green arrows show upward and red arrows downward adjustments.*

Result 3 Beliefs adjust systematically upwards as a result of social exchange, particularly for those in the low IQ group. Social exchange thus worsens bias on average. No similar effects occur in the absence of social interaction.

Figure 3 provides more detailed evidence on adjustment over the course of phase 2. The Figure plots each matched pair in the dataset (each horizontal position is slotted for a different pair), broken up by treatment. Solid dots plot the belief of the more pessimistic subject in the pair and hollow dots the belief of the more optimistic subject at the 44th second. These dots are normalized to be symmetrically arranged around zero and ordered from the largest initial gap in beliefs to smallest from left to right. Finally, the Figure includes arrows emanating from each dot showing the movement of that subject's belief between the end of phase 1 and the end of phase 2 (where there is no arrow there is no adjustment). Upward adjustments are plotted in green and downward

the change (in beliefs) from the end of phase 1 to the end of phase 2. We also establish that the results are robust to alternative definitions including using the last ten second of phase 1 vs. phase 2 (rather than the very last second) or, instead of focusing on relative pessimism/optimism, looking at whether one's counterpart has a higher or lower belief than median belief.

adjustments in red.

In E-M, most (71% of) pessimistic subjects adjust upward to some degree (by contrast only 2% adjust downward) and these adjustments are typically large: the average pessimistic subject adjusts 53% of the way towards her optimistic counterpart's belief which is statistically different from 0 ($p < 0.01$). By contrast, downward (red) arrows are comparatively rare, even among optimistic subjects. Indeed, optimistic subjects are nearly as likely to adjust upward (away from their counterpart) as downward (19% vs. 32%, which are statistically not different using a t-test) and on average they adjust only 9% of the way towards their pessimistic counterpart which is not statistically different from 0. This pattern is *caused* by the social exchange of beliefs: in the No Exchange-Motivation treatment relatively few (13% of) subjects adjust at all over the course of phase 2 with no systematic direction to the change.

We summarize our results from this section below:

Result 4 The majority of pessimistic subjects adjust significantly upwards towards their more optimistic counterpart in Exchange-Motivation. By contrast only a minority of optimistic subjects adjust downwards towards their pessimistic counterpart (and almost as many adjust away from their pessimistic counterpart).

3.3 Phase 3: The Effect of Public Signals

At the beginning of phase 3, we provide subjects with a noisy binary signal about their group assignment that is accurate with probability 0.75. Throughout the analysis we denote the signal s as h or l (high vs. low, indicative of being in the high IQ vs. low IQ group).[33] Figure 4 plots time series of beliefs from phase 2 and 3, broken up by signal type. The upper (blue) series in each panel plots subjects who (at the beginning of phase 3) receive an h signal and the lower (red) series plots subjects who receive an l signal.

Because our main question concerns the *corrective effects* of public signals on beliefs worsened by social exchange, it is important to separately examine the impact of signals on pessimistic subjects (who, recall, change their beliefs in response to social exchange) from optimistic subjects (who overwhelmingly do not respond to social exchange at all). For this reason, Figure 4 is divided into two panels: the left panel plots beliefs for relatively pessimistic subjects (within a pair), the right for relatively optimistic subjects.

[33]Note that in the experimental implementation a signal was referred to as a test result and matched the color of the assigned group with probability 0.75.

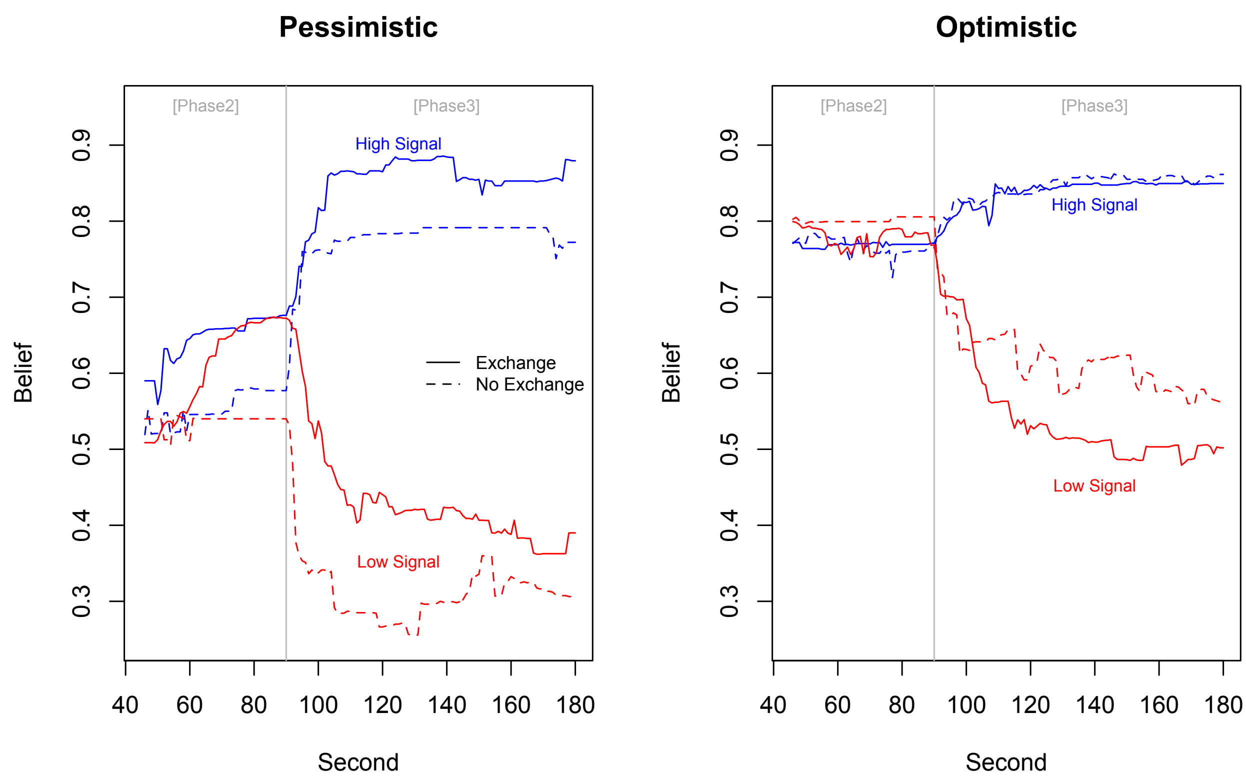

Figure 4: Evolution of average beliefs in phases 2 and 3 for the E-M and NE-M treatments. *Notes: Panels break up data by relatively pessimistic and optimistic counterparts.The upper series (in blue) plots data for subjects who receive an h signal (indicative of being in the high IQ group) of at the beginning of phase 3 and the lower series (in red) for subjects who receive an l signal (indicative of being in the low IQ group) .*

Focusing on Phase 3 (to the right of the vertical gray line) we observe strong responses to these signals. Signals indicative of being in the high IQ group cause subjects' beliefs to substantially increase and signals indicative of being in the low IQ group significantly lower beliefs relative to those beliefs held prior to social exchange in phase 2. In aggregate, by the end of phase 3 (180th second) beliefs of those subjects who received the h signal are 41% higher than those who received the l signal in the Exchange-Motivation treatment. A similar pattern occurs in the No Exchange-Motivation treatment where the corresponding value is 38%. In Section 4 we examine this updating in more detail and provide statistical evidence that responses to these two types of signals are symmetric and, indeed, are collectively consistent with a Bayesian response to the public signal.

Result 5 Subjects respond strongly to public information.

Do these public signals eliminate the additional bias created by social exchange in phase 2, or does this amplification of bias persist? Each panel of Figure 4 includes data from the No Exchange-Motivation treatment to provide some insight on this. We highlight several important points about the contrast between data from the two treatments.

First, pessimistic subjects in Exchange-Motivation – the subjects whose beliefs changed most radically during phase 2 – show some evidence of a persistent bias relative to their counterparts in No Exchange-Motivation. Solid lines are persistently above dashed lines for both high and low signals. Second, this effect is removed or reversed for optimistic subjects. We provide detailed statistical analysis in Online Appendix D on treatment differences. To summarize, we find that, conditioning on the public signal, social exchange has no significant impact on the overall change in beliefs from the end of phase 1 to end of phase 3. This suggests that the public signal is effective to a large extent at removing the initial impact of social exchange. However, looking more closely at the data, we find some evidence that at least for certain groups of subjects, social exchange may have persistent effects. For example, conditioning on the public signal, we find that pessimistic subjects in E-M increase their beliefs from the 44 to the 180th second by an additional 10 percentage points relative to those in NE-M ($p = 0.03$). This effect is mainly driven by the behavior of pessimistic subjects in low IQ group for whom this necessarily leads to a decline in accuracy – precisely the group for which we found the sharpest impact of social exchange in phase 2. A natural interpretation of this pattern is that initially pessimistic subjects retain (at least partially) biases generated by social exchange even after receiving public signals.[34,35]

Result 6 Public information mostly corrects for the amplification of bias generated by social exchange of beliefs. However, there is some evidence to suggest that social exchange may have persistent effects for some subjects.

4 Mechanism

In this section, we examine the evolution of beliefs in more detail to better understand the mechanism by which social exchange worsens bias. Our conclusion based on this analysis is that this

[34] Figure 4 is suggestive of some optimistic subjects responding more strongly to low signals than they otherwise would have due to their history of observing the beliefs of their pessimistic counterparts in phase 2, but given heterogeneity in priors we are not powered to make statistical statements.

[35] We also replicate these results in Online Appendix D focusing on the change in average beliefs from the last ten seconds of phase 1 to the last ten seconds of phase 3.

effect likely arises because, in social exchange, (i) others' beliefs contain *some* information (as documented in Section 3.1) but (ii) subjects have little basis for determining the accuracy of this information (beliefs in our data are noisy and biased). Under this interpretation, subjects fill in missing information about the accuracy of others' beliefs by attributing greater accuracy to beliefs that reinforce their own motivation (e.g. support the belief that have relatively high IQ). This "motivated assignment of accuracy to others' beliefs," in turn, causes subjects to respond much more strongly to optimistic counterparts' beliefs than pessimistic counterpart's beliefs, worsening bias in social exchange. This mechanism is conceptually related to but ultimately different from confirmation bias: in our data we do not find evidence that subjects with pessimistic priors update more strongly in response to pessimistic than optimistic signals, as predicted by standard version of confirmation bias.

To make this case, we first show that it is difficult to rationalize asymmetric responses to social information as Bayesian behavior in our data. In Section 4.1, we show that relative to both naive and sophisticated Bayesian benchmarks, subjects fail to update in a Bayesian fashion and put dramatically more weight on others' beliefs when these beliefs point in the direction of one's motivation. By contrast, in Section 4.2 we show that when we remove ingredient (ii) by giving subjects *unambiguous* public signals, subjects respond in a symmetric and (broadly) Bayesian fashion. Likewise, in Section 4.3 we use data from our Exchange-No Motivation treatment to remove ingredient (i), eliminating private information from social exchange and again find broadly Bayesian behavior, ruling out the alternative possibility that some other framing bias associated with our experimental design generates these effects. Finally, we show in Section 4.4 that this pattern is difficult to reconcile with the predictions of standard formulations of confirmation bias. In Section 5 we argue that this series of findings leaves "motivated assignment of accuracy" as the most likely mechanism driving our results.

