Are Biases Contagious?
The Influence of Communication on Motivated Beliefs[a]

Andreas Grunewald Victor Klockmann Alicia von Schenk
Ferdinand A. von Siemens[b]

May 21, 2024

Abstract

This paper examines the potential reinforcement of motivated beliefs when individuals with identical biases communicate. We propose a controlled online experiment that allows to manipulate belief biases and the communication environment. We find that communication, even among like-minded individuals, diminishes motivated beliefs if it takes place in an environment without previously declared external opinions. In the presence of external plural opinions, however, communication allows motivated beliefs to persist. Our results indicate a potential drawback of a plurality of opinions – it may create communication environments wherein motivated beliefs become contagious within social networks.

JEL: C91, C92, D83
Keywords: Belief bias, Social interaction, Motivated beliefs

---

[a]Financial support by the DFG, grant FOR 5392 project number 462020252, and by the Leibniz Institute for Financial Research SAFE, grants 232303 and 232408, is gratefully acknowledged. We thank seminar participants at DICE Düsseldorf, FAU Erlangen-Nürnberg, Goethe-University Frankfurt, Thurgau Institute of Economics, Maastricht University, University of Hamburg, University of Heidelberg, and University of Würzburg for valuable comments and suggestions. The paper has also benefited from discussions with Andrea Amelio, Russell Golman, Michael Kosfeld, Ricardo Perez-Truglia, Frederik Schwerter, Roberto Weber, and Florian Zimmermann. We are grateful for excellent research assistance by Jonathan Gehle and Zhuokun Liu.

[b]Grunewald: Frankfurt School of Finance and Management (email: a.grunewald@fs.de). Klockmann: Julius-Maximilians-Universität Würzburg and Max Planck Institute for Human Development (email: victor.klockmann@uni-wuerzburg.de). von Schenk: Julius-Maximilians-Universität Würzburg and Max Planck Institute for Human Development (email: alicia.vonschenk@uni-wuerzburg.de). von Siemens: Goethe University Frankfurt, CESifo and Leibniz Institute for Financial Research SAFE (email: vonsiemens@econ.uni-frankfurt.de).

1 Introduction

A significant portion of our beliefs exhibit systematic distortions. We display overconfidence regarding our abilities or outward appearance (Moore and Healy; 2008; Kogan et al.; 2021; Huffman et al.; 2022), endorse fabricated statistics or false information about political outcomes (Flynn et al.; 2017), and hold self-serving beliefs about individuals outside our social circles (Di Tella et al.; 2015; Ging-Jehli et al.; 2020). Importantly, these biased beliefs are not formed in isolation but are often shaped by social interactions and communication with other individuals (Oprea and Yuksel; 2022; Enke et al.; 2023; Amelio; 2023). Moreover, individuals tend to seek out communication partners with similar characteristics, beliefs and, arguably, similar belief biases (McPherson et al.; 2001; Gentzkow and Shapiro; 2011; Bakshy et al.; 2015) – a tendency that appears to be reinforced by the matching algorithms of various social media platforms (Cinelli et al.; 2021). A major concern is that in combination with beliefs biases, such selective communication may aggravate belief distortions rather than accumulate information, potentially undermining social cohesion by promoting extremism, polarization of political beliefs, violence, political gridlock, and social immobility (Levy and Razin; 2019; Sunstein; 2017).

One form of bias that has come under particular scrutiny in this respect is the tendency to form motivated beliefs. Extensive research has shown that individuals often hold these biases to enhance their self-image or gain anticipatory utility. Importantly, achieving this frequently requires individuals to engage in racial discrimination (Eyting; 2022), adopt specific opinions and political ideologies (Schwardmann et al.; 2022; Sprengholz et al.; 2023), or embrace biased negative perceptions of outsiders (Di Tella et al.; 2015; Ging-Jehli et al.; 2020). Thus, the dissemination of these biases through communication within like-minded communities is thought to significantly influence policy preferences, political ideologies, and polarization (Bénabou and Tirole; 2006; Bénabou; 2015; Levy and Razin; 2019; Sprengholz et al.; 2023). Despite the severe consequences that reinforcement of motivated beliefs may thus have, causal evidence on whether communication can indeed lead to an accumulation of such biases rather than an aggregation of information is still absent.

Given that individuals maintain motivated beliefs for self-serving purposes, the effect of communication on the extent of the bias is, in fact, unclear. On the one hand, individuals' motivation to maintain their beliefs may lead them to selectively ignore, discount, or forget parts of a conversation that contradict their self-serving views (Bénabou; 2015; Golman et al.;

2017; Zimmermann; 2020; Thaler; 2023b). On the other hand, expressing motivated beliefs may damage the speaker's reputation or conflict with prevailing social norms (Loury; 1994; Braghieri; 2022; Bursztyn et al.; 2020; Golman; 2023). The relative importance of these countervailing effects presumably depends on the communication environment, in particular the presence of external plural opinions that individuals can draw upon and refer to. Such plural opinions can provide justifications for upholding one's self-serving convictions, and they may offer social cover so that individuals feel comfortable expressing socially ostracized views that support their motivated beliefs (Masser and Phillips; 2003; Bursztyn, Egorov, Haaland, Rao and Roth; 2023). Indeed, the ease with which individuals can find rationales even for the most extreme, unpopular, or factually incorrect opinions is a major concern with respect to many online communication environments.

In light of these considerations, this paper develops an experimental paradigm to analyze the propagation of motivated beliefs through communication. We embrace the notion that communication may be particularly problematic if belief distortions are shared within a homogeneous communication network, and we explore the importance of the communication environment for the diffusion of motivated beliefs. Therefore, we study two main questions: (i) Does communication between individuals with the same motivated beliefs reinforce or reduce their biases? (ii) Which communication environments are particularly conducive to the formation of motivated beliefs?

Our experimental design has three important features which are essential to investigate these questions. First, the experiments take place online, which allows us to implement communication between participants in a natural but controlled way through free-form chats. Second, the setup enables us to exogenously induce motivated beliefs without affecting the information participants hold. This feature is indispensable for drawing inferences about whether communication among individuals with similar biases leads to a proliferation of those biases. In particular, naturally occurring belief biases are typically correlated with preferences, information, communication habits, and how individuals process new information. Therefore, pairing participants with similar natural belief biases would confound the effect of the chat partner's bias in beliefs with that of her characteristics and information. It would thus be impossible to determine whether communication leads to an accumulation of information or a reinforcement of biases. Third, we can exogenously manipulate whether or not participants see opinions beyond those expressed in the chat. This feature enables us to study how the

plurality of opinions available to individuals affects the extent to which the communication environment reinforces belief distortions.

In our experiment, we implement a simple decision environment in which participants are randomly assigned to groups of two, with one Player A and one Player B. Each group plays two dictator games, and each player is once the dictator and once the recipient. In the first dictator game, Player A is the dictator and can distribute an endowment either fairly or keep most of it for herself. In the second dictator game, Player B is the dictator and distributes another endowment. However, the exact options that Player B can choose from are randomly determined. In 50% of the cases, Player B has the same options as Player A – a fair split or keeping most for herself. In the other 50% of the cases, Player B has no options to choose from, and the equal distribution is automatically implemented. This variation in Player B's choice options constitutes our first treatment dimension. The crucial element of the experiment is its information structure. All players are equally informed about the structure and payoffs of the two dictator games as well as the two different possible choice sets that Player B may face, but they receive no information about the other players' choices. Furthermore, Player A receives no information on which of the choice sets materialized for Player B, and this is common knowledge.

Our main object of interest is the belief that Player B holds about the behavior of the matched Player A. This belief should, in principle, not depend on Player B's own choice options, which are commonly known to be unknown to Player A. However, the literature on motivated beliefs (Di Tella et al.; 2015; Zimmermann; 2020) suggests that it might be easier for Player B to take most of the endowment for herself if she believes Player A also did so. Because creating such a justification is only advantageous if the endowment can be split unevenly, Player B should only adopt negative perceptions of Player A in a self-serving manner when having this option available. In other words, the variation in choice options induces an exogenous distortion in the beliefs of Player B without providing information – a prerequisite for delineating how communication aggregates biases in beliefs.

Building upon this choice paradigm, we implement two additional randomly assigned treatment dimensions. First, we introduce free-form chats that allow some Players B to communicate with another Player B who faces the same choice options. In other words, participants who hold, on average, the same motivated beliefs communicate with each other. Second, we manipulate the communication environment. For this purpose, we present external plu-

ral opinions to participants apart from those expressed in the chat. In particular, we show two opposing rationales to Player B concerning the behavior of Player A before Player B states her beliefs and before communication takes place (if there is any). This plurality of external opinions reflects a communication environment wherein individuals can draw upon, refer to, and subscribe to previously declared opinions (often in a self-serving manner). If subjects harbor concerns about holding or communicating potentially unpopular opinions, these additional opinions can furnish justifications, even for opinions others might perceive as biased or unpopular (Loury; 1994; Golman; 2023). Based on the ideas of Masser and Phillips (2003) and Bursztyn, Egorov, Haaland, Rao and Roth (2023), this form of social cover is a crucial aspect of a communication environment and can significantly affect an individual's willingness to dissent and openly express own perspectives.