To conduct this analysis, we make use of the following property of Bayes' rule.[36] For any signal s (indicating the high IQ vs. the low IQ group), an implication of Bayes' rule (written in log form) is that

$$\log\left(\frac{p}{1-p}\right) = \log\left(\frac{p_0}{1-p_0}\right) + \log\left(\lambda_s\right) \tag{1}$$

where p_0 is the prior, p is the posterior, and λ_s is the likelihood ratio of observing s if the subject is indeed in the high IQ vs. the low IQ group.[37] To test the extent to which updating behavior

[36]The regression approach we use was first introduced in Grether (1980), and since has been used in many empirical papers studying updating behavior. The regressions estimate responsiveness to different types of signals (controlling for their informational value) and the prior (controlling for its strength).

[37]For example, given that p and p_0 represent beliefs in being in the high IQ group, for a binary signal with precision

across different phases of the main elicitation task is consistent with this property, we estimate regressions of the following form:

$$\log\left(\frac{p}{1-p}\right) = \alpha \log\left(\frac{p_0}{1-p_0}\right) + \beta_h \mathbf{1}_{(\lambda_s>1)} \log(\lambda_s) + \beta_l \mathbf{1}_{(\lambda_s \leq 1)} \log(\lambda_s) \quad (2)$$

where $\mathbf{1}_{(\cdot)}$ is an indicator function, α measures responsiveness to the prior and β_h and β_l separately measure responsiveness to signals that are indicative of being in the high IQ group vs. the low IQ group.[38] λ_s is the likelihood ratio of observing signal s, which will vary in the specifications below. If updating behavior is consistent with Bayes' rule in the aggregate (given assumptions on λ_s), we would expect $\alpha = \beta_h = \beta_l = 1$. More generally, if subjects respond symmetrically to *good* vs. *bad* signals (signals which respectively increase or decrease optimism), we would expect $\beta_h = \beta_l$ in regression estimates. Any asymmetry with respect to these estimates implies that updating behavior suffers from a systematic directional bias.

4.1 Updating In Response to Social Signals

We first analyze how subjects respond to the social signal they receive at the beginning of phase 2, by examining how subjects change their beliefs from the 44th second (just prior to observing their counterpart's belief) to the 89th second (just prior to receiving the public signal). The subject's beliefs in the 44th second gives us her prior belief p_0, and her belief in the 89th second gives us her posterior belief p. The likelihood ratio attributed to each signal (λ_s) depends on how sophisticated subjects are in assigning accuracy to initial beliefs, so we consider two natural benchmarks. First we consider the possibility that subjects are *sophisticated*, assigning correct (at the population level) accuracy to their counterparts' beliefs. Second, we consider the possibility that subject are *naive*, presupposing their counterpart's belief to be unbiased and thus believing their probabilisitic belief about the likelihood of being in the high IQ group to be accurate. We show that relative to either benchmark, subjects are asymmetric in their updating: they are more responsive to signals that are indicative of being in the high IQ group relative to other signals that move beliefs in the opposite direction.

In order to form a benchmark for sophistication, we assume that beliefs (at end of phase 1 in the 44th second) are distributed according to a normal distribution truncated to the unit interval with mean and variance parameters (μ, σ). Importantly, we allow these parameters to vary with

---

q, $\lambda_h = \frac{1}{\lambda_l} = \frac{q}{1-q}$.

[38] With a binary signal, β_h and β_l simply measure responsiveness to the h and l signals.

the state, i.e. we allow for the distribution of beliefs to be different for those subjects in the high IQ and the low IQ groups. Using maximum likelihood, we estimate (μ_h, σ_h) and (μ_l, σ_l) (for high and low IQ groups) that best fit the beliefs observed in the data.[39] The best fit distributions for each group differ sufficiently for social exchange of beliefs to enable significant information transmission, with counterparts' high (low) beliefs probabilistically indicating membership in the high (low) IQ group. These distributions provide us with a benchmark on how much a sophisticated Bayesian agent would update her beliefs about being in the high IQ group based on observing her counterpart to have any belief $\tilde{p} \in [0, 1]$ at the end of phase 1. That is, for any $\tilde{p}$ the two distributions imply a likelihood ratio $\lambda_{\tilde{p}}$.[40] We then estimate the following variation on specification (2) adapted to this case:

$$\log\left(\frac{p}{1-p}\right) = \alpha \log\left(\frac{p_0}{1-p_0}\right) + \beta_h \mathbf{1}_{(\lambda_{\tilde{p}}>1)} \log(\lambda_{\tilde{p}}) + \beta_l \mathbf{1}_{(\lambda_{\tilde{p}}\leq 1)} \log(\lambda_{\tilde{p}}) \qquad (3)$$

where $\lambda_{\tilde{p}}$ being greater (less than) 1 denotes that the social signal is indicative of being in the high (low) IQ group according to the sophisticated benchmark.

In order to form a benchmark for naive Bayesians, we simply assume that subjects take their counterpart's beliefs at "face value." We assume that each subject considers her counterpart's belief $\tilde{p}$ as a signal on the state of the world (being in the high vs. low IQ group) with a likelihood ratio of $\lambda_{\tilde{p}} = \frac{\tilde{p}}{1-\tilde{p}}$. That is, if a subject's counterpart reports his belief of being in the high group as $\tilde{p}$, the subject considers this report to be generated with probability $\tilde{p}$ if they are in the high IQ group and $1-\tilde{p}$ if they are in the low IQ group. We then estimate specification (3) using this naive benchmark.

Results for both specifications are reported in Table 2. As discussed above, in both cases, if updating behavior is consistent with Bayes' rule in the aggregate (given assumptions on $\lambda_{\tilde{p}}$ based on either the sophisticated or the naive benchmark), we would expect $\alpha = \beta_h = \beta_l = 1$. Furthermore, to provide us with a baseline benchmark, we also estimate the same regression for subjects in the No-Exchange- Motivated treatment where (since subjects cannot observe their counterparts'

[39] We plot the distributions corresponding to these best fits (as well as the empirical distributions in the data) in Online Appendix E.1. According to the maximum likelihood estimation the following two truncated normals best fit the distribution of beliefs at the end of Phase 1: For the High Group: $(\mu, \sigma) = (0.91, 0.29)$; For the Low Group: $(\mu, \sigma) = (1, 0.64)$. Since the distributions are truncated to the unit interval, the implied mean is 0.73 and 0.59 for the High and Low Group, respectively.

[40] Let (μ_h, σ_h) and (μ_l, σ_l) characterize the truncated normal distributions in the high and low state. $\lambda_{\tilde{p}} := \frac{\frac{1}{k_h\sigma_h}\phi\left(\frac{\tilde{p}-\mu_h}{\sigma_h}\right)}{\frac{1}{k_l\sigma_l}\phi\left(\frac{\tilde{p}-\mu_l}{\sigma_l}\right)}$ where for $i \in \{h, l\}$, $k_i = \Phi\left(\frac{1-\mu_i}{\sigma_i}\right) - \Phi\left(\frac{0-\mu_i}{\sigma_i}\right)$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution.

Table 2: OLS Estimation (Dependent Variable: Log Posterior Odds Ratio End of Phase 2)

	Sophisticated Benchmark		*Naive Benchmark*	
	Exchange	No-Exchange	Exchange	No-Exchange
α	0.721***	1.034***	0.773***	1.039***
	(0.117)	(0.0235)	(0.110)	(0.0215)
β_h	2.808***	0.212	0.363***	0.0290
	(0.693)	(0.239)	(0.0832)	(0.0354)
β_l	0.0655	0.0292	0.0334	0.0406
	(0.0776)	(0.0256)	(0.0589)	(0.0334)
$H_0 : \beta_h = \beta_l$	0.001	0.462	0.007	0.075
Observations	100	60	100	60

Estimation results are on updating from end of phase 1 to 2.

Standard errors in parentheses; clustered at the pair level in E-M and subject level in NE-M.

The second to last row shows p-values associated with testing $\beta_h = \beta_l$.

***1%, **5%, *10% significance. Data from Motivation treatments.

beliefs) we expect $\alpha = 1$ and $\beta_h = \beta_l = 0$.

Results for both the sophisticated and the naive benchmarks show that under Exchange subjects are highly responsive to "high" signals that are indicative of being in the high IQ group (their counterpart having a high belief), but do not respond at all to "low" signals that are indicative of being in the low IQ group (their counterpart having a low belief): β_h is large and significant in both cases, while β_l is small and insignificant. α is slightly smaller than 1 (indicating slight under weighting of the prior — partial base rate neglect), and both β_h and β_l are far from 1 in both cases.[41] These patterns suggest highly non-Bayesian responses to signals. Note also that the robustness of this result with respect to benchmark suggests that the asymmetric response pattern we observe is not driven by a specific choice concerning how to categorize beliefs as "high" vs. "low" signals.[42]

[41] For both benchmarks, under Exchange, α is significantly different from 1 ($p < 0.05$), β_l is not different from 0 and β_h is different than 1 ($p < 0.05$ in sophisticated benchmark and $p < 0.01$ in naive benchmark) .

[42] In the Online Appendix, we also estimate an alternative sophisticated benchmark where we code a subject's counterpart's belief as high vs. low depending on whether it is strictly higher than the median belief. We then look at the empirical likelihood ratio of observing each type of signal in the high vs. low IQ groups to evaluate the informational value. This creates another benchmark (one in which $\tilde{p}$ is reduced to a binary signal) for how a sophisticated Bayesian should update beliefs from end of phase 1 to phase 2. Regression analysis based on this model

In particular the near-zero estimate for β_l suggests subjects attribute virtually no informational value to observations of others' beliefs that are indicative of being in the low IQ group. By contrast, estimates for β_h greater than 1 under the sophisticated benchmark and less than 1 in the naive benchmark suggest that subjects attribute much higher accuracy to their counterparts' initial beliefs than the actual informational content warrants, but less accuracy than a fully naive subject who takes others' beliefs at 'face value' would assign. In the No Exchange case, the observation that the estimated value of α is near 1 and the values for the β coefficients are small and insignificant acts as a useful sanity check: subjects mostly hold onto their priors and do not act as if they have access to their counterparts' beliefs when they do not see these beliefs.

Result 7 The effect of social exchange on beliefs cannot be explained as (i) a Bayesian response to correct beliefs about the accuracy of initial beliefs, nor can it be explained as (ii) a Bayesian response to a naive failure to account for bias in others' beliefs.

Of course, these specifications are hardly exhaustive and it is natural to wonder whether *any* model of the distribution of beliefs (regardless of how different from the actual distribution) could rationalize behavior under a Bayesian lens. In Online Appendix E.3, we perform such an exercise. We show that we would need to attribute highly flexible priors to subjects that differ fundamentally from the actual empirical distribution of elicited beliefs in order to rationalize the asymmetry as Bayesian. In this exercise, we first estimate the distribution of prior beliefs[43] for which observed beliefs at the end of phase 2 (after social exchange) would be most consistent with Bayesian updating.[44] This subjective distribution looks substantially different from the actual distribution of elicited beliefs: to account for the asymmetry in updating, the rationalizing distributions require high log likelihood ratios (in absolute value) for high realizations of $\tilde{p}$ (belief of one's counterpart) and counterfactually low log likelihood ratios (in absolute value) for low realizations of $\tilde{p}$. This is achieved in the following way. According to the estimated distributions, subjects believe others—and particularly those in the low IQ group—have highly pessimistic beliefs. This implies even relatively optimistic beliefs (which includes almost all empirical observations of $\tilde{p}$) to be highly informative about being in the high IQ group, justifying an overall strong positive response to such beliefs. Of course, precisely in this respect, the subjective distribution differs substantially from

reveals the same pattern observed in Table 2 where subjects are highly responsive to high signals (indicative of being in the high IQ group) and dismissive of low signals (indicative of being in the low IQ group).