We start the analysis by focusing on the conditions without communication and without additional opinions. These conditions show that the variation in choice options for Player B indeed causes substantial motivated beliefs. In particular, the fraction of Players B believing that their Player A chose the unequal split increases from 42% if Players B do not have choice options to 62% if they have choice options – an increase by roughly 50 percent. This 20pp difference in beliefs constitutes our measure for the prevalence of the induced motivated beliefs of Players B who have choice options and form their beliefs in isolation. Note that our measurement of motivated beliefs is a between-subject measure. This approach has three advantages. First, any within-subject measure would have to elicit beliefs once without a self-serving motivation and once with a self-serving motivation from the same subject—an approach that seems hardly feasible. Second, measuring between subjects also creates an ideal control group that is, on average, identical to the treatment group regarding their characteristics and information. We leverage this feature to separate the formation and aggregation of biases from the accumulation of information through communication. Third, we do not have to assume that Players B hold unbiased beliefs about the behavior of Players A in those conditions where they cannot make an allocation choice. Rather, our approach measures the bias in beliefs that we induce exogenously.

The second set of results addresses whether communication between individuals holding similar belief biases fosters the propagation of motivated beliefs. Again, we first consider the conditions without additional opinions. With communication, we find that the fraction of Players B believing that their Player A chose the unequal split increases from 37% if Players B do not have choice options to 44% if they have choice options. Hence, the belief distortions

of Players B with choice options are significantly smaller with communication (7pp) than without communication (20pp), implying that communication reduces motivated beliefs in this communication environment.

Next, we repeat the same analysis in the conditions with plural opinions. In these conditions, the effect of communication on motivated beliefs reverses: communication leads to an increase in the measure of distortion in beliefs from 10pp without to 16pp with communication. While the figures imply an economically relevant increase in distortions by 60%, the estimates are imprecisely estimated and the difference between them is statistically insignificant. Nevertheless, the estimates clearly show that the existence of external plural opinions at least nullifies or even reverses the reduction in motivated beliefs by communication. Supporting this insight, we further find that the propagation of motivated beliefs through communication is significantly more pronounced with than without a plurality of external opinions: The triple difference estimator, which compares the 6pp *increase* in the bias under plurality of opinions with the 20pp *decrease* without plurality, is statistically significant. Overall, our findings strongly emphasize the role of the communication environment in the reinforcement of motivated beliefs.

To understand why communication under plural opinions is significantly more conducive to the formation of motivated beliefs, we analyze how communication differs between treatments with and without plural opinions. While we did not design our experiment to elucidate the exact mechanisms, the chat content and the correlation of beliefs between chat partners provide some suggestive evidence. Various analyses suggest that the plurality of opinions does not affect the chats' content and tone. For example, chats do not differ in length, the frequency with which participants state a given opinion, the timing when they first mention an opinion, or the discussed topics. Nevertheless, we find that the correlation of beliefs within chat groups is much stronger without plural opinions. Hence, the influence of the chat partner on beliefs seems to be stronger without plural opinions. The chat content suggests that this weaker correlation of beliefs under plural opinions is particularly pronounced when the partner expresses an opinion that would contradict a motivated belief. Our data thus indicate that a plurality of opinions selectively disrupts the link between communication that corrects self-serving belief biases and the beliefs themselves.

2 Related Literature

Our paper draws upon an existing literature that has extensively documented instances of motivated reasoning in controlled experimental settings (Di Tella et al.; 2015; Engelmann et al.; 2019; Ging-Jehli et al.; 2020; Zimmermann; 2020; Drobner; 2022), big life decisions (Müller; 2022), competitive debating environments (Schwardmann et al.; 2022), policy preferences (Sprengholz et al.; 2023; Thaler; 2023b), racial discrimination (Eyting; 2022), and among management professionals (Huffman et al.; 2022). Because upholding such self-serving motivated beliefs often requires individuals to adopt negative perceptions about outsiders (see e.g. Di Tella et al.; 2015; Ging-Jehli et al.; 2020) or certain policy reforms (Sprengholz et al.; 2023; Herz et al.; 2022), motivated reasoning is thought to be a significant catalyst for political polarization and the formation of biased political opinions (Bénabou and Tirole; 2006; Bénabou; 2015; Levy and Razin; 2019). While the existing studies show that motivated beliefs play an important role in individual decisions, the idea that they can drive significant social trends implicitly assumes that social interactions reinforce motivated beliefs or at least allows them to persist. Addressing these ideas, our paper complements the studies in individual decision contexts by investigating under which conditions communication among individuals with, on average, the same motivated beliefs indeed reinforces or reduces the biases themselves.

We also contribute to a recent literature that studies the aggregation of belief biases through various institutions of stylized social interactions. Most of this literature has focused on the transmission of cognitive mistakes rather than motivated biases. Enke et al. (2023) study how various cognitive mistakes propagate through social interactions such as auctions, betting markets, and committee experiments, and Amelio (2023) shows how similar cognitive mistakes are transmitted through social learning. Complementary, Graeber et al. (2024) study the transmission of truths and falsehoods in the context of financial decisions via recorded audio explanations. All of these papers show that in situations with an objectively correct answer, social interactions can reduce biases. In contrast, we study the spread of motivated beliefs through two-sided communication. For motivational biases, the effect of communication on the prevalence of biases is especially unclear, because individuals hold motivated beliefs for a self-serving reason and may therefore be inclined to uphold and communicate them. The potential resistance to social interactions combined with their relevance for political views is also an important reason for why motivated beliefs have come under particular scrutiny. In fact, we show that motivated negative perceptions about outsiders persist after two-sided

communication if there exist plural external opinions in the communication environment.

Closest to our work are two papers that explicitly study the connection between motivated reasoning and social interaction. Oprea and Yuksel (2022) allow subjects to use sliders to signal their probabilistic assessment that they and their partner both have above median or below median IQ. They find that subjects respond to one another's beliefs asymmetrically, causing an amplification of overconfidence. Thaler (2023a) analyzes the effect of incentivizing senders to be perceived as truthful. In response, senders become more likely to supply information in line with the motivated political beliefs of receivers. Our paper differs from these articles in three important ways. First, we study the effect of natural free-form communication on the aggregation of motivated beliefs.[1] Second, we consider motivated negative perceptions about outsiders. It is the spread of such negative motivated beliefs that is deemed to shape policy preferences and to amplify racial discrimination. Third, a key methodological contribution of our work lies in the exogenous manipulation of individuals' biases. This methodology is especially advantageous when studying the aggregation of biases through social interactions. In particular, naturally occurring belief biases are typically correlated with confounding factors such as individuals' information, updating, and communication habits (Benjamin et al.; 2013; Oprea and Yuksel; 2022; Enke et al.; 2023). In contrast to a design that leverages naturally occurring biases, our approach thus allows to cleanly distinguish whether social interactions reinforce the induced bias or simply aggregate information.

From a more general perspective, our findings add to an ongoing discussion about the potential adverse effects of individuals' inclination to communicate with like-minded others. Levy and Razin (2019) and Sunstein (2017) argue that the endogenous selection of communication partners can not only result in political polarization but also contribute to political gridlock, hinder social mobility, and eventually pose a threat to democracy. Their arguments inherently build on the idea that selected social interactions induce a proliferation of biases in beliefs and not just an accumulation of information. While there is ample evidence that individuals actively select their communication network online (Bakshy et al.; 2015; Cinelli et al.;

[1]There is also an experimental literature that studies how communication affects group choices. Papers have documented that communication makes group decisions display less risk aversion (Stoner; 1961; Teger and Pruitt; 1967), leads to more prosocial behavior (Cason and Mui; 1998; Bartling et al.; 2022), improves the efficiency of collective decisions (Goeree and Yariv; 2011), and makes decisions reflect more closely the predictions of Nash Equilibrium (Bornstein and Yaniv; 1998; Schotter; 2003; Cooper and Kagel; 2005; Kocher and Sutter; 2005). In contrast to these papers, we study distortions in individuals' beliefs and consider decisions taken in isolation and not together as a group.

2021) and offline (Verbrugge; 1977; McPherson et al.; 2001; Gentzkow and Shapiro; 2011; Barnidge; 2017; Jackson et al.; 2023), causal evidence on the extent to which communication in these selected networks reinforces belief biases instead of information is still scarce. To collect evidence on the reinforcement of biases when these are shared among communication partners, it is inevitable to exogenously manipulate the communication environment as well as the belief bias itself. Our paper is, to the best of our knowledge, the first that implements such a procedure. Leveraging this approach, we show that the effect of communication on motivated beliefs is contingent upon the communication environment. Specifically, communication is significantly more conducive to the formation of motivated beliefs if individuals can draw upon external plural opinions.

This latter finding also speaks to an ongoing discourse surrounding the impact of fake news on social media (Allcott and Gentzkow; 2017; Barrera et al.; 2020; Bursztyn, Rao, Roth and Yanagizawa-Drott; 2023). While previous studies primarily examine the direct effects of false information, our research uncovers an additional, indirect effect: If there are plural opinions that allow to support even factually wrong or otherwise ostracized arguments, individuals seem to be able to uphold their motivated beliefs in conversations. Therefore, a possibly positive bias-attenuating effect of communication is nullified or even reversed in such communication environments. These findings support arguments favoring debunking fake news before it spreads through social networks to eliminate communication environments that might reinforce motivated beliefs. Complementing the arguments by Bursztyn, Egorov, Haaland, Rao and Roth (2023) that fake news can serve as social cover to express otherwise stigmatized opinions, we thus demonstrate that fake news can shield individuals from revising their self-serving beliefs.