[43]Focusing attention on the normal distribution truncated to the unit interval, we estimate mean and variance parameters conditional on each state.

[44]We do a least squares estimation, details of which are provided in the Online Appendix. Figures and regressions relating to this exercise are also included there.

the actual empirical distribution.[45,46]

4.2 Updating In Response to Public Signals

Table 3: OLS Estimation (Dependent Variable: Log Posterior Odds Ratio End of Phase 3)

	No-Exchange	Exchange
α	0.857***	0.798***
	(0.119)	(0.123)
β_h	0.999***	1.295***
	(0.205)	(0.325)
β_l	1.030***	1.254***
	(0.191)	(0.234)
$H_0 : \beta_h = \beta_l$	0.914	0.927
Observations	60	100

Estimation results are on updating from end of phase 2 to 3.

Standard errors in parentheses, clustered at the pair level in E-M and subject level in NE-M.

The second to last row shows p-values associated with testing $\beta_h = \beta_l$.

***1%, **5%, *10% significance. Data from Motivation treatments.

It is possible that the asymmetry documented in the previous subsection has little to do with social exchange per se, but is instead a generic biased response to information in settings with motivated beliefs. We can test this possibility by studying the updating that occurs in response to public signals between seconds 89 (the end of phase 2, just before subjects receive the public signal) and 189 (the end of phase 3). That is, considering subjects' beliefs at the 89th second as their prior, we study how they update their beliefs in response to the public signal. Here the nature of signals ($s = h$ or $s = l$) and the accuracy associated with these signals are clear and so (unlike in the case of social exchange) there is no ambiguity on λ_h and λ_l.[47] We therefore can directly estimate equation (2) using objective values of λ_h and λ_l; where Bayesianism predicts $\alpha = \beta_h = \beta_l = 1$ for

[45] Gentzkow & Shapiro (2006) study the effect of uncertainty about the trustworthiness of news sources on information acquisition strategies of voters and consequently distort incentives in the media market to generate biased reporting. More recently, Gentzkow et al. (2018) use a dynamic model to study how uncertainty about trustworthiness of new sources can combine with systematically biased feedback to generate persistent belief errors even among Bayesian agents.

[46] See footnote 53 and 54 for further discussion about what types of non-Bayesian responses—when there is uncertainty about informational value of a signal—can generate asymmetric updating.

[47] $\lambda_h = \frac{1}{\lambda_l} = \frac{0.75}{1-0.75} = 3$ where 0.75 is the accuracy of the public signal.

both treatments.[48]

Results reported in Table 3 (again for both No-Exchange and Exchange subjects) reject this hypothesis by revealing a sharp contrast between the response to public signals and the response to social signals. Most importantly, the asymmetry in response to high versus low signals disappears under objective public signals: estimates for β_h and β_l are nearly identical in each case and statistically not different. Moreover, both estimates are consistent with Bayesian updating. Subjects in the No-Exchange treatment give us the clearest picture, uncontaminated by the effects of prior social exchange in phase 2 and estimates here are almost exactly 1, consistent with Bayesian updating. Estimates from the Exchange treatment hint at symmetric over-response to both types of signals (perhaps revealing a continued effect of social exchange), but we cannot statistically distinguish these estimates from 1.[49]

Crucially, subjects do not respond to public signals significantly differently in the Exchange treatment than they do in the No-Exchange treatment. This suggests that social exchange only influences beliefs when there is some potential for *private information* to be revealed in the exchange (as in phase 2). In phase 3 the only new information comes from the public signal, and there we find no effect of social exchange. This suggests that subjects (quite reasonably) absorb only the information contained in the public signal itself, ignoring their counterparts' interpretation of that signal.

Result 8 In contrast to social exchange, there is no asymmetry in response to objective public signals that increase versus decrease optimism (reinforce or undermine the subject's motivation). Moreover, the response to public signals is highly consistent with Bayesian updating.

4.3 Social Exchange in the Absence of Private Information and Motivation

In order to rule out the possibility that other, superficial elements of the decision environment or the physical layout of the display are drivers of our main findings, we ran our Exchange-No Motivation (E-NM) control treatment in which group assignment is random rather than driven by performance in the IQ test. Specifically, out of the 20 subjects in each session of this treatment 14

[48]Note that by comparing updating behavior with social exchange to updating behavior in response to the public signal, we are also controlling for other behavioral forces that could act as confounding factors. For example, it could be that subjects with more extreme beliefs (possibly due to over/under-confidence or censoring of beliefs) generally update differently (controlling for the strength of their prior). If such patterns exist, they should not play a differential role in response to social-exchange.

[49]Estimates for α in both treatments are also not significantly different from 1.

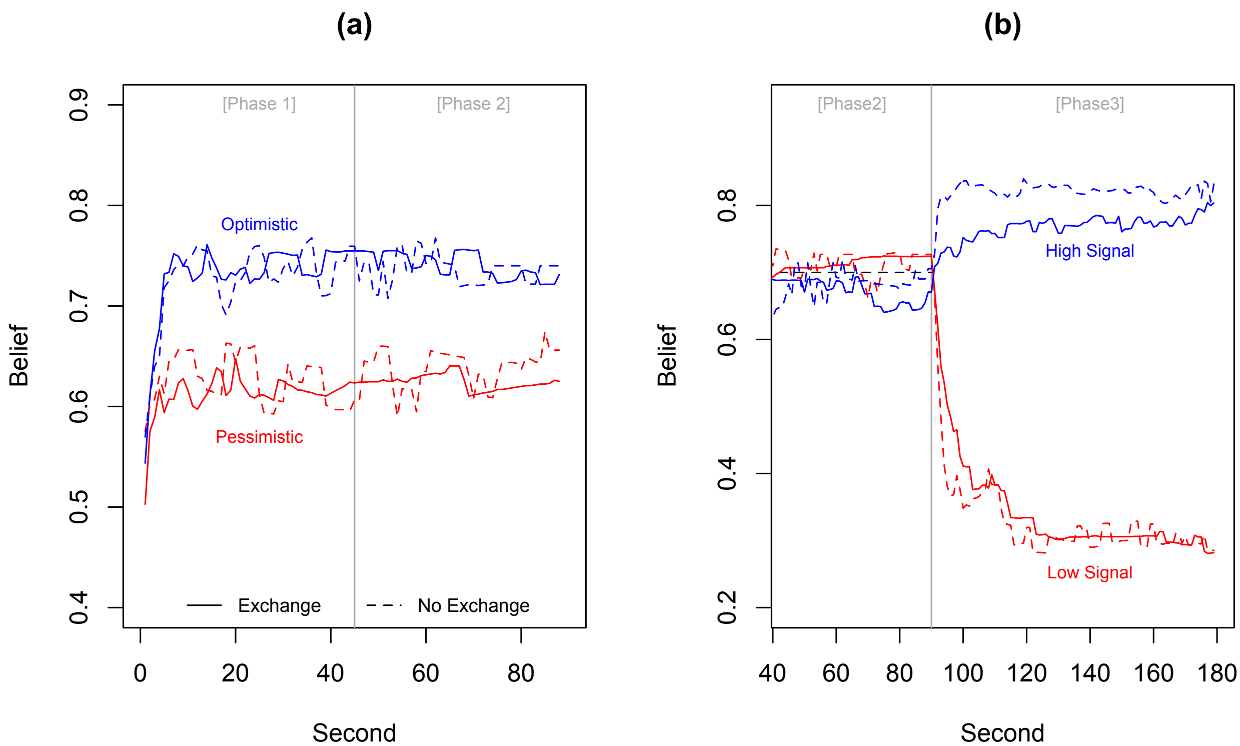

Figure 5: Behavior in No Motivation Elicitation Tasks. *Notes: Panel (a) replicates Figure 2 aggregating over subjects in both groups and (b) replicates Figure 4 aggregating over optimistic and pessimistic subjects.*

were randomly assigned into the green group and 6 were assigned to the red group.[50] Note that in E-NM the prior associated with the likelihood of being in either group is exogenously determined and clearly communicated to the subjects. Thus, this treatment theoretically removes the possibility that others' initial beliefs can have any informational value beyond the commonly known prior. Since group assignment is random, this treatment also removes any scope for motivation. Hence, this treatment provides us a with a baseline measure of how much social exchange impacts beliefs in the absence of *both* of the key ingredients we hypothesize as necessary for bias amplification, at the beginning of this section.

Figure 5 replicates Figures 2 and 4 using E-NM data (solid lines) from the main elicitation task. For a No Exchange control, we use the second practice period (from the same treatment) which features exactly the same probabilities but has no social exchange of belief (in this practice period, in contrast to the main elicitation task, beliefs in the more-likely-to-occur group are submitted by moving the slider leftward rather than rightward). Again, we highlight several observations.

[50]Recall that we calibrated the prior of being in the green group in E-NM to 70% to match average end-of-phase-1 beliefs in the E-M treatment to facilitate such comparisons.

First, and most importantly, in panel (a) it is clear that there is no systematic upward movement in phase 2 beliefs in the Exchange condition. Subjects, on average, continue with their old beliefs. This is consistent with subjects considering their counterparts' beliefs to carry no meaningful information. More importantly, in contrast to our Motivation treatments, there is no difference between the Exchange and No-Exchange conditions. Second, and for completeness, in panel (b) we observe strong responses to public signals as in the Motivation treatments. There is a stronger response to low signals.[51] We observe slightly lower beliefs for subjects receiving the high signal under Exchange than No Exchange though this seems to be driven by (random) differences in submitted beliefs prior to phase 3 between the Exchange and No Exchange conditions.

Result 9 Social exchange does not have a significant effect on beliefs when there is no private information or scope for motivation.

4.4 Confirmation Bias

Finally, we consider the possibility that the asymmetric response to optimistic versus pessimistic social information in the E-M treatment is driven by *confirmation bias*: the tendency to put more weight on signals that reinforce one's prior relative to signals that undermine one's prior. After all, as we show above, subjects' prior beliefs (as measured at the end of phase 1), in aggregate, reveal overconfidence and therefore tend to already lean in the direction of the subject's motivation. In our experiment, confirmation bias would predict that subjects attribute higher information value to others' beliefs when these beliefs are directionally consistent with the subject's own prior.[52,53,54]

[51] In Online Appendix E.4 we provide further analysis on this and show evidence of significant base-rate neglect under both conditions.

[52] There is a growing body of work that studies different forms of confirmation bias. See Nickerson (1998), Klayman & Ha (1987), Skov & Sherman (1986), Slowiaczek et al. (1992) Baron et al. (1988) and Rabin & Schrag (1999) for a review of the literature. Studies in this literature tend to find that people are drawn to uninformative evidence consistent with the initial hypothesis at the expense of informative evidence that could falsify the hypothesis. Most recently, Charness et al. (2020) find evidence of confirmation bias in how subjects choose between information sources in an abstract setting where scope for motivated beliefs is minimized.

[53] Several recent papers propose theoretical foundations for confirmation bias. Operating in the world of uncertainty, Koçak (2018) and Cheng & Hsiaw (2019) study models in which agents update their beliefs when there is uncertainty about the credibility of the information source. They show that simple deviations from Bayes' rule (sequential updating about the state and the credibility leading to "double dippping" of the data instead of joint updating) can generate patterns consistent with confirmation bias.

[54] If there is ambiguity about the credibility of the information source, maximum likelihood updating can also generate behavior consistent with confirmation bias. Examples include Gilboa & Schmeidler (1993), Fryer Jr et al.

To investigate this possibility, we separate out subjects in the E-M treatment into two groups based on the direction of their prior at the end of phase 1 (i.e. based on whether subjects believe they are strictly more likely to be in the high vs. the low IQ group) and estimate equation (3) under both the sophisticated and naive benchmarks. If subjects display confirmation bias, we would expect *opposite* asymmetries in updating for these two groups of subjects. Specifically, we would expect those with initially high beliefs ($p_0 > 0.5$) to put more weight on their counterpart's beliefs when their counterpart also has high beliefs (indicative of being in the high IQ group) and the opposite pattern for those subjects who start with low beliefs ($p_0 < 0.5$).

Results, reported in Table 4 are not supportive of this explanation. Subjects with initially high beliefs ($p_0 > 0.5$) put more weight on "high" signals (i.e. on beliefs that indicate membership in the high IQ group). We observe a similar pattern among subjects who start with low initial beliefs ($p_0 < 0.5$): low-prior subjects also tend to adjust in a biased fashion in the direction of their motivation rather than in the direction of their low prior beliefs. However, we have few subjects (only 16) with initially low beliefs ($p_0 < 0.5$) and so are underpowered to make statistical statements (we cannot reject the hypothesis that these subjects put equal weight on both signals). Nonetheless, there is no compelling evidence such subjects put more weight on "low" signals (i.e. on beliefs that indicate membership in the high IQ group) as required by standard formulations of confirmation bias.

We summarize our results as follows:

Result 10 The asymmetric updating we observe in social exchange is not supportive of confirmation bias. There is no evidence that subjects with initially low priors respond more strongly to signals that are in the same direction as their prior that decrease optimism (undermine subject's motivation).

5 Discussion

We find that when subjects have motivated beliefs, social exchange of these beliefs worsens biases on average over time. In revising their own beliefs, subjects tend to put substantial weight on the beliefs of their counterpart when it reinforces their motivation, but dismiss the beliefs of their

(2019), Baliga et al. (2013), Suleymanov (2018). Given that it is difficult to form precise beliefs about the informational content in others beliefs, it is reasonable to expect to such mechanisms to potentially play a role in how beliefs evolve through social exchange in our experiment.

Table 4: OLS Estimation (Dependent Variable: Log Posterior Odds Ratio End of Phase 2)

	Sophisticated Benchmark		*Naive Benchmark*	
	$p_0 < 0.5$	$p_0 > 0.5$	$p_0 < 0.5$	$p_0 > 0.5$
α	0.509	0.786***	0.526	0.873***
	(0.712)	(0.0842)	(0.622)	(0.0723)
β_h	3.476	2.527***	1.036	0.328***
	(3.706)	(0.554)	(0.947)	(0.0680)
β_l	0.543	0.0142	0.584	0.0221
	(0.630)	(0.0593)	(0.663)	(0.0479)
$H_0 : \beta_h = \beta_l$	0.656	0.000	0.756	0.001
Observations	16	76	16	76

Estimation results are on updating from end of phase 1 to 2.

Bootstrapped standard errors (clustered at the pair level) in parentheses.

The second to last row shows p-values associated with testing $\beta_h = \beta_l$.

***1%, **5%, *10% significance. Data from E-M.

counterpart when it does not. The data provides several important pieces of evidence on why this occurs.

First, it is difficult to explain these results as a Bayesian response to a reasonable set of prior beliefs. Under multiple ways of specifying such beliefs, we still find evidence of a strong asymmetric response to others' beliefs.

Second, we find no evidence that this is a generic biased response to information that reinforces or opposes one's motivation. Subjects respond to unambiguous, objective public signals in a symmetric and Bayesian manner.

Third, this result cannot be a simple response to the passage of time. In control treatments without social exchange of beliefs we find no systematic adjustment in beliefs at all.

Fourth, we rule out the possibility that the results are due to framing or implementation details. In a control treatment in which subjects lack private information (or motivation) we see no such effects of the social exchange of beliefs.

Fifth, the effect is unlikely to be driven by standard versions of confirmation bias. Subjects who initially hold directionally pessimistic beliefs do not seem to asymmetrically respond more to others' beliefs when these beliefs align with their prior. The asymmetry, whenever it is observed, is apparently driven by the subject's motivation, not their prior beliefs.

Aggregating this evidence, then, we find that subjects selectively attribute high informational value to social signals that reinforce their motivation while unwarrantedly dismissing information that would move their beliefs in the opposite direction. Importantly, they do this precisely when (and only when) the informational value of signals is difficult to gauge, as it is under social exchange. We hypothesize therefore that the combination of uncertainty about information quality (often present in the social exchange of beliefs) with motivation to hold some beliefs over others produces a particularly ripe environment for the emergence and amplification of bias. When subjects are forced to assign accuracy to signals themselves due to ambiguity, they tend to use their motivation to guide this assignment, causing them to overweight information supporting beliefs they'd prefer to hold. In settings in which agents with similar motivation exchange beliefs, these paired forces – ambiguity/uncertainty and motivation – will tend to cause biases latent in the population to amplify. We call this mechanism "motivated assignment of accuracy," and we speculate that it may be an important channel, distinct from confirmation bias, by which motivated beliefs may produce and worsen biases.[55]

Responses to the post experiment survey provide some further evidence that "motivated assignment of accuracy" may be important for generating asymmetric updating in our data. As documented in Section 3.1, while subjects are predominantly overconfident in their relative ability, when asked to indicate the likelihood that they scored higher on the IQ test than their counterpart, those subjects who are matched with a counterpart with strictly higher beliefs (at the 44th second) are less confident (16 percentage points) than those who are matched with a counterpart with lower beliefs. (The difference is significant, $p < 0.01$). Moreover, subjects who are less confident about their IQ score (relative to their counterpart) are more likely to report that they were influenced by their counterpart's belief in Phase 2, and more likely to have increased their beliefs in this phase.[56]

This mechanism, if borne out in future research, may give us important clues as to the conditions under which biases can amplify in social settings and of how policies might be constructed to combat this effect. Ambiguity about the value of socially transmitted information, and motivation to hold some beliefs over others, seem to be two crucial ingredients. While we may be able to do little

[55] This mechanism relies on ambiguity about the accuracy of counterparts' beliefs, but it is natural to wonder whether ambiguous priors may also play a role here. For instance, subject's introspection about how well her performance is on the IQ questions might be consistent with multiple priors. Given degrees of freedom in how ambiguity in priors is introduced, it is not possible to completely rule out this channel. However, highly Bayesian responses to the public signal suggests multiple priors likely play a limited role in driving the amplification of bias.

[56] See regression analysis in Online Appendix E.5 for details.

about motivated beliefs (motivation may simply be part of how human belief-formation works) we may be able to take policy steps to reduce ambiguities in socially transmitted information. In our setting, subjects lack a social history with their counterparts, which likely creates scope for ambiguity over the accuracy of others' beliefs. The relative anonymity of many online forums and of prices in financial markets, may mirror this lack of social history and create similar scope for ambiguity about accuracy. It may be possible to design alternative institutions in which social histories create clearer track records and less opportunity for discussants to fill in uncertainties about accuracies with their own motivation. Similarly, we find a strong corrective effect of objective public information on socially amplified biases. Producing and protecting sources of information with shared and objective credibility may have similar bias-mitigating effects in important naturally occurring settings.

References

Asparouhova, E., Bossaerts, P., Eguia, J. & Zame, W. (2015), 'Asset pricing and asymmetric reasoning', *Journal of Political Economy* **123**(1), 66–122.

Aumann, R. J. (1976), 'Agreeing to disagree', *The annals of statistics* pp. 1236–1239.

Baliga, S., Hanany, E. & Klibanoff, P. (2013), 'Polarization and ambiguity', *American Economic Review* **103**(7), 3071–83.

Barber, B. M. & Odean, T. (2001), 'Boys will be boys: Gender, overconfidence, and common stock investment', *The quarterly journal of economics* **116**(1), 261–292.

Baron, J., Beattie, J. & Hershey, J. C. (1988), 'Heuristics and biases in diagnostic reasoning: Ii. congruence, information, and certainty', *Organizational Behavior and Human Decision Processes* **42**(1), 88–110.

Barron, K. (2020), 'Belief updating: does the ?good-news, bad-news? asymmetry extend to purely financial domains?', *Experimental Economics* pp. 1–28.

Bénabou, R. (2013), 'Groupthink: Collective delusions in organizations and markets', *Review of Economic Studies* **80**(2), 429–462.

Bénabou, R. & Tirole, J. (2002), 'Self-confidence and personal motivation', *The Quarterly Journal of Economics* **117**(3), 871–915.

Bénabou, R. & Tirole, J. (2016), 'Mindful economics: The production, consumption, and value of beliefs', *Journal of Economic Perspectives* **30**(3), 141–64.

Brunnermeier, M. K. & Parker, J. A. (2005), 'Optimal expectations', *American Economic Review* **95**(4), 1092–1118.

Burks, S. V., Carpenter, J. P., Goette, L. & Rustichini, A. (2013), 'Overconfidence and social signalling', *Review of Economic Studies* **80**(3), 949–983.

Buser, T., Gerhards, L. & Van Der Weele, J. (2018), 'Responsiveness to feedback as a personal trait', *Journal of Risk and Uncertainty* **56**(2), 165–192.

Camerer, C. (1998), 'Bounded rationality in individual decision making', *Experimental economics* **1**(2), 163–183.

Chaiken, S. & Maheswaran, D. (1994), ‘Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment.’, *Journal of personality and social psychology* **66**(3), 460.

Charness, G. & Dave, C. (2017), ‘Confirmation bias with motivated beliefs’, *Games and Economic Behavior* **104**, 1–23.

Charness, G., Oprea, R. & Yuksel, S. (2020), ‘Biases over biased information structures: Confirmation, contradiction and certainty seeking behavior in the laboratory’, *Journal of European Economic Association* .

Charness, G., Rustichini, A. & Van de Ven, J. (2018), ‘Self-confidence and strategic behavior’, *Experimental Economics* **21**(1), 72–98.

Cheng, I.-H. & Hsiaw, A. (2019), ‘Distrust in experts and the origins of disagreement’, *SSRN Scholarly Paper ID* **2864563**.

Coutts, A. (2019), ‘Good news and bad news are still news: Experimental evidence on belief updating’, *Experimental Economics* **22**(2), 369–395.

De Filippis, R., Guarino, A., Jehiel, P. & Kitagawa, T. (2017), Updating ambiguous beliefs in a social learning experiment, Technical report, cemmap working paper.

DellaVigna, S. & Malmendier, U. (2006), ‘Paying not to go to the gym’, *american economic Review* **96**(3), 694–719.

Eil, D. & Rao, J. M. (2011), ‘The good news-bad news effect: asymmetric processing of objective information about yourself’, *American Economic Journal: Microeconomics* **3**(2), 114–38.

Ertac, S. (2011), ‘Does self-relevance affect information processing? experimental evidence on the response to performance and non-performance feedback’, *Journal of Economic Behavior & Organization* **80**(3), 532–545.