3 Experimental Design

Our study investigates how motivated beliefs propagate through communication within like-minded communities. Studying this question requires a choice paradigm with three essential features. First, we need to cleanly measure a meaningful, behaviorally relevant belief of participants in an incentivized manner. Second, the paradigm must allow us to exogenously induce motivated beliefs. In particular, naturally occurring biases in beliefs are inherently correlated with unobserved personal experiences, preferences, and economic circumstances.

If communication occurred among participants who naturally hold the same belief biases, it would be impossible to delineate whether the communicating partner's biases, information, or personal characteristics drive posterior beliefs. Third, manipulating beliefs in the treatment group must not affect participants' information. If the treatment manipulation was changing information, it would not be possible to disentangle whether any induced shift in beliefs was due to a bias in beliefs or to the informational part of the treatment. This latter part is non-trivial because we most often associate belief changes with updating through incoming information. The following paragraphs describe the implemented choice paradigm and treatments in light of these necessary features.

3.1 Choice Paradigm

We implement a simple decision environment in which participants are randomly matched into groups of two, with one Player A and one Player B. Each group plays two binary dictator games without feedback, and each player is once the dictator and once the recipient. In the first dictator game, Player A is the dictator and has two options to distribute an endowment of £5: she can either choose an equal split of £2.50 for each or allocate £4 to herself and £1 to Player B. In the second dictator game, Player B is the dictator, allocating another additional endowment of £5 to the two players. The exact options that Player B can choose from are randomly determined. In 50% of the cases, Player B faces the same options as Player A, that is, an equal split of £2.50 for each or £4 to herself and £1 for Player A. In the remaining 50% of the cases, Player B has no option to choose from, and the allocation is automatically the equal split.

The crucial element of the experiment is its information structure. All players are equally informed about the structure and payoffs of the two dictator games and the two different possible choice sets that Player B may face. However, they do not receive any information about the other players' choices or choice options. When making her choice, Player A knows neither the choice of Player B nor Player B's choice options. Equally, when making her choice, Player B does not know the choice of Player A. This information structure is essential because it immediately implies that it is common knowledge that the choices of Player A cannot depend on the randomly determined and unknown choice options of Player B.

Our primary outcome measure in all treatments is the belief that Player B holds about the prior behavior of Player A. After the allocation, we ask Player B whether or not the specific

Player A matched with her selected the equal split. If the answer is correct, Player B earns an additional £2.50. Our main outcome variable *"unfavorable belief"* is a dummy variable indicating whether Player B believes Player A to have chosen the unfair allocation.[2] Note that this belief is behaviorally meaningful because it will affect the decision of Player B, and we can cleanly measure it in an incentivized way – it thus fulfills the first requirement for the choice paradigm described above.

We implement the simple binary dictator games to create an obvious tension between self-gratification and social behavior for the participants. While our information structure implies that Player B's belief about the behavior of Player A should not depend on her choice set, the literature on motivated beliefs (Di Tella et al.; 2015; Zimmermann; 2020; Drobner; 2022) suggests that it nevertheless does. In particular, it might be psychologically convenient for Player B to take the £4 for herself if she believes Player A did the same. Generating such a justification to resolve the tension between self-gratification and social behavior self-servingly is only advantageous or necessary if the unfair option is available. Hence, Players B without choice options and those with choice options are identical on average, except that participants with the larger choice set are more motivated to adopt a negative perception of Player A. Any self-serving bias in beliefs induced by the treatment manipulation is therefore exogenous to participants' characteristics and information. The induction of motivated beliefs, therefore, fulfills requirements number two and three of the choice paradigm raised above – it is exogenous and uninformative.[3]

[2]We also elicit two alternative belief measures. First, we obtain a *"general"* belief by asking Player B how many of the previous 100 Players A picked the unequal option. If the answer deviates by at most 5 in absolute terms from the correct answer, the subject earns £2.50. For this belief, we find similar but weaker and sometimes insignificant results (see Appendix C.1). This difference is consistent with Di Tella et al. (2015), who also find more substantial effects for specific than general beliefs. It is also consistent with the idea of motivated beliefs: It is sufficient to hold an unfavorable belief about one's partner to justify unfair behavior. Believing that the full population of participants is selfish is not necessary. Second, in the conditions with Quotes we additionally ask participants to state their certainty about the stated belief regarding the decision of Player A (in %). In Appendix C, we use this question to obtain a continuous measure of beliefs. All results for the Quotes conditions are qualitatively the same for this continuous measure and the binary measure that we use in the main text.

[3]We also elicited the analogous beliefs of Player A as well as her second-order beliefs about Player B's beliefs. If Player B has no choice, we ask Player A only to report her second-order belief concerning Player B. In this case, we inform Player A after her allocation choice that Player B has no choice.

Table 1: Treatment Overview

	No Quotes		Quotes	
	No Chat	Chat	No Chat	Chat
No Choice	NoChoice-NoChat	NoChoice-Chat	NoChoice-NoChat-Quotes	NoChoice-Chat-Quotes
	$n = 171$	$n = 324$	$n = 143$	$n = 270$
Choice	Choice-NoChat	Choice-Chat	Choice-NoChat-Quotes	Choice-Chat-Quotes
	$n = 164$	$n = 333$	$n = 146$	$n = 270$

Note: The table provides an overview of the different treatments of the experiment and the number of Players B in each treatment after taking out the fastest 15% (see below). We balanced the number of independent observations and thus collected about twice as many subjects in the communication treatments as compared to the no communication treatments.

3.2 Treatments

Table 1 depicts an overview of the treatments implemented in our 2x2x2 design. As described above, our first treatment dimension varies Player B's choice set. In the treatment arms labeled Choice, Player B has the option to allocate the endowment unevenly. This option implies a motivation to distort her beliefs regarding Player A's behavior self-servingly. In contrast, Player B in the treatment arms labeled NoChoice has no choice options and, therefore, no incentive to distort her beliefs. From the perspective of Player A, these two treatments are identical because this player is not informed about the choice options of Player B. The average differences in beliefs of Players B between the Choice-NoChat and NoChoice-NoChat conditions measure the extent to which individuals in the Choice treatment hold induced motivated beliefs, i.e, the prevalence of induced motivated beliefs in the absence of communication. This between-subject approach to measurement does not inherently assume that participants in the No Choice conditions hold unbiased beliefs about the behavior of Players A. Rather we track the exogenously induced bias in beliefs and how it propagates via communication in different environments.

Our second treatment dimension varies whether or not participants communicate with another subject in the same role. Players B in the treatment arms labeled NoChat do not

interact with any other participant during the experiment. Players B in the treatment arms labeled CHAT enter a surprise communication stage immediately after reading the instructions and before making their allocation decisions (if any) and reporting their beliefs.[4] The free-form chat lasts for three minutes, and we encourage participants to discuss the previous behavior of Players A and their intended own decisions before the chat starts. We seek to study the propagation of motivated beliefs within like-minded communities. At the communication stage, we thus group Players B together that face the same choice options; they are both either in a CHOICE or a NOCHOICE treatment arm, i.e,. both communication partners hold the same motivation to distort their beliefs. Overall, participants in the conditions NOCHOICE-CHAT and CHOICE-CHAT both communicate and are, on average, identical in all aspects except for their motivation to distort their beliefs. The difference in average beliefs between NOCHOICE-CHAT and CHOICE-CHAT thus allows us to measure the extent to which individuals in CHOICE-CHAT hold induced motivated beliefs after communicating. In analogy, the difference between CHOICE-NOCHAT and NOCHOICE-NOCHAT quantifies the prevalence of induced motivated beliefs in the absence of communication (see above). Importantly, when we compare these two differences, we can identify to what extent communication with individuals holding similar biases propagates these biases. Therefore, when describing the results in Section 4, this difference-in-differences estimator is a major outcome of interest.

Our third treatment dimension is designed to analyze which characteristics of the communication environment foster or mitigate the propagation of motivated beliefs. An essential aspect in this regard is the availability of plural opinions that subjects can draw upon and refer to in the conversations. If subjects have concerns about holding or communicating potentially unpopular opinions (see Loury; 1994; Morris; 2001; Golman; 2023), the availability of such plural opinions may provide social cover and can thereby significantly affect an individual's inclination to communicate and uphold their desired beliefs (Masser and Phillips; 2003; Bursztyn, Egorov, Haaland, Rao and Roth; 2023). To create a communication environment that reflects the presence of multiple external opinions, in the treatment arms with QUOTES, we display two opposing quotes to all Players B after the instructions, i.e, before a potential chat and before their allocation decision. These quotes are

> *"I think we are living in selfish times."*
>
> Javier Bardem, Hollywood actor and Oscar winner.

[4] A screenshot of the chat window can be found in Figure 3 in the appendix.

and

> *"I'm just thankful I'm surrounded by good people."*
> Jon Pardi, singer and songwriter.

All participants in the conditions with QUOTES receive both quotes, also participants who do not have the opportunity to chat. Hence, the idea of providing the quotes is not to steer participants' beliefs in one particular direction but to generate a plurality of opinions that may allow participants to find social cover for holding and expressing either opinion in the chat. While we provide two opposing quotes, it is still plausible that they affect individuals' beliefs in all treatment cells. For example, one of the quotes could be more convincing or one of the actors more popular. Therefore, to examine how the communication environment affects the propagation of motivated beliefs, we need to employ a triple difference estimator, i.e., we study how the difference-in-differences described above depends on the communication environment. In the conditions with NOQUOTES, we simply do not provide the quotes. Because we think of the treatments without the quotes as our baseline, we omit the label NOQUOTES in their acronyms for expositional clarity.

3.3 Experimental Procedures

We conducted the experiment online using Prolific and oTree (Chen et al.; 2016).[5] At the beginning of the experiment, participants received written on-screen instructions explaining the rules and details of the experiment.[6] Afterward, they answered control questions, ensuring a basic understanding of the experiment. We excluded the 22% of registered participants who did not answer all of these questions correctly. We informed the remaining participants whether they were Player A or B. The main part of the experiment started once we matched each Player B to a partner by arrival time. Due to the large active subject pool, the average waiting time was only a few seconds. To hold waiting times between the introduction and the main part constant across treatments, we formed pairs of Players B in all conditions, although only those in the CHAT conditions interacted with each other. At the end of the

[5] We received prior ethics approval from the joint ethics committee of Goethe University Frankfurt and Johannes Gutenberg University Mainz.

[6] A translated version of the instructions for all parts and all treatments of the experiment can be found in Appendix E

experiment, all participants answered short surveys on demographics, their education, and social media usage.[7] For the participants in the conditions with QUOTES, we additionally elicited social preferences by standard questionnaire items.[8]

For all treatments, we first implemented sessions with only the participants in the role of Player A. We informed them that we would later match them with a participant in the role of Player B and that they would receive their payment resulting from the choices of Player B in a second tranche. After collecting the decision data from Players A, we ran sessions with the participants in the role of Player B. We matched each participant in these sessions with one Player A from the first part of the experiment. After this second part of the experiment, we transferred the remaining payoff from Player B's decision to the corresponding Player A.

Our main object of interest is the belief of Player B concerning the behavior of the Player A who is matched with her. To obtain approximately the same number of independent observations in each treatment, we oversampled the treatments in the CHAT conditions by a factor of two. A potential problem with online experiments is that participants might click through the experiment without paying attention to the instructions. To improve data quality, we dropped the fastest 15% of participants in each treatment, overall 316 observations. Our main results in Table 2 below are qualitatively robust to including these observations (see Table 6 in the appendix).

Participants took, on average, about 8 minutes to complete the experiment and earned on average £6.00. In total, 4316 subjects participated in the study. All participants were located in the U.S. and between 18 and 60 years old. We conducted the experiment in two waves. The first wave elicited observations for all treatments in the conditions with NOQUOTES, and the second wave for all treatments in the conditions with QUOTES. Table 1 provides an overview of the number of observations for each treatment. We held all key aspects of the experiment constant across waves. In particular, the order and content of the experimental tasks and instructions were identical across all participants, treatments, and waves of the experiment. Moreover, when recruiting participants, we imposed identical constraints on the

[7] In terms of social media usage, we elicited whether participants actively create content on social media, how often they use social media, and whether they share their political views on social media.

[8] In analogy to Falk et al. (2018), we measure positive reciprocity and general trust by asking, on an eleven-point Lickert scale, whether participants are willing to return a favor and whether they believe that people have only the best intentions. To measure altruism, we asked how many of unexpectedly received £1000 they would be willing to donate to a good cause.

characteristics of the participant pool. Table 5 in the appendix summarizes the balance checks showing that these procedures resulted in an overall well-balanced sample of participants across all treatment cells depicted in Table 1 with two exceptions. Participants in the first wave are less likely to be female and more likely to be older than the median age of 32 years than in the second wave. Most empirical tests we conduct are within waves, so the differences do not affect the corresponding results. For all analyses using data from both waves, we always present the raw treatment differences and specifications controlling for age and gender to account for the wave-specific differences in our empirical analysis.

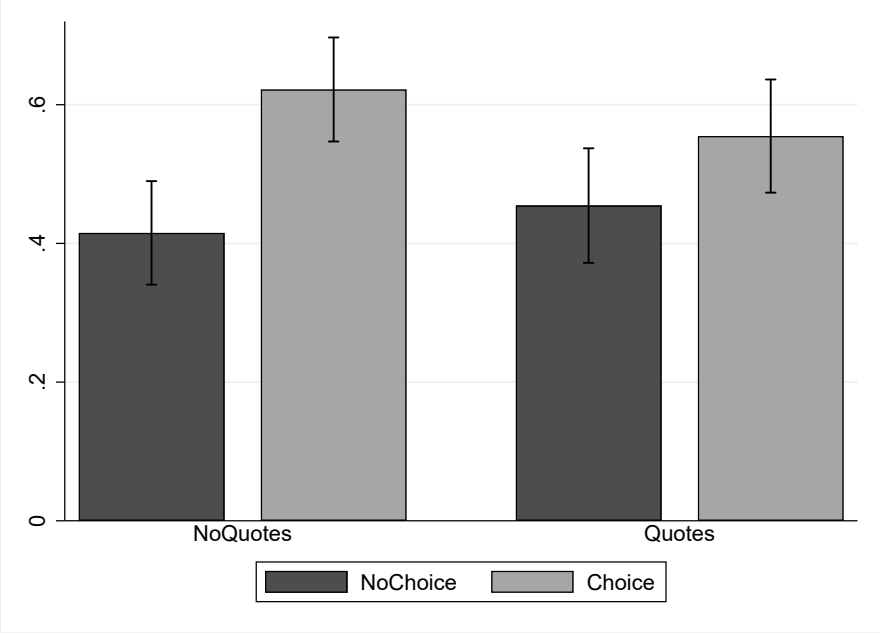
4 Results

Our empirical analysis first investigates the presence of motivated beliefs without communication. For this purpose, we study the difference between the NoChoice and Choice conditions in the NoChat treatment arms. We then analyze how communication with other participants holding, on average, the same motivated beliefs affects the prevalence of motivated beliefs. Hence, we estimate how the difference between the NoChoice and Choice conditions changes with communication. Finally, we explore what kind of communication environments are more or less susceptible to reinforcing motivated beliefs, i.e., how the difference-in-differences depends on the availability of plural opinions in the communication environment.

4.1 Motivated Beliefs without Communication

A prerequisite for our analysis is that our experimental intervention exogenously shifts the beliefs of Player B concerning Player A. We test this presumption by comparing the Choice and NoChoice conditions in the scenarios without communication. Figure 1 shows the average beliefs of Player B in the four treatments without communication. It demonstrates that we induce strongly motivated beliefs for individuals in the Choice conditions. In particular, the share of Players B with unfavorable beliefs about their Player A increases statistically significantly from 42% in NoChoice-NoChat to 62% in Choice-NoChat (Rank-sum test, $p < 0.01$). Similarly, the share of Players B with unfavorable beliefs about their Players A increases from 45% in NoChoice-NoChat-Quotes to 55% in Choice-NoChat-Quotes (Rank-sum test, $p = 0.09$). These effects are also economically sizable. In the treatments

Figure 1: Average Unfavorable Beliefs Without Communication



Notes: The figure reports for the NoChat treatments the fraction of Players B who believe that their Player A has chosen the unfair allocation. The dark gray bars report the fractions in the NoChoice and the light gray bars the fractions in the Choice conditions.

with NoQuotes, the share of unfavorable beliefs increases by roughly 20pp or 48% and in those with Quotes by roughly 10pp or 22%. Overall, the treatment intervention Choice thus increases the share of participants attributing negative intentions to their partner: it induces a self-serving negative perception of other participants.

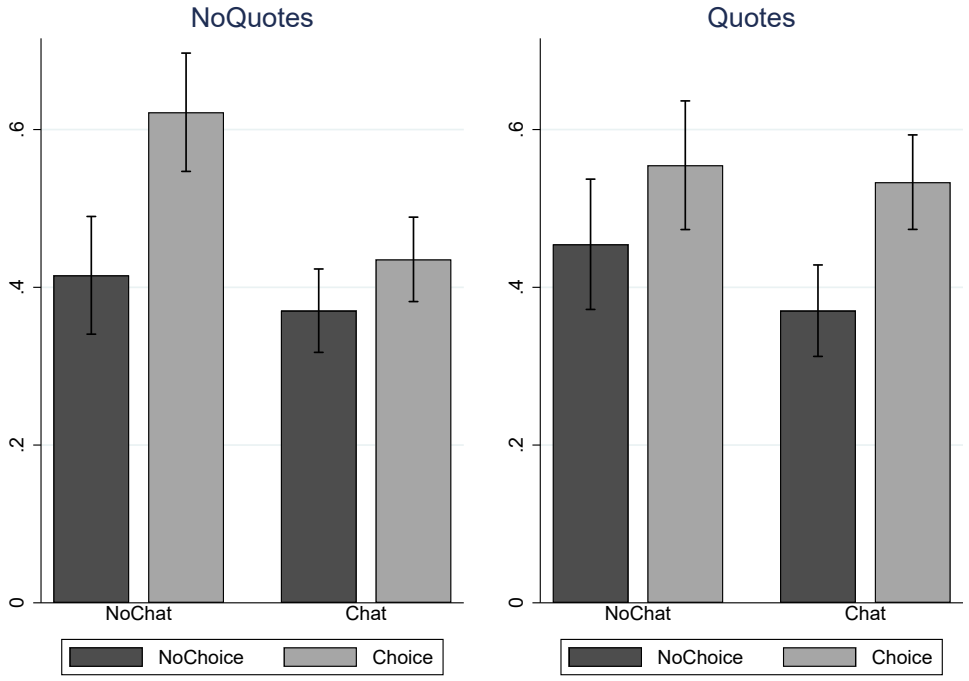
While the effect of the Choice conditions on beliefs differs slightly between the conditions with Quotes and NoQuotes, this difference is not statistically significant. More specifically, Rank-sum tests show that beliefs do neither differ significantly between Choice-NoChat and Choice-NoChat-Quotes nor between NoChoice-NoChat and NoChoice-NoChat-Quotes (Rank-sum test, $p = 0.23$ for the Choice conditions and $p = 0.48$ for the NoChoice conditions). If beliefs are formed in isolation, the opposing quotes we provide to participants thus do not appear to shift beliefs in one particular direction.