Ewers, M. & Zimmermann, F. (2015), ‘Image and misreporting’, *Journal of the European Economic Association* **13**(2), 363–380.

Fryer Jr, R. G., Harms, P. & Jackson, M. O. (2019), ‘Updating beliefs when evidence is open to interpretation: Implications for bias and polarization’, *Journal of the European Economic Association* **17**(5), 1470–1501.

Gentzkow, M. & Shapiro, J. M. (2006), 'Media bias and reputation', *Journal of political Economy* **114**(2), 280–316.

Gentzkow, M., Wong, M. B. & Zhang, A. T. (2018), Ideological bias and trust in information sources, Technical report, Working Paper.

Gilboa, I. & Schmeidler, D. (1993), 'Updating ambiguous beliefs', *Journal of economic theory* **59**(1), 33–49.

Glaeser, E. L. & Sunstein, C. R. (2009), 'Extremism and social learning', *Journal of Legal Analysis* **1**(1), 263–324.

Gotthard-Real, A. (2017), 'Desirability and information processing: An experimental study', *Economics Letters* **152**, 96–99.

Greiner, B. (2015), 'Subject pool recruitment procedures: organizing experiments with orsee', *Journal of the Economic Science Association* **1**(1), 114–125.

Grether, D. M. (1980), 'Bayes rule as a descriptive model: The representativeness heuristic', *The Quarterly journal of economics* **95**(3), 537–557.

Grether, D. M. (1992), 'Testing bayes rule and the representativeness heuristic: Some experimental evidence', *Journal of Economic Behavior & Organization* **17**(1), 31–57.

Grossman, Z. & Owens, D. (2012), 'An unlucky feeling: Overconfidence and noisy feedback', *Journal of Economic Behavior & Organization* **84**(2), 510–524.

Hoffman, M. & Burks, S. V. (2020), 'Worker overconfidence: Field evidence and implications for employee turnover and firm profits', *Quantitative Economics* **11**(1), 315–348.

Holt, C. A. & Smith, A. M. (2009), 'An update on bayesian updating', *Journal of Economic Behavior & Organization* **69**(2), 125–134.

Hossain, T. & Okui, R. (2013), 'The binarized scoring rule', *Review of Economic Studies* **80**(3), 984–1001.

Huffman, D., Raymond, C. & Shvets, J. (2019), 'Persistent overconfidence and biased memory: Evidence from managers'.

Jain, S. P. & Maheswaran, D. (2000), 'Motivated reasoning: A depth-of-processing perspective', *Journal of Consumer Research* **26**(4), 358–371.

Kahan, D. M. (2015*a*), 'The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it', *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource* pp. 1–16.

Kahan, D. M. (2015*b*), 'The politically motivated reasoning paradigm, part 2: Unanswered questions', *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource* pp. 1–15.

Kahan, D. M., Jenkins-Smith, H. & Braman, D. (2011), 'Cultural cognition of scientific consensus', *Journal of risk research* **14**(2), 147–174.

Kahneman, D. & Tversky, A. (1972), 'Subjective probability: A judgment of representativeness', *Cognitive Psychology* **3**(3), 430–454.

Klayman, J. & Ha, Y.-W. (1987), 'Confirmation, disconfirmation, and information in hypothesis testing.', *Psychological review* **94**(2), 211.

Koçak, K. (2018), Sequential updating: A behavioral model of belief change, Technical report, Working Paper.

Koehler, J. J. (1993), 'The influence of prior beliefs on scientific judgments of evidence quality', *Organizational behavior and human decision processes* **56**(1), 28–55.

Köszegi, B. (2006), 'Ego utility, overconfidence, and task choice', *Journal of the European Economic Association* **4**(4), 673–707.

Kuhnen, C. M. (2015), 'Asymmetric learning from financial information', *The Journal of Finance* **70**(5), 2029–2062.

Larry, E. & Halevy, Y. (2019), 'Hard-to-interpret signals'.

Liang, Y. (2019), 'Learning from unknown information sources', *Available at SSRN 3314789* .

Lord, C. G., Ross, L. & Lepper, M. R. (1979), 'Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence.', *Journal of personality and social psychology* **37**(11), 2098.

Malmendier, U. & Tate, G. (2005), 'Does overconfidence affect corporate investment? ceo overconfidence measures revisited', *European financial management* **11**(5), 649–659.

Mobius, M. M., Niederle, M., Niehaus, P. & Rosenblat, T. S. (2011), Managing self-confidence: Theory and experimental evidence, Technical report, National Bureau of Economic Research.

Nickerson, R. S. (1998), 'Confirmation bias: A ubiquitous phenomenon in many guises', *Review of general psychology* **2**(2), 175–220.

Oster, E., Shoulson, I. & Dorsey, E. (2013), 'Optimal expectations and limited medical testing: evidence from huntington disease', *American Economic Review* **103**(2), 804–30.

Park, Y. J. & Santos-Pinto, L. (2010), 'Overconfidence in tournaments: evidence from the field', *Theory and Decision* **69**(1), 143–166.

Peysakhovich, A. & Karmarkar, U. R. (2016), 'Asymmetric effects of favorable and unfavorable information on decision making under ambiguity', *Management Science* **62**(8), 2163–2178.

Rabin, M. & Schrag, J. L. (1999), 'First impressions matter: A model of confirmatory bias', *The Quarterly Journal of Economics* **114**(1), 37–82.

Roux, N. & Sobel, J. (2015), 'Group polarization in a model of information aggregation', *American Economic Journal: Microeconomics* **7**(4), 202–32.

Russo, J. E., Medvec, V. H. & Meloy, M. G. (1996), 'The distortion of information during decisions', *Organizational behavior and human decision processes* **66**(1), 102–110.

Scheier, M. E. & Carver, C. S. (1987), 'Dispositional optimism and physical well-being: The influence of generalized outcome expectancies on health', *Journal of personality* **55**(2), 169–210.

Schwardmann, P. & Van der Weele, J. (2019), 'Deception and self-deception', *Nature human behaviour* **3**(10), 1055–1061.

Shishkin, D. & Ortoleva, P. (2019), 'Ambiguous information and dilation: An experiment'.

Skov, R. B. & Sherman, S. J. (1986), 'Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation', *Journal of Experimental Social Psychology* **22**(2), 93–121.

Slowiaczek, L. M., Klayman, J., Sherman, S. J. & Skov, R. B. (1992), 'Information selection and use in hypothesis testing: What is a good question, and what is a good answer?', *Memory & Cognition* **20**(4), 392–405.

Suleymanov, E. (2018), 'Robust maximum likelihood updating'.

Taber, C. S. & Lodge, M. (2006), 'Motivated skepticism in the evaluation of political beliefs', *American journal of political science* **50**(3), 755–769.

Tappin, B. M., Pennycook, G. & Rand, D. G. (2020), 'Thinking clearly about causal inferences of politically motivated reasoning: Why paradigmatic study designs often undermine causal inference', *Current Opinion in Behavioral Sciences* **34**, 81–87.

Thaler, M. (2019), The "fake news" effect: An experiment on motivated reasoning and trust in news, Technical report.

Wilson, A. & Vespa, E. (2018), 'Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language', *Working paper* .

Zimmermann, F. (2020), 'The dynamics of motivated beliefs', *American Economic Review* **110**(2), 337–61.

Online Appendix for

Social Exchange of Motivated Beliefs

Ryan Oprea   Sevgi Yuksel

CONTENTS:

A Behavior in Parts 2 & 3

Bayesian benchmark: In Part 2, beliefs should be 0.60 at seconds 44 and 89, and should be 0.82 or 0.33 (conditional on $s = h$ or $s = l$) in second 180; in Part 3, beliefs should be 0.70 at seconds 44 and 89, and should be 0.88 or 0.44 (conditional on $s = h$ or $s = l$) in second 180.[57]

Table 5: OLS Estimation (Dependent Variable: Beliefs in Part 2)

	Second =44 All	Second =89 All	Second =180 $s = h$	Second =180 $s = l$
NE-M	0.0149	0.0556*	-0.0214	-0.0180
	(0.0339)	(0.0317)	(0.0421)	(0.0501)
E-NM	0.0168	0.0412	0.0283	-0.0503
	(0.0312)	(0.0281)	(0.0355)	(0.0463)
Constant	0.580***	0.582***	0.772***	0.274***
	(0.0215)	(0.0203)	(0.0258)	(0.0262)
Observations	220	220	120	100

Standard errors in parentheses.

Errors clustered at the individual level for second 44 and at the pair level later.

***1%, **5%, *10% significance.

Constant shows beliefs in E-M; NE-M and E-NM capture difference relative to E-M.

Table 6: OLS Estimation (Dependent Variable: Beliefs in Part 3)

	Second =44 All	Second =89 All	Second =180 $s = h$	Second =180 $s = l$
NE-M	0.0471**	0.0159	0.00588	0.0936*
	(0.0227)	(0.0217)	(0.0239)	(0.0535)
E-NM	-0.0177	0.00473	-0.00714	0.00468
	(0.0288)	(0.0226)	(0.0269)	(0.0488)
Constant	0.701***	0.715***	0.845***	0.281***
	(0.0176)	(0.0159)	(0.0137)	(0.0321)
Observations	220	220	120	100

Standard errors in parentheses.

Errors clustered at the individual level for second 44 and at the pair level later.

***1%, **5%, *10% significance.

Constant shows beliefs in E-M; NE-M and E-NM capture difference relative to E-M.

[57] Beliefs conditional on signals (at the 180th second) are indicative of some degree of base-rate neglect in these parts.

B Behavior in Phase 1 (Part 4: Main Elicitation Task)

Table 7: OLS Estimation (Dependent Variable: Beliefs in Second 44)

	All	All	Low Group	High Group
Exchange	0.00310	0.00310	0.0267	-0.0205
	(0.0381)	(0.0359)	(0.0594)	(0.0407)
High Group		0.145***		
		(0.0365)		
Constant	0.665***	0.592***	0.577***	0.752***
	(0.0284)	(0.0346)	(0.0430)	(0.0301)
Observations	160	160	80	80

Standard errors (clustered at the subject level) in parentheses.

***1%, **5%, *10% significance. Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

High Group takes value 1 for High Group (0 otherwise).

Table 8: OLS Estimation (Dependent Variable: Average Beliefs in Seconds 35-44)

	All	All	Low Group	High Group
Exchange	0.00289	0.00289	0.0290	-0.0232
	(0.0380)	(0.0359)	(0.0592)	(0.0408)
High Group		0.143***		
		(0.0364)		
Constant	0.663***	0.592***	0.575***	0.751***
	(0.0284)	(0.0344)	(0.0427)	(0.0305)
Observations	160	160	80	80

Standard errors (clustered at the subject level) in parentheses.

***1%, **5%, *10% significance. Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

High Group takes value 1 for High Group (0 otherwise).

C Behavior in Phase 2 (Part 4: Main Elicitation Task)

Table 9: OLS Estimation (Dependent Variable: Absolute Distance in Beliefs within Pair in Seconds 44 to 89)

	E-M	NE-M	E-R
89th Second	-0.143***	-0.00500	-0.0107
	(0.0250)	(0.00753)	(0.0175)
Constant	0.291***	0.237***	0.153***
	(0.0291)	(0.0359)	(0.0315)
Observations	96	60	54

Standard errors in parentheses.