Result 1 *The availability of different choice options induces substantially motivated self-serving beliefs. This finding holds true irrespective of the existence of a plurality of previously declared opinions.*

While Result 1 summarizes the main insights from the treatments without communication, there are two more points worth noting. First, our data provides additional evidence that the CHOICE condition indeed induced motivated beliefs. In particular, our induction of motivated beliefs builds on the idea of Di Tella et al. (2015), who argue that individuals employ motivated beliefs in a self-serving manner to resolve a trade-off between self-gratification and social behavior. Our choice data are consistent with this argument. In both CHOICE conditions without communication, there is a strong correlation for Players B between holding an unfavorable belief about Player A's behavior and choosing an unfair allocation themselves. The respective Spearman correlations are 0.72 and 0.63 in CHOICE-NOCHAT and CHOICE-NOCHAT-QUOTES ($p < 0.01$ in both cases). Moreover, the self-serving nature of induced motivated beliefs also implies that the link between beliefs and decisions should be particularly pronounced for Players B that choose an unfair allocation. In NOCHAT, 90% of Players B that choose the unfair allocation indeed also state the belief that player A has chosen the unfair allocation. In contrast, only 77% of Players B who choose the fair allocation state a favorable belief. In line with the idea that individuals distort the beliefs in order to choose the unfair allocation, the difference between these shares is statistically significant ($p = 0.001$, Chi-squared test).

Second, while not being our main outcome of interest, we can also look at the realized behavior of Players A. Remember that from the perspective of Player A, all treatment conditions are identical. Across all treatments, the share of Players A that choose the unfair allocation is 51%. While this number is an interesting benchmark, the distance between the average beliefs of Players B and the average actions of Players A may be influenced by the shape of the underlying distribution of prior and posterior beliefs or by other cognitive biases. We, therefore, focus on the difference between the CHOICE and NOCHOICE conditions to measure the induced motivated beliefs of Players B.[9]

Figure 2: Average Beliefs of Player B



Notes: The figure reports for all treatments the fraction of Players B who believe that their Player A has chosen the unfair allocation. The dark gray bars report the fractions in the NoChoice and the light gray bars the fractions in the Choice conditions.

4.2 Communication and Motivated Beliefs

As argued in the previous section, participants in the Choice conditions without communication hold substantial motivated beliefs. The main objective of our experiment is to investigate to what extent communication among individuals with the same motivation affects these motivated beliefs. Our results show that the answer crucially depends on the communication environment. Consider first the NoQuotes conditions. The left panel of Figure 2 shows that participants in the Choice condition also hold motivated beliefs if they communicate: The fraction of participants reporting an unfavorable belief about their partner increases from 37% in NoChoice-Chat to 44% in Choice-Chat. However, this 7pp difference is less than

[9]With the caveats mentioned above in mind, we can investigate the differences between Players A behavior and Players B beliefs to compile suggestive evidence on the amount of uncertainty and the aggregate belief biases concerning the behavior of Player A. In line with Ging-Jehli et al. (2020), players with NoChoices seem to perceive the behavior of Players A slightly too positive, while subjects in the Choice conditions exhibit a too negative view on Players A.

half as large as the 20pp difference in the conditions without communication. Communication without external plural opinions – even among participants with similar belief distortions – thus attenuates rather than aggravates belief biases.[10] The regression analysis we present in Column (1) of Table 2 confirms this finding. Taking individual beliefs as the dependent variable, the coefficient of the difference-in-differences is negative and statistically significant ($p = 0.04$, t-test).

Result 2 *In the absence of external plural opinions, communication reduces motivated beliefs.*

Next, we consider the conditions with QUOTES. In line with the idea that they express opposing views, the QUOTES did not significantly affect individuals' average beliefs in the conditions without communication (Rank-sum tests, $p = 0.48$ for NOCHOICE-NOCHAT versus NOCHOICE-NOCHAT-QUOTES and $p = 0.23$ for CHOICE-NOCHAT versus CHOICE-NOCHAT-QUOTES). Importantly, however, in this environment, the attenuating effect of communication on motivated beliefs is nullified or even reversed (see the right panel of Figure 2). Remember that the difference in average beliefs of Players B between the CHOICE and NOCHOICE conditions without communication but with the additional QUOTES is 10pp. As discussed above, individuals in the CHOICE condition thus hold induced motivated beliefs if they form beliefs in isolation. If participants communicate, the difference between the CHOICE and the NOCHOICE conditions increases by 60% to 16pp. Hence, communication aggravates motivated beliefs in the environment with QUOTES. Accordingly, the coefficient of the difference-in-differences in Column (2) of Table 2 is positive. As it is imprecisely estimated, however, it turns out to be statistically insignificant ($p = 0.40$). We therefore conclude that the availability of plural opinions apart from the chat content nullifies or, if anything, even reverses the attenuating effect of communication on motivated beliefs.

Columns (3) and (4) speak to our second main research question, i.e., whether the effect of communication on motivated beliefs differs across communication environments with and

[10]Communication inherently requires individuals to deliberate on their beliefs and to articulate them. Additionally, it also exposes them to others' opinions. In Section 4.3.2, we demonstrate a strong correlation between the beliefs of individuals within chat groups. Therefore, the impact of communication extends beyond mere expression of opinions in our setup. Instead, the interactive nature of communication and the exposure to others' viewpoints leads participants to align their beliefs within chats.

Table 2: Regression Results Unfavorable Beliefs

	NoQuotes	Quotes	All	
	(1)	(2)	(3)	(4)
Choice	0.21***	0.10*	0.21***	0.21***
	(0.05)	(0.06)	(0.05)	(0.05)
Chat	-0.04	-0.08	-0.04	-0.04
	(0.05)	(0.05)	(0.05)	(0.05)
Chat × Choice	-0.14**	0.06	-0.14**	-0.14**
	(0.07)	(0.07)	(0.07)	(0.07)
Quotes			0.04	0.04
			(0.06)	(0.06)
Quotes × Choice			-0.11	-0.11
			(0.08)	(0.08)
Quotes × Chat			-0.04	-0.04
			(0.07)	(0.07)
Quotes × Chat × Choice			0.20**	0.20*
			(0.10)	(0.10)
Older Than 32 Years				-0.01
				(0.02)
Female				-0.08***
				(0.02)
Constant	0.42***	0.45***	0.42***	0.46***
	(0.04)	(0.04)	(0.04)	(0.04)
Number of Observations	992	829	1821	1821
adjusted R^2	0.03	0.02	0.02	0.03

Notes: The table reports the results of OLS regressions. The dependent variable is an indicator whether Player B holds the unfavorable belief on the behavior of Player A in all specifications. Column (1) analyses the NoQuotes conditions, column (2) the Quotes conditions, and columns (3) and (4) all data jointly. We report in parenthesis the standard errors clustered at the chat level for those who chat. Stars indicate significance at the 1%, 5%, and 10% level.

without external plural opinions. To answer this question, we compare the 6pp *increase* in motivated beliefs in the environment with QUOTES to the 13pp *decrease* in the environment with NOQUOTES. The coefficient of the corresponding triple interaction in Column (3) is positive and statistically significant ($p = 0.046$). Column (4) of Table 2 confirms that this result also holds if we control for the age and gender of participants. In fact, all coefficients literally remain identical after rounding to two digits.[11] Hence, communication environments with a plurality of previously declared opinions are significantly more conducive to the formation of motivated beliefs than communication in environments where such opinions are absent.

Result 3 *In the presence of external plural opinions, communication allows motivated beliefs to persist or even aggravates them. Furthermore, communication with like-minded partners is significantly more conducive to the formation of motivated beliefs in the presence of external plural opinions.*

4.3 Mechanisms

Result 3 documents that the characteristics of the communication environment are crucial in determining both the direction and magnitude of the impact of communication on motivated beliefs. This result immediately raises the question of how communication itself and participants' reactions to it differ between the environments with and without external plural opinions. In general, there are two possible reasons why the effect of communication on motivated beliefs may differ across the two environments. First, previously declared opinions may influence participants' attitudes toward communication and, thus, the content or tone of the chats. In particular, they might be more willing to express biased, unpopular beliefs if they can draw upon or refer to such declared opinions—facilitating communication that reinforces motivated beliefs. Second, available opinions besides the chat may change subjects' responses to the chat content. Specifically, external opinions may allow participants to uphold self-serving beliefs even when their chat partner communicates a contrary point of view. In line with Oprea and Yuksel (2022), a plurality of opinions may then facilitate

[11] In further regressions not reported here in detail, we fully interact the model with age or gender to test whether any of the coefficients above depends on these characteristics. None of the interactions is statistically significant ($p > 0.20$). While the statistical power of these regressions is lower, our coefficients of interest are again almost identical to the ones in Table 2.

selective updating such that participants adjust their beliefs more strongly when statements in the chat are consistent with their self-serving biases.

While we did not explicitly design our treatment variations to quantify the contribution of each of these factors to our results, the chat content provides indications for their relative importance. In the following analysis, we will investigate how communication differs between the conditions with and without available plural opinions. To improve statistical power, we pool all chat data from the conditions with Quotes and compare it to the pooled data from the conditions with NoQuotes.[12] In analogy to above, we drop all observations from chats in which at least one of the chat partners belongs to the fastest 15% of participants in their treatment.[13]

4.3.1 Willingness to Reveal Unfavorable Beliefs

We first study how the presence of a plurality of opinions affects the content and tone of the chats. We compare the chat content and tone across conditions with four different approaches, but we never find any significant differences between the scenarios with and without Quotes. First, there is no significant difference in the length of chats regarding the number of typed words ($p = 0.17$, Kolmogorov–Smirnov test).[14] Second, we hired two research assistants who coded the chat content in various aspects. Most importantly, they recorded whether a participant mentioned an unfavorable belief about the behavior of her Player A, and whether such an unfavorable belief was the first belief to be mentioned in the chat.[15] Note that these mentioned beliefs do not have to, and often do not, coincide with the incentivized stated beliefs after the chat. We find no significant difference in the number of times a participant

[12] We get qualitatively the same results when we do not pool the data but analyze the differences between the Quotes and NoQuotes conditions separately for the Choice and NoChoice conditions.

[13] Compared to the analysis in the previous section, this filtering eliminates data from further 136 participants, leaving 541 chats from 1082 participants. On average, each participant writes 4.9 messages or 34.9 words, yielding an average chat length of 9.9 messages or 69.8 words.

[14] The chats comprise, on average, 10.2 messages or 68.5 words in the NoQuotes treatments and 9.5 messages or 71.4 words in the Quotes treatments.

[15] Both research assistants first coded the chats independently, on the chat and individual levels. There was a high degree of agreement in the relevant variables between the two codings (the same coding in approximately 80% of all chats, Cramér's V above 0.7). Afterward, we asked them to provide a consolidated version of the coding by resolving any differences in their coding via discussion. For all analyses, we use this consolidated version of the coding.

mentions the unfavorable belief (34.3% for NoQuotes conditions vs. 36.5% for Quotes conditions, $p = 0.47$, Chi-squared test) or in the number of times the unfavorable belief is the first belief to be mentioned in a chat (34.0% for NoQuotes conditions vs. 37.7% for Quotes conditions, $p = 0.37$, Chi-squared test). Third, we defined two word lists, which include words indicating that a subject expresses a favorable or unfavorable belief (see Appendix D.1 for details). There are no significant differences in the number of chats that contain at least one word from one of the lists or from both lists across conditions with and without Quotes ($p = 0.48$ for words expressing a favorable belief, $p = 0.44$ for an unfavorable belief, $p = 1.00$ for both, Chi-squared tests). Fourth, we ran bigram and trigram analyses, capturing the chats' most frequently mentioned pairs and triples of words. After merging similar word combinations, we ended up with three topics – splitting fairly, thinking about others, and wishing good luck (for more details, see Appendix D.2). Although subjects seem slightly more likely to wish their chat partner good luck at the end of the chat in the NoQuotes treatments, we do not find substantial differences in the frequency with which the three topics are mentioned across the conditions with and without quotes ($p = 0.16$ for fair split, $p = 0.75$ for thinking about others, $p = 0.10$ for wishing good luck, Chi-squared tests). Each of these four analyses by itself, but in particular all of them jointly, strongly suggest that the presence of plural opinions neither substantially changed the content nor the tone of the chats.

Result 4 *The content or tone of the chats does not differ substantially between the conditions with and without external plural opinions.*

4.3.2 Immunity to Inconvenient Opinions

Next, we analyze how participants respond to the content of the chats when updating their beliefs. A critical measure in this regard is the correlation of elicited beliefs within chat groups. If participants did not react to the observed chat, their beliefs should be uncorrelated within a chat group. Instead, we find that beliefs are correlated within chat groups in all four Chat conditions. The Spearman correlation coefficients are 0.42 in the NoQuotes conditions and 0.24 in the conditions with Quotes ($p < 0.001$ for both correlations). Overall, the strong correlations between chat partners' beliefs show that communication induces an alignment of beliefs among chat partners.

Importantly, the correlation of chat partners' beliefs is weaker in the Quotes conditions than in the NoQuotes conditions. Confirming this observation, Table 3 shows linear regressions

Table 3: Regression Results Belief Correlations

	NoQuotes	Quotes	All	
	(1)	(2)	(3)	(4)
Other's Belief	0.42***	0.24***	0.42***	0.42***
	(0.05)	(0.06)	(0.05)	(0.06)
Quotes			0.12**	0.13**
			(0.05)	(0.05)
Other's Belief × Quotes			-0.18**	-0.18**
			(0.08)	(0.08)
Older than 32 Years				-0.12**
				(0.06)
Female				-0.06
				(0.06)
Constant	0.23***	0.35***	0.23***	0.32***
	(0.03)	(0.04)	(0.03)	(0.06)
Number of Observations	297	244	541	541
adjusted R^2	0.18	0.06	0.13	0.13

Notes: The table reports the results of OLS regressions. The dependent variable is an indicator whether Player B holds the unfavorable belief about the behavior of Player A in all specifications. Column (1) analyses the conditions with NoQuotes, Column (2) the conditions with Quotes, and Column (3) and (4) all data jointly. Other's Belief is an indicator whether the chat partner holds the unfavorable belief about the behavior of Player A. The control variables Older than 32 Years and Female are the averages within the respective chat group. There is no clustering of standard errors at the chat level because there is only one observation per chat. Stars indicate significance at the 1%, 5%, and 10% level.

with the subject's beliefs as the dependent variable. Indeed, columns (3) and (4) show that the existence of a plurality of external opinions significantly reduces the correlation between the chat partners' beliefs. The weakened link implies that participants are more likely to stick to their own beliefs rather than adopting their chat partner's point of view when they can also refer to or subscribe to alternative opinions.

The weaker correlations show that participants generally discount the chat content in the presence of plural opinions. However, this finding does not reveal why a communication

Table 4: Stated Beliefs Conditional on Chat Partner's Mentioned Beliefs

	NoQuotes	Quotes	All	
	(1)	(2)	(3)	(4)
Partner did not mention unfavorable belief	-0.41***	-0.29***	-0.41***	-0.41***
	(0.05)	(0.05)	(0.05)	(0.05)
Quotes			-0.03	-0.03
			(0.05)	(0.05)
Partner not mention unfav. belief × Quotes			0.12*	0.12*
			(0.07)	(0.07)
Older Than 32 Years				-0.04
				(0.03)
Female				-0.09***
				(0.03)
Constant	0.67***	0.65***	0.67***	0.74***
	(0.04)	(0.04)	(0.04)	(0.04)
N	594	488	1082	1082
adj. R^2	0.16	0.08	0.12	0.13
Quotes + Partner not ment. × Quotes = 0			0.018	0.030

Notes: The table reports the results of OLS regressions. The dependent variable is an indicator whether Player B holds the unfavorable belief about the behavior of Player A in all specifications. Column (1) analyses the conditions with NoQuotes, Column (2) the conditions with Quotes, and Column (3) and (4) all data jointly. We report in parenthesis the standard errors clustered at the chat level. Stars indicate significance at the 1%, 5%, and 10% level.

environment with external plural opinions is more susceptible to cause a reinforcement of motivated beliefs. To investigate this question further, we study whether information discounting is selective in a self-serving fashion. In particular, we analyze whether the Quotes enable participants to maintain their motivated beliefs even when their partner expresses an opinion contradicting their self-serving bias, we look at the chat content. Table 4 illustrates

the correlation between the beliefs a participant states in the incentivized elicitation following the chat and the opinions articulated by their partner during the conversation. Reiterating the findings from above, this correlation is significantly weaker in the QUOTES condition than in the NOQUOTES condition. Importantly, however, plural opinions seem to asymmetrically impede the connection between observed statements in the conversation and beliefs: If the chat partner's statements align with the motivated belief, plural opinions do not have a significant impact on beliefs, as indicated by the coefficients of QUOTES ($p > 0.5$, t-tests). In contrast, if the chat partner refrains from supporting the self-serving bias or even declares an opinion against the motivated belief, individuals are significantly more likely to uphold self-serving, negative beliefs in the presence of plural opinions ($p < 0.03$, post estimation Wald tests). Overall, the availability of plural opinions thus appears to disrupt the link between statements in the chat that correct self-serving biases and participants' ex-post beliefs.