Errors clustered at the pair level.

***1%, **5%, *10% significance.

Table 10: OLS Estimation (Dependent Variable: Change in Beliefs from Seconds 44 to 89)

	All	Low Group	High Group
Exchange	0.0481***	0.0738***	0.0224
	(0.0162)	(0.0247)	(0.0205)
Constant	0.00650	0.00200	0.0110
	(0.00492)	(0.00234)	(0.00950)
Observations	160	80	80

Standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Table 11: OLS Estimation (Dependent Variable: Change in Belief Errors from Seconds 44 to 89)

	All	Low Group	High Group
Exchange	0.0257	0.0738***	-0.0224
	(0.0177)	(0.0247)	(0.0205)
Constant	-0.00450	0.00200	-0.0110
	(0.00499)	(0.00234)	(0.00950)
Observations	160	80	80

Standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Belief error is equal to 1- belief for high group,

and equal to belief for low group.

Table 12: OLS Estimation (Dependent Variable: Change in Probability of Winning Prize from Seconds 44 to 89)

	All	Low Group	High Group
Exchange	-0.0384**	-0.0870***	0.0101
	(0.0164)	(0.0271)	(0.0118)
Constant	0.000105	-0.00304	0.00325
	(0.00235)	(0.00336)	(0.00314)
Observations	160	80	80

Standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Expected winning probability is equal to 1- (1- belief)^2 for high group,

and equal to 1- belief^2 for low group.

Table 13: OLS Estimation (Dependent Variable: Change in Average Beliefs from Seconds 35-44 to 80-89)

	All	Low Group	High Group
Exchange	0.0520***	0.0718***	0.0322*
	(0.0148)	(0.0238)	(0.0173)
Constant	0.00718	0.00210	0.0123
	(0.00613)	(0.00226)	(0.0120)
Observations	160	80	80

Standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Table 14: OLS Estimation (Dependent Variable: Change in Beliefs from Seconds 44 to 89)

	Optimistic All	Pessimistic All	Optimistic High Group	Optimistic Low Group	Pessimistic High Group	Pessimistic Low Group
Exchange	-0.0221	0.124***	-0.0257	-0.0184	0.0752***	0.174***
	(0.0140)	(0.0267)	(0.0241)	(0.0157)	(0.0250)	(0.0463)
Constant	0.00167	0.0113	0.00267	0.000667	0.0193	0.00333
	(0.00112)	(0.00840)	(0.00206)	(0.00138)	(0.0183)	(0.00364)
Observations	82	78	41	41	39	39

Bootstrapped standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Table 15: OLS Estimation (Dependent Variable: Change in Average Beliefs from Seconds 35-44 to 80-89)

	Optimistic All	Pessimistic All	Optimistic High Group	Optimistic Low Group	Pessimistic High Group	Pessimistic Low Group
Exchange	-0.0103	0.120^{***}	-0.000274	-0.0203	0.0683^{**}	0.172^{***}
	(0.0100)	(0.0305)	(0.0143)	(0.0151)	(0.0318)	(0.0597)
Constant	0.00117^{*}	0.0132	0.00147	0.000867	0.0231	0.00333
	(0.000701)	(0.0132)	(0.00117)	(0.00123)	(0.0232)	(0.00307)
Observations	82	78	41	41	39	39

Bootstrapped standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Table 16: OLS Estimation (Dependent Variable: Change in Beliefs from Seconds 44 to 89)

	Low Partner All	High Partner All	Low Partner High Group	Low Partner Low Group	High Partner High Group	High Partner Low Group
Exchange	-0.00836	0.112^{***}	-0.0300	0.00694	0.0657^{***}	0.182^{***}
	(0.0165)	(0.0249)	(0.0320)	(0.0196)	(0.0235)	(0.0684)
Constant	0.000625	0.0132	0.000909	0.000476	0.0168	0.00556
	(0.000859)	(0.00985)	(0.00101)	(0.00104)	(0.0152)	(0.00480)
Observations	85	75	33	52	47	28

Bootstrapped standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Low (High) partner refers to cases where partner has a belief weakly below (stricly above) the median belief at the end of Phase 1.

Table 17: OLS Estimation (Dependent Variable: Change in Average Beliefs from Seconds 35-44 to 80-89)

	Low Partner All	High Partner All	Low Partner High Group	Low Partner Low Group	High Partner High Group	High Partner Low Group
Exchange	0.00262	0.108***	-0.00255	0.00625	0.0619**	0.178***
	(0.0118)	(0.0307)	(0.0122)	(0.0163)	(0.0249)	(0.0536)
Constant	0.000719	0.0146	0.000909	0.000619	0.0188	0.00556
	(0.000734)	(0.0138)	(0.000846)	(0.00101)	(0.0149)	(0.00627)
Observations	85	75	33	52	47	28

Bootstrapped standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Low (High) partner refers to cases where partner has a belief weakly below (stricly above) the median belief at the end of Phase 1.

Table 18: OLS Estimation (Dependent Variable: Adjustment in Beliefs from Seconds 44 to 89)

	E-M Optimistic	E-M Pessimistic
Constant	-0.0853	0.526***
	(0.0941)	(0.105)
Observations	48	48

Standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Adjustment represents change in beliefs from 44th to 89th second,

divided by absolute difference in beliefs within a pair at 44th second.

D Behavior in Phase 3 (Part 4: Main Elicitation Task)

D.1 Regressions

Table 19: OLS Estimation (Dependent Variable: Change in Beliefs from Seconds 44 to 180)

	All	Low Group	High Group
Exchange	0.0236	0.0235	0.0258
	(0.0335)	(0.0541)	(0.0412)
High Signal	0.390***	0.408***	0.386***
	(0.0327)	(0.0503)	(0.0491)
Constant	-0.235***	-0.232***	-0.241***
	(0.0313)	(0.0473)	(0.0432)
Observations	160	80	80

Standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Table 20: OLS Estimation (Dependent Variable: Change in Average Beliefs from Seconds 35-44 to 171-180)

	All	Low Group	High Group
Exchange	0.0138	-0.00217	0.0311
	(0.0331)	(0.0543)	(0.0406)
High Signal	0.390***	0.403***	0.380***
	(0.0330)	(0.0578)	(0.0480)
Constant	-0.232***	-0.225***	-0.237***
	(0.0312)	(0.0476)	(0.0431)
Observations	160	80	80

Standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Table 21: OLS Estimation (Dependent Variable: Change in Beliefs from Seconds 44 to 180)

	Optimistic All	Pessimistic All	Optimistic High Group	Optimistic Low Group	Pessimistic High Group	Pessimistic Low Group
Exchange	-0.0418	0.0973*	-0.0375	-0.0451	0.0891	0.107**
	(0.0445)	(0.0514)	(0.0536)	(0.0777)	(0.0625)	(0.0493)
High Signal	0.362***	0.423***	0.279***	0.420***	0.493***	0.425***
	(0.0416)	(0.0402)	(0.0574)	(0.0715)	(0.0814)	(0.0678)
Constant	-0.252***	-0.220***	-0.178***	-0.295***	-0.304***	-0.178***
	(0.0476)	(0.0424)	(0.0611)	(0.0729)	(0.0812)	(0.0342)
Observations	82	78	41	41	39	39

Bootstrapped standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

Table 22: OLS Estimation (Dependent Variable: Change in Average Beliefs from Seconds 35-44 to 171-180)

	Optimistic All	Pessimistic All	Optimistic High Group	Optimistic Low Group	Pessimistic High Group	Pessimistic Low Group
Exchange	-0.0411	0.0766*	-0.0282	-0.0533	0.0899	0.0623
	(0.0482)	(0.0457)	(0.0523)	(0.0680)	(0.0618)	(0.0594)
High Signal	0.355***	0.430***	0.267***	0.416***	0.494***	0.417***
	(0.0429)	(0.0508)	(0.0548)	(0.0656)	(0.0672)	(0.0941)
Constant	-0.247***	-0.220***	-0.172***	-0.289***	-0.305***	-0.169***
	(0.0515)	(0.0455)	(0.0579)	(0.0609)	(0.0754)	(0.0491)
Observations	82	78	41	41	39	39

Bootstrapped standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Data from E-M and NE-M.

Exchange takes value 1 for E-M (0 otherwise).

D.2 Looking at Updating More Closely

We estimate the following equation

$$\log\left(\frac{p}{1-p}\right) = \alpha \log\left(\frac{p_0}{1-p_0}\right) + \beta_h \mathbf{1}_{(s=h)} \log\left(\lambda_h\right) + \beta_l \mathbf{1}_{(s=l)} \log\left(\lambda_l\right) + \gamma \mathbf{1}_{(p_0<\tilde{p})} \tag{4}$$

where p (p_0) is a subject's belief at the end of Phase 3 (Phase 1) and $\mathbf{1}_{(\cdot)}$ is an indicator function. α measures responsiveness to the prior, β_h and β_l measure responsiveness to the h and l signals. λ_s is the odds ratio of observing signal s.[58] γ captures whether subject with low prior (relative to their counterpart) update their beliefs differently. If updating behavior is consistent with Bayes' rule in aggregate and the bias amplification observed through social-exchange is eliminated after the introduction of the public signal, we would expect $\alpha = \beta_h = \beta_l = 1$ and $\gamma = 0$. We also estimate the same regression for NE-M to provide a benchmark. Note that the coefficient for γ is positive and significantly different from 0 suggesting social-exchange to have some persistence.

Table 23: OLS Estimation (Dependent Variable: Log Posterior Odds Ratio in Second 180)

	No-Exchange	Exchange
α	0.945***	0.873***
	(0.139)	(0.0930)
β_h	0.926***	1.030***
	(0.259)	(0.302)
β_l	1.221***	1.356***
	(0.276)	(0.287)
γ	0.261	0.890***
	(0.364)	(0.251)
$H_0 : \beta_h = \beta_l$	0.522	0.500
Observations	60	100

Standard errors in parentheses, clustered at the subject level in NE-M, pair level in E-M. The second to last row shows p-values associated with testing $\beta_h = \beta_l$.

***1%, **5%, *10% significance.

[58] $\lambda_h = \frac{1}{\lambda_l} = \frac{0.75}{1-0.75} = 3$.

E Mechanisms

E.1 Estimating Distribution of Beliefs in Phase 1

In this section we provide details of how we estimate the distribution of beliefs (focusing on the normal distribution truncated to the unit interval) that is most consistent with stated beliefs at the end of Phase 1 (second 44).

Let (μ_h, σ_h) and (μ_l, σ_l) denote the parameters that characterize the distribution of beliefs at the end of phase 1 conditional on the state (high vs. low group). Let p_0^i denote the stated belief of subject i at this point.

We do a maximum likelihood estimation to find values of (μ_h, σ_h) and (μ_l, σ_l) that are most consistent with observed beliefs at the end of Phase 1.[59] That is we search for values that maximize

$$\sum_{i\in\{1,\ldots,100\}} \mathbf{1}_{(i\in h)} \log\left(\frac{1}{k_h\sigma_h}\phi\left(\frac{p_0-\mu_h}{\sigma_h}\right)\right) + \mathbf{1}_{(i\in l)} \log\left(\frac{1}{k_l\sigma_h}\phi\left(\frac{p_0-\mu_l}{\sigma_l}\right)\right)$$

where $\mathbf{1}_{(\cdot)}$ is an indicator function denoting whether the subject is in the high vs. low group. For $g \in \{h, l\}$, $k_g = \Phi\left(\frac{1-\mu_g}{\sigma_g}\right) - \Phi\left(\frac{0-\mu_g}{\sigma_g}\right)$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution.