Result 5 *In the presence of external plural opinions, participants are less responsive to communication. This insensitivity to communication is more pronounced if the chat content inconveniently does not coincide with their self-serving motivated beliefs.*

5 Conclusion

This paper presents a controlled online experiment studying the effects of communication on the formation and propagation of biased beliefs. Our findings reveal that communication reduces motivated beliefs even if communication takes place among individuals with the same motivated beliefs. However, this finding only holds in communication environments that are free of previously declared plural opinions to which individuals can refer and subscribe to. In the presence of plural opinions, communication allows motivated beliefs to persist. This finding highlights the important role of the communication environment in the proliferation of biases. Our evidence indicates that the availability of plural opinions enables individuals to selectively ignore information that does not support their desired self-serving beliefs.

Preserving a plurality of opinions is vital to the freedom of expression and, therefore, to a healthy democratic discourse. At the same time, our paper suggests that the presence of opinions supporting political views that are otherwise socially ostracized may also involve costs: They may create communication environments in which biased beliefs stemming from individuals' motivations can thrive and even become contagious. Therefore, our findings indicate

that social media platforms, and society in general, may have a reason to regulate certain statements, particularly extreme opinions based on fake news, that are counterproductive when forming factually correct public opinions.

An important advantage of our setting is that it allows exogenous manipulation of motivated beliefs and communication environment in a simple experimental paradigm. It is, therefore, well suited to study how other biases in beliefs spread through different forms of communication in social networks. Our paper focuses on the spread of motivated beliefs through communication in two-person chats. Extending this analysis to multi-person chats, forums, endogenously selected communication partners, other forms of belief bias, chatbots, or large language models seems to be a rich, largely unexplored, and important area for future research. Such future research will hopefully provide guidance on how to strike a balance between countering the spread of belief biases and preserving the plurality of opinions vital to a liberal society.

References

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election, *Journal of Economic Perspectives* **31**(2): 211–236.

Amelio, A. (2023). Social learning, behavioral biases and group outcomes, *working paper* .

Bakshy, E., Messing, S. and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook, *Science* **348**(6239): 1130–1132.

Barnidge, M. (2017). Exposure to political disagreement in social media versus face-to-face and anonymous online settings, *Political Communication* **34**(2): 302–321.

Barrera, O., Guriev, S., Henry, E. and Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics, *Journal of Public Economics* **182**: 104123.

Bartling, B., Valero, V., Weber, R. A. and Yao, L. (2022). Public discourse and socially responsible market behavior, *Working paper*, University of Zurich. Available at SSRN: https://ssrn.com/abstract=3677968.

Bénabou, R. (2015). The economics of motivated beliefs, *Revue d'économie politique* **125**(5): 665–685.

Bénabou, R. and Tirole, J. (2006). Belief in a just world and redistributive politics, *The Quarterly Journal of Economics* **121**(2): 699–746.

Benjamin, D. J., Brown, S. A. and Shapiro, J. M. (2013). Who is 'behavioral'? cognitive ability and anomalous preferences, *Journal of the European Economic Association* **11**(6): 1231–1255.

Bornstein, G. and Yaniv, I. (1998). Individual and group behavior in the ultimatum game: Are groups more "rational" players?, *Experimental Economics* **1**(1): 101–108.

Braghieri, L. (2022). Political correctness, social image, and information transmission, *Working paper*, Stanford University.

Bursztyn, L., Egorov, G. and Fiorin, S. (2020). From extreme to mainstream: The erosion of social norms, *American Economic Review* **110**(11): 3522–3548.

Bursztyn, L., Egorov, G., Haaland, I., Rao, A. and Roth, C. (2023). Justifying dissent, *The Quarterly Journal of Economics* **138**(3): 1403–1451.

Bursztyn, L., Rao, A., Roth, C. and Yanagizawa-Drott, D. (2023). Opinions as facts, *The Review of Economic Studies* **90**(4): 1832–1864.

Cason, T. N. and Mui, V.-L. (1998). Social influence in the sequential dictator game, *Journal of Mathematical Psychology* **42**(2-3): 248–265.

Chen, D. L., Schonger, M. and Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments, *Journal of Behavioral and Experimental Finance* **9**: 88–97.

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. and Starnini, M. (2021). The echo chamber effect on social media, *Proceedings of the National Academy of Sciences* **118**(9): e2023301118.

Cooper, D. J. and Kagel, J. H. (2005). Are two heads better than one? Team versus individual play in signaling games, *American Economic Review* **95**(3): 477–509.

Di Tella, R., Perez-Truglia, R., Babino, A. and Sigman, M. (2015). Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism, *American Economic Review* **105**(11): 4316–3442.

Drobner, C. (2022). Motivated beliefs and anticipation of uncertainty resolution, *American Economic Review: Insights* **4**(1): 89–105.

Engelmann, J., Lebreton, M., Schwardmann, P., van der Weele, J. J. and Chang, L.-A. (2019). Anticipatory anxiety and wishful thinking.

Enke, B., Graeber, T. and Oprea, R. (2023). Confidence, self-selection, and bias in the aggregate, *American Economic Review* **113**(7): 1933–1966.

Eyting, M. (2022). Why do we discriminate? The role of motivated reasoning, *SAFE Working paper 356*. Available at SSRN: https://ssrn.com/abstract=3819096.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. and Sunde, U. (2018). Global evidence on economic preferences, *The Quarterly Journal of Economics* **133**(4): 1645–1692.

Flynn, D. J., Nyhan, B. and Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics, *Political Psychology* **38**: 127–150.

Gentzkow, M. and Shapiro, J. M. (2011). Ideological segregation online and offline, *The Quarterly Journal of Economics* **126**(4): 1799–1839.

Ging-Jehli, N. R., Schneider, F. H. and Weber, R. A. (2020). On self-serving strategic beliefs, *Games and Economic Behavior* **122**: 341–353.

Goeree, J. K. and Yariv, L. (2011). An experimental study of collective deliberation, *Econometrica* **79**(3): 893–921.

Golman, R. (2023). Acceptable discourse: Social norms of beliefs and opinions, *European Economic Review* **160**: 104588.

Golman, R., Hagman, D. and Loewenstein, G. (2017). Information avoidance, *Journal of Economic Literature* **55**(1): 96–135.

Graeber, T., Roth, C. and Schesch, C. (2024). Explanations, *Working paper*, University of Bonn and University of Cologne, Germany.

Herz, H., Kistler, D., Zehnder, C. and Zihlmann., C. (2022). Hindsight bias and trust in government: Evidence from the united states, *CESifo Working Paper* (No. 9767): 1–56.

Huffman, D., Raymond, C. and Shvets, J. (2022). Persistent overconfidence and biased memory: Evidence from managers, *American Economic Review* **112**(10): 3141–75.

Jackson, M. O., Nei, S. M., Snowberg, E. and Yariv, L. (2023). The dynamics of networks and homophily, *Technical report*, National Bureau of Economic Research.

Kocher, M. G. and Sutter, M. (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games, *The Economic Journal* **115**(500): 200–223.

Kogan, S., Schneider, F. H. and Weber, R. A. (2021). Self-serving biases in beliefs about collective outcomes, *Working paper*, University of Zurich. Available at SSRN: https://ssrn.com/abstract=3819096.

Levy, G. and Razin, R. (2019). Echo chambers and their effects on economic and political outcomes, *Annual Review of Economics* **11**: 303–328.

Loury, G. C. (1994). Self-censorship in public discourse: A theory of "political correctness" and related phenomena, *Rationality and Society* **6**(4): 428–461.

Masser, B. and Phillips, L. (2003). "What do other people think?"—The role of prejudice and social norms in the expression of opinions against gay men, *Australian Journal of Psychology* **55**(3): 184–190.

McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001). Birds of a feather: Homophily in social networks, *Annual review of sociology* **27**(1): 415–444.

Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence, *Psychological Review* **115**(2): 502.

Morris, S. (2001). Political correctness, *Journal of Political Economy* **109**(2): 231–265.

Müller, M. W. (2022). Selective memory around big life decisions, *Working paper*.

Oprea, R. and Yuksel, S. (2022). Social exchange of motivated beliefs, *Journal of the European Economic Association* **20**(2): 667–699.

Schotter, A. (2003). Decision making with naive advice, *American Economic Review* **93**(2): 196–201.

Schwardmann, P., Tripodi, E. and Van der Weele, J. J. (2022). Self-persuasion: Evidence from field experiments at international debating competitions, *American Economic Review* **112**(4): 1118–1146.

Sprengholz, P., Henkel, L., Böhm, R. and Betsch, C. (2023). Historical narratives about the covid-19 pandemic are motivationally biased, *Nature* **623**: 588–593.

Stoner, J. A. F. (1961). *A comparison of individual and group decisions involving risk*, PhD thesis, Massachusetts Institute of Technology.

Sunstein, C. R. (2017). *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press, Princeton.

Teger, A. I. and Pruitt, D. G. (1967). Components of group risk taking, *Journal of Experimental Social Psychology* **3**(2): 189–205.

Thaler, M. (2023a). The supply of motivated beliefs, *working paper* .

Thaler, M. (2023b). The "fake news" effect: An experiment on motivated reasoning and trust in news, *American Economic Journals: Microeconomics, forthcoming* pp. 1–50.

Verbrugge, L. M. (1977). The structure of adult friendship choices, *Social forces* **56**(2): 576–597.