Figure **??** plots the estimated distributions. Figure 7 displays the empirical distribution of beliefs in the two groups.

[59]We restrict $\mu_h \in [0, 1]$ and $\mu_l \in [0, 1]$.

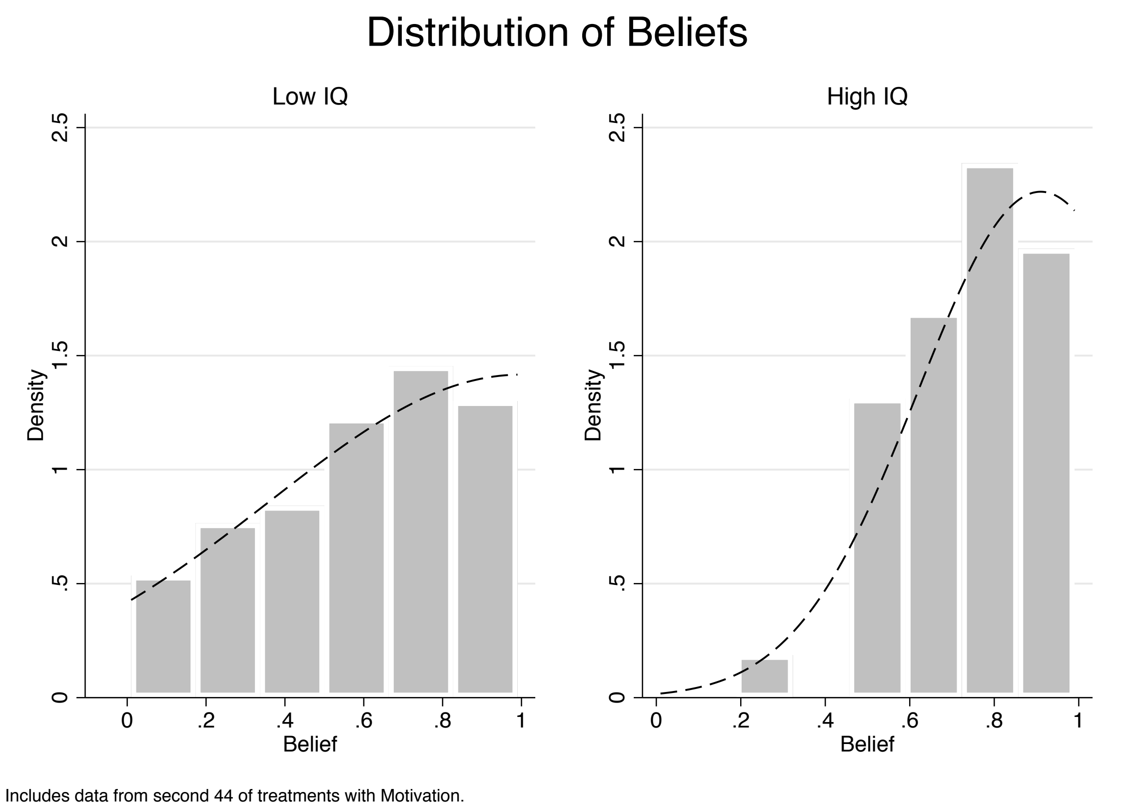

Figure 6: Empirical Distribution of Beliefs with Best Fit at End of Phase 1.

Notes: Columns represent histogram of actual beliefs. Best fit distribution (truncated normal) shown with dashed line: For the High Group: (μ, σ) *= (0.91, 0.29); For the Low Group:* (μ, σ) *= (1, 0.64)*

E.2 Coding Other's Belief as a Binary Signal

The median belief in E-M treatment is 0.7. We code a subject's counterpart's beliefs $\tilde{p}$ at the end of Phase 1 (in second 44) as a binary signal ("good" vs. "bad") indicating whether it is strictly above the median belief or not.[60] When we look at the empirical distribution of this binary signal conditional on state (high vs. low group), we find that λ_h the likelihood ratio of observing the good signal in high vs. low group is 1.47, while the likelihood ratio associated with the bad signal λ_l is 0.71. This confirms that the when $\tilde{p}$ is coded as a binary signal it carries substantial information about which group a subject is in. We then estimate the following regression:

$$\log\left(\frac{p}{1-p}\right) = \alpha \log\left(\frac{p_0}{1-p_0}\right) + \beta_h \mathbf{1}_{(\tilde{p}>0.7)} \log\left(\lambda_h\right) + \beta_l \mathbf{1}_{(\tilde{p}\leq 0.7)} \log\left(\lambda_l\right) \tag{5}$$

where, as before, p (p_0) is a subject's belief at the end of phase 2 (phase 1) and $\mathbf{1}_{(\cdot)}$ is an indicator function. We differentiate between "good" and "bad" signal as defined above. With this specification α measures responsiveness to the prior while β_h and β_l measure responsiveness to the good and bad signals. If updating behavior is consistent with Bayes' rule in aggregate, we would expect $\alpha = \beta_h = \beta_l = 1$. To provide us with a benchmark, we also estimate the same regression for subjects in the No-Exchange- Motivation treatment where (since subjects cannot observe their counterpart's beliefs) we expect $\alpha = 1$ and $\beta_h = \beta_l = 0$. Table 24 reports the results of this estimation.

Table 24: OLS Estimation (Dependent Variable: Log Posterior Odds Ratio End of Phase 2)

	Exchange	No Exchange
α	0.726***	1.030***
	(0.127)	(0.0210)
β_h	3.009***	0.256
	(0.772)	(0.246)
β_l	-0.720	-0.0354
	(0.535)	(0.0551)
$H_0 : \beta_h = \beta_l$	0.003	0.243
Observations	100	60

Standard errors in parentheses, clustered at the subject level in NE-M, pair level in E-M. The second to last row shows p-values associated with testing $\beta_h = \beta_l$.

***1%, **5%, *10% significance.

[60] The unconditional probability of observing a good signal is 47%.

E.3 Mental Model of Beliefs Most Consistent with Bayesianism

In this section we provide details of how we estimate the distribution of beliefs (focusing on the normal distribution truncated to the unit interval) that is most consistent with subjects in aggregate updating their beliefs according to Bayes' rule in phase 2 of E-M.

Let (μ_h, σ_h) and (μ_l, σ_l) denote the parameters that characterize the subjective beliefs about the distribution of beliefs at the end of phase 1 conditional on the state (high vs. low group). We assume all subjects share this common prior. That is, if a subject observes her counterpart to have a belief of $\tilde{p}$, by Bayes' rule, the posterior of the subject conditional on this event would be equal to p_B which would satisfy the following equation:

$$\frac{p_B}{1-p_B} = \left(\frac{p_0}{1-p_0}\right) \frac{\frac{1}{k_h \sigma_h} \phi\left(\frac{\tilde{p}-\mu_h}{\sigma_h}\right)}{\frac{1}{k_l \sigma_l} \phi\left(\frac{\tilde{p}-\mu_l}{\sigma_l}\right)} \tag{6}$$

where p_0 is the prior of the subject at the end of Phase 1. For $g \in \{h, l\}$, $k_g = \Phi\left(\frac{1-\mu_g}{\sigma_g}\right) - \Phi\left(\frac{0-\mu_g}{\sigma_g}\right)$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution.

We do a least squares estimation to find value of (μ_h, σ_h) and (μ_l, σ_l) that are most consistent with observed beliefs at the end of phase 2.[61] That, is we search for values that minimize

$$\sum_{i \in \{1,\ldots,100\}} (p^i - p_B^i)^2$$

where the summation (using i) is over the 100 subjects in our E-M treatment, p^i is the belief of the subject at second 89 and p_B^i the Bayesian belief implied by Equation 6 for a subject with prior p_0^i given parameters (μ_h, σ_h) and (μ_l, σ_l).

[61] We restrict $\mu_h \in [0, 1]$ and $\mu_l \in [0, 1]$.

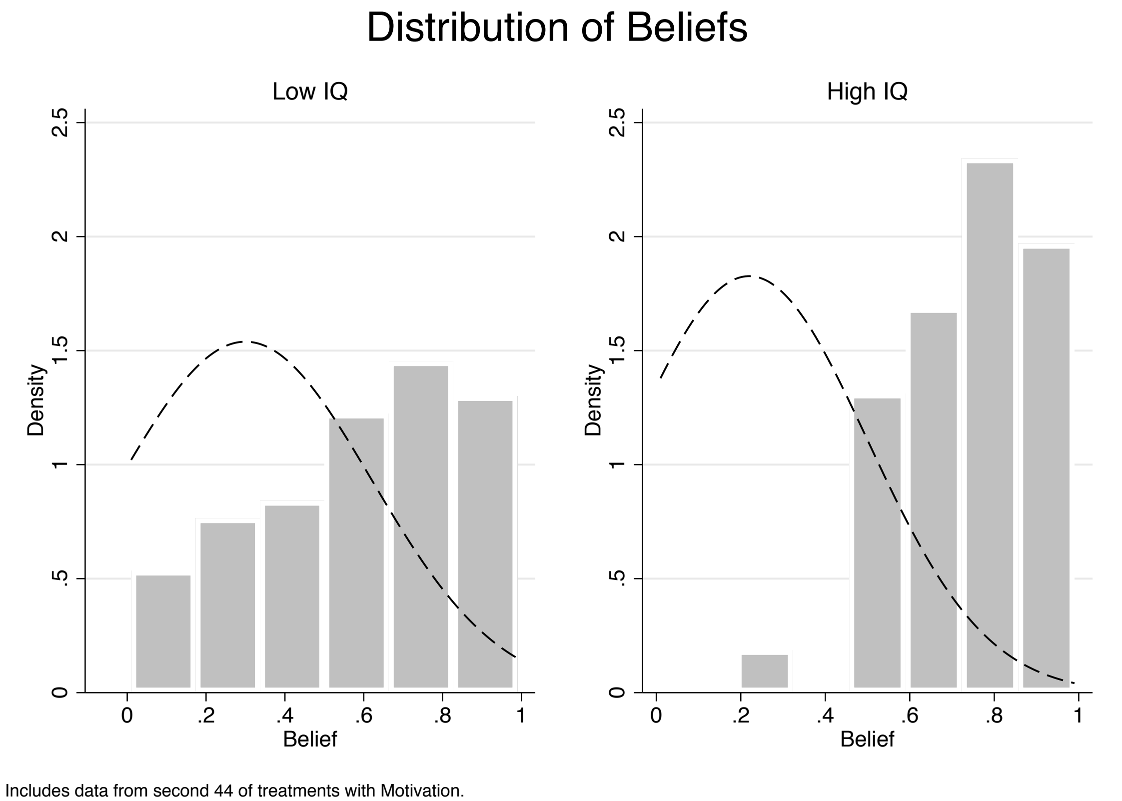

Figure 7: Empirical Distribution of Beliefs with Rationalizing Distribution at End of Phase 1.

Notes: Columns represent histogram of actual beliefs. Best fit rationalizing distribution (truncated normal) shown with dashed line: For the High Group: (μ, σ) *= (0.30, 0.32); For the Low Group:* (μ, σ) *= (0.22, 0.28).*

Figure E.3 plots the estimated distributions. Using the estimated values for (μ_h, σ_h) and (μ_l, σ_l), in Table 25 we report results of the following regression:

$$\log\left(\frac{p}{1-p}\right) = \alpha \log\left(\frac{p_0}{1-p_0}\right) + \beta_h \mathbf{1}_{(\lambda_{\tilde{p}}>1)} \log(\lambda_{\tilde{p}}) + \beta_l \mathbf{1}_{(\lambda_{\tilde{p}}\leq 1)} \log(\lambda_{\tilde{p}}) \tag{7}$$

where, as before, p (p_0) is a subject's belief at the end of phase 2 (phase 1) and $\mathbf{1}_{(\cdot)}$ is an indicator function. We differentiate between "good" and "bad" signals: values of $\tilde{p}$ which are indicative of being in the high vs. low group (depends on whether $\lambda_{\tilde{p}}$ is greater than 1). To be clear, we are not treating $\tilde{p}$ as a binary signal: $\lambda_{\tilde{p}}$ assigns a different informational value (likelihood ratio) for each $\tilde{p}$.[62]

[62] $\lambda(\tilde{p}) := \frac{\frac{1}{k_h \sigma_h}\phi\left(\frac{\tilde{p}-\mu_h}{\sigma_h}\right)}{\frac{1}{k_l \sigma_l}\phi\left(\frac{\tilde{p}-\mu_l}{\sigma_l}\right)}$ where for $i \in \{h, l\}$, $k_i = \Phi\left(\frac{1-\mu_i}{\sigma_i}\right) - \Phi\left(\frac{0-\mu_i}{\sigma_i}\right)$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution.

Table 25: OLS Estimation (Dependent Variable: Log Posterior Odds Ratio End of Phase 2)

	Exchange	No-Exchange
α	0.723***	1.032***
	(0.115)	(0.0227)
β_h	1.174***	0.103
	(0.271)	(0.111)
β_l	0.690	0.213
	(0.692)	(0.199)
$H_0 : \beta_h = \beta_l$	0.553	0.663
Observations	100	60

Estimation results are on updating from end of phase 1 to 2. Standard errors in parentheses, clustered at the subject level in NE-M, pair level in E-M. The second to last row shows p-values associated with testing $\beta_h = \beta_l$.

***1%, **5%, *10% significance.

With this specification α measures responsiveness to the prior while β_h and β_l measure responsiveness to the good and bad signals. If updating behavior is consistent with Bayes' rule in aggregate, we would expect $\alpha = \beta_h = \beta_l = 1$. To provide us with a benchmark, we also estimate the same regression for subjects in the No-Exchange- Motivation treatment where (since subjects cannot observe their counterpart's beliefs) we expect $\alpha = 1$ and $\beta_h = \beta_l = 0$.

E.4 Behavior in Exchange-No Motivation Treatment

Table 26: OLS Estimation (Dependent Variable: Log Posterior Odds Ratio End of Phase 3)

	No-Exchange	Exchange
α	0.392**	-0.327
	(0.167)	(0.227)
β_h	1.615***	1.725***
	(0.270)	(0.218)
β_l	1.411***	0.790***
	(0.221)	(0.285)
$H_0 : \beta_h = \beta_l$	0.637	0.034
Observations	60	60

Estimation results are on updating from end of phase 2 to 3.

Standard errors in parentheses, clustered at the subject level for

No-Exchange, pair level for Exchange. The second to last row shows p-values

associated with testing $\beta_h = \beta_l$.

***1%, **5%, *10% significance. Data from E-NM.

E.5 Studying Factors that Impact Influence

Table 27: OLS Estimation

	(1)	(2)
Pessimistic	-0.163***	0.0434
	(0.0416)	(0.0444)
Constant	0.604***	0.508***
	(0.0331)	(0.0259)
Observations	100	100

Standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Pessimitic denotes subjects whose belief at the end of Phase 1 is lower than their counterpart.

Dependent variable in (1) Belief of higher IQ performance relative to counterpart; in (2)

Belief of higher performance in Part 2 (first practice elicitation task) relative to counterpart;

Table 28: OLS Estimation (Dependent Variable: How Much Subject is Influenced by Counterpart (Self-declared))

	(1)
Relative performance (IQ)	-1.283***
	(0.467)
Relative performance (Part 2)	0.492
	(0.531)
Constant	2.744***
	(0.358)
Observations	100

Standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

Relative performance (IQ): Belief of higher IQ performance relative to counterpart.

Relative performance (Part 2): Belief of higher performance in Part 2 relative to counterpart.

Table 29: OLS Estimation (Dependent Variable: Change in Beliefs from Seconds 44 to 180)

	(1)	(2)
Pessimistic	0.140***	1.666***
	(0.0241)	(0.270)
Raven Score	-0.00161	-0.253***
	(0.00763)	(0.0882)
Score on Other Cognitive Tasks	-0.0176	-0.0916
	(0.0112)	(0.102)
Male	0.00584	-0.253
	(0.0339)	(0.289)
STEM Major	0.0266	-0.365
	(0.0290)	(0.300)
Constant	0.0667	1.563**
	(0.0545)	(0.775)
Observations	100	100

Standard errors (clustered at the pair level) in parentheses.

***1%, **5%, *10% significance.

(1): Change in beliefs; (1) Change is log posterior odds ratio.

F Instructions for E-M

Welcome:

You are about to participate in an experiment in the economics of decision-making. Follow these instructions carefully. In this experiment, you can earn a **CONSIDERABLE AMOUNT OF MONEY**, which will be **PAID TO YOU IN CASH** at the end of the experiment.

Please put away cell phones now. Please do not interact with other participants.

The experiment will consist of 5 parts. One of these parts will be randomly selected for payment at the end of the experiment. In addition to your earnings from the experiment, you will also receive a show up fee of $10 for participating in the experiment.

Instructions for Part 1

- You will be asked to solve 10 questions in this part.

- You will have at most 75 seconds to spend on each question.

How is your payment from this part determined?

- If you are paid for this part, the computer will randomly pick one of the 10 questions. You will receive $10 if you answered that question correctly and 0 otherwise.

Instructions for Part 2

- There are 20 people in this session. The computer randomly assigned people into two groups. 12 people are in the Turquoise group and 8 people are in the Purple group.

- Your task is to submit your belief about how likely it is that you are in the Turquoise group vs. the Purple group.

- To indicate your beliefs, you'll use a slider. Where you move the slider will represent your best assessment of the likelihood (expressed as a chance out of 100) that you are in the Turquoise group or you are in the Purple group.

    ▪ *Example:* By moving the slider almost all the way to the right you can indicate that you believe that you are in the Turquoise group with 5% chance, and the Purple group with 95% chance.

    ▪ *Example:* By moving the slider almost all the way to the left you can indicate that you believe that you are in the Turquoise group with 95% chance, and the Purple group with 5% chance.

- You will have, in total, 3 minutes to adjust your belief using the slider. One of these seconds will be chosen randomly (all are equally likely to be selected), and your decision at that point will be implemented by the computer. (You are free to adjust your belief as many times as you like, but it is in your best interest to always report your best assessment.)

- 90 seconds after the timer has started, to aid your decision, the computer will present results of a "test" to provide some information on whether you are in the Purple or Turquoise group. The test result is either *Purple* or *Turquoise* and has a reliability of 75%. That is,

    o If you are in the Turquoise group, the test result will be *Turquoise* with 75% probability and the test result will be *Purple* with 25% probability.
    o If you are in the Purple group, the test result will be *Purple* with 75% probability and the test result will be *Turquoise* with 25% probability.

How is your payment from this part determined?

- If you are paid for this part, your earnings will depend on how accurate the belief you submit with the slider is at a randomly selected second. The more accurate the belief you submit is, the more money you will earn on average.

- To determine your payment, the computer will randomly draw two numbers. For each draw, all numbers between 0 and 100 (including decimal numbers) are equally likely to be selected. Draws are independent in the sense that the outcome of the first draw in no way affects the outcome of the second draw.

- If you are in the Turquoise group and the number you indicated as the chance that you are in the Turquoise group is larger than either of the two draws, you will get $10.

- If you are in Purple group and the number you indicated as the chance that you are in the Purple group is larger than either of the two draws, you will get $10.

The main takeaway is that the rules were purposefully designed so you maximize your chances of winning $10 by submitting a belief that coincides with your best assessment of the likelihood that you are in each group.

Instruction for Part 3

- The structure of this part is very similar to the previous part.
  - As before, your task will be to submit your belief about how likely you think that you are in one of two groups.
  - You will have 3 minutes to adjust your belief.
  - After 90 seconds, to aid your decision, the computer will present results of a test to provide some information which group you are in. The reliability of the test result is 75%, just as in the previous part.

- The only difference in this part are in the two groups. There are 20 people in this session. The computer randomly assigned people into two groups. **14 people are in the Orange group and 6 people are in the Blue group.**

How is your payment from this part determined?

- If you are paid for this part, your earnings will be calculated exactly as described in part 2.

- To remind you, the belief you submit will impact your chance of winning $10 and you maximize your chances of winning this prize by submitting a belief that coincides with your best assessment of the likelihood that you are in each group.

Instructions for Part 4

- The structure of this part is similar to the previous part.
    - As before, your task will be to submit your belief about how likely you think it is that you are in each of two groups.
    - You will have 3 minutes to adjust your belief.
    - After 90 seconds, to aid your decision, the computer will present the results of a test to provide some information about which group you are in. The reliability of the test result is 75% as before.

Here are the ways in which this part is different from the previous part:

- There are 20 people in this session.

- Questions asked in Part 1 of the experiment are often used as **non-verbal tests of intelligence**, and the purpose of these questions was to form an estimate on your IQ. Using scores on this part of the experiment, we ranked everyone in terms of their intelligence.

- The people with the top 10 intelligence scores were put in the Green group and the people with the bottom 10 intelligence scores were put in the Red group. (Ties were broken randomly.)

- In this part, you will also be matched with one other person from your own group. This means that if you are in the Green group, the person you are matched with is also in the Green group, and if you are in the Red group, the person you are matched with is also in the Red group.

- For the first 45 seconds, you will submit your belief on your own. After this, both of you will observe each other's submitted beliefs in the remaining time.

- While you will be able to change only your own submitted beliefs on the slider (at the top of the screen), you will also observe the submitted beliefs of the person you are matched with (on the slider that will be located on the bottom of the screen). You will instantly see those beliefs change if the other player moves his or her slider, and the other player will see any slider movements you make.

- The test is done only **ONCE**. Both you and the person you are matched with will observe the **SAME** test result, which will be made public to you both at the same time on your screens.

How is your payment from this part determined?

- If you are paid for this part, your earnings will be calculated exactly as described in part 2.

- The decision of the person you are matched with will have no impact on your earnings.

- To remind you the belief you submit will impact your chance of winning $10 and you maximize your chances of winning this prize by submitting a belief that coincides with your best assessment of the likelihood that you are in either group.

Instructions for Part 5

In this part, you will be asked several questions. Some of them are puzzle questions, some of them are belief questions about the performance of others in this experiment. The remaining are survey questions.

How is your payment from this part determined?

- If you are paid for this part, one of the puzzle or belief questions will be chosen randomly. If the randomly picked question is a puzzle question, you will receive $10 if you answer that question correctly and 0 otherwise. If this is a belief question, your earnings will be calculated exactly as described in Part 2 (i.e. you maximize your chance of earning $10 by submitting a belief that coincides with your best assessment).