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

Appendix A Randomization Checks

This section presents the randomization checks for all treatment conditions. To ease their exposition, we binarize several variables. Concerning gender, 2% of our participants report to be neither male nor female. When creating the dummy variable female, we pool those observations with those reporting to be male. Second, we generate a dummy variable "old" that is equal to one for all participants reporting an age at or above the median age of 33 years. Third, we create a dummy variable "No university degree" that is zero if and only if the highest reported degree is a high school degree. Fourth, we generate a dummy "Daily Social Media Use" that is equal to one if and only if the participant reports using social media daily.

Table 5 contains the summary statistics of Players B in all treatments. It reports the p-values referring to Chi-squared tests concerning the null hypothesis of no differences across (i) the four treatments in the NoQuotes condition (p^{NQ}) (ii) the four treatments in the Quotes condition (p^{Q}), and (iii) the eight treatments in both conditions (p^{all}). The p-values p^{Q} and p^{NQ} show that we have an overall well-balanced sample within the Quotes and NoQuotes conditions. The p^{all}-values show significant differences across all treatments only in gender and age. These differences arise because participants in the Quotes conditions are older and less likely to be female. As explained in Section 3, we conducted the experiment in two waves: All conditions with NoQuotes were conducted between the 30th of November and 16th of December 2021, and all conditions with Quotes between the 5th and 13th of December 2022. This timing may have created the differences in the characteristics across waves. However, most of our analyses are within waves, so these differences do not affect the corresponding results. For the few analyses containing data from both waves, we always present the raw treatment differences and specifications controlling for age and gender to account for these wave-specific differences in our empirical analysis.

Table 5: Summary Statistics

	NoQuotes					Quotes					
	NoChoice		Choice			NoChoice		Choice			
	NoChat	Chat	NoChat	Chat	p^{NQ}	NoChat	Chat	NoChat	Chat	p^{Q}	p^{all}
Socio-Demographics											
Female	0.51	0.55	0.49	0.59	0.13	0.43	0.41	0.43	0.40	0.89	0.00
Older Than 32 Years	0.46	0.49	0.54	0.47	0.48	0.61	0.67	0.55	0.59	0.09	0.00
No University Degree	0.38	0.40	0.37	0.43	0.56	0.41	0.41	0.42	0.35	0.37	0.63
Social Media and Peers											
Daily Social Media Use	0.49	0.49	0.48	0.50	0.99	0.50	0.49	0.51	0.53	0.82	0.98
Creates Content on Social Media	0.25	0.30	0.27	0.28	0.63	0.24	0.34	0.30	0.29	0.21	0.41
Shares Political Views on Social Media	0.26	0.30	0.27	0.25	0.44	0.23	0.29	0.19	0.29	0.09	0.23
Similar Political Orientation as Friends	0.73	0.76	0.77	0.74	0.74	0.77	0.72	0.77	0.76	0.52	0.83
Number of Observations	171	324	164	333		143	270	146	270		

Notes: The table reports summary statistics for Players B in all treatments. The p-values p^{NQ}, p^{Q}, and p^{all} refer to Chi-squared tests testing whether the distributions of variables are identical across the NoQuotes conditions, across the Quotes conditions, and across all conditions, respectively.

Appendix B Robustness Check Using Full Sample

Table 6: Regression Results Unfavorable Beliefs (Full Sample)

	NoQuotes	Quotes	All	
	(1)	(2)	(3)	(4)
Choice	0.20***	0.11**	0.20***	0.20***
	(0.05)	(0.05)	(0.05)	(0.05)
Chat	-0.04	-0.09*	-0.04	-0.04
	(0.05)	(0.05)	(0.05)	(0.05)
Chat × Choice	-0.11*	0.06	-0.11*	-0.11*
	(0.06)	(0.07)	(0.06)	(0.06)
Quotes			0.03	0.03
			(0.05)	(0.05)
Quotes × Choice			-0.08	-0.09
			(0.07)	(0.07)
Quotes × Chat			-0.04	-0.05
			(0.07)	(0.07)
Quotes × Chat × Choice			0.18*	0.17*
			(0.10)	(0.10)
Older Than 32 Years				-0.01
				(0.02)
Female				-0.06***
				(0.02)
Constant	0.42***	0.45***	0.42***	0.45***
	(0.04)	(0.04)	(0.04)	(0.04)
Number of Observations	1163	974	2137	2137
adjusted R^2	0.02	0.02	0.02	0.03

Notes: The table reports the results of OLS regressions using the full sample without excluding the fastest 15%. The dependent variable is an indicator whether Player B holds the unfavorable belief on the behavior of Player A in all specifications. Column (1) analyses the NoQuotes conditions, column (2) the Quotes conditions, and columns (3) and (4) all data jointly. We report in parenthesis the standard errors clustered at the chat level for those who chat. Stars indicate significance at the 1%, 5%, and 10% level.

Appendix C Results on Alternative Belief Measures

This section reports results on two alternative measures of beliefs that we elicited in the experiment. While the regressions based on the alternative measures exhibit somewhat lower statistical power, all results from the main text also hold directionally for both of them.

Appendix C.1 General Beliefs

Next to our main outcome variable, we also elicited a "general" belief for Player B. In particular, we asked Players B how many of the previous 100 Players A picked the unequal option. If the answer deviated by at most 5 in absolute terms from the correct answer, the subject earned £2.50. Note that this belief is not immediately related to the specific Player A that is matched to Player B, so the motivation to distort this belief to justify self-gratification instead of social behavior is considerably smaller than for the specific belief that we discuss in the main text (Di Tella et al.; 2015). Consistent with this idea, we find qualitatively identical but overall considerably weaker patterns when repeating the analysis of the main text with this general belief as the dependent variable.

To be more specific, Table 7 summarizes the findings of our main specifications when we use the general instead of the specific belief as the dependent variable. First, when analyzing the formation of beliefs in isolation, the availability of Choice options induces motivated beliefs in both the Quotes and NoQuotes conditions. However, only the increase in average beliefs in the Quotes conditions is statistically significant, see Columns (1) and (2). Second, we find that communication via Chats slightly reduces induced motivated beliefs in the NoQuotes conditions and slightly aggravates them in the Quotes conditions. Both coefficients of the corresponding interaction terms turn out to be statistically insignificant, see Columns (3) and (4). Controlling for age and gender has very little effect on our coefficients of interest. Overall, the results for our general beliefs are qualitatively identical to our main results but considerably weaker. This difference reiterates previous findings from the literature. For example, Di Tella et al. (2015) also find substantially weaker effects when considering a general belief instead of a specific belief. These common findings also align with the broader idea underlying motivated beliefs: To justify self-gratification, it is convenient to believe that the matched Player A chose unfairly, while it is not essential to believe that the entire population of Players A did so.

Table 7: Regression Results General Unfavorable Beliefs

	NoQuotes	Quotes	All	
	(1)	(2)	(3)	(4)
Choice	3.91	4.96*	3.91	3.88
	(2.93)	(2.97)	(2.93)	(2.94)
Chat	-7.39***	-3.31	-7.39***	-7.08**
	(2.75)	(2.86)	(2.75)	(2.75)
Chat × Choice	-0.24	1.76	-0.24	-0.01
	(3.82)	(3.94)	(3.82)	(3.81)
Quotes			-1.93	-2.20
			(3.03)	(3.03)
Quotes × Choice			1.05	1.00
			(4.17)	(4.17)
Quotes × Chat			4.08	3.71
			(3.96)	(3.94)
Quotes × Chat × Choice			2.00	1.69
			(5.49)	(5.46)
Older than 32 Years				-1.26
				(1.25)
Female				-6.12***
				(1.24)
Constant	58.64***	56.71***	58.64***	62.33***
	(2.10)	(2.18)	(2.10)	(2.30)
Number of Observations	992	829	1821	1821
adjusted R^2	0.02	0.01	0.02	0.03

Notes: The table reports the results of OLS regressions of treatment differences in the general beliefs of Players B. Column (1) analyses the NoQuotes conditions, column (2) the Quotes conditions, and columns (3) and (4) all data jointly. We report in parenthesis the standard errors clustered at the chat level for those who chat. The stars indicates significance at the 1%, 5%, and 10% level, respectively.

Appendix C.2 Constructed Continuous Beliefs

In the conditions with QUOTES, we included an additional question after the elicitation of beliefs asking how certain the participant was about the stated binary belief regarding the decision of Player A (in %). Building on the answer to this question and the binary belief, we can construct a continuous belief indicating the likelihood with which Player B believes that Player A has chosen the unfair option. In particular, we define the continuous belief to be equal to the certainty measure in case of a stated unfavorable belief, and 100 minus the certainty measure in case of a stated favorable belief.

Table 8 shows regression results where the dependent variable is this constructed continuous belief. We report specifications analogous to the analyses depicted in Tables 2 and 7. All results are qualitatively the same as the ones for the QUOTES treatment in the main text (cp. column (2) of Table 2). First, the availability of CHOICE options induces motivated beliefs. Second, communication in the presence of QUOTES allows motivated beliefs to persist or even aggravates them.

Table 8: Regression Results Continuous Unfavorable Beliefs

	Quotes	
	(1)	(2)
Choice	5.58*	5.35*
	(3.21)	(3.21)
Chat	-6.14**	-5.95*
	(3.08)	(3.08)
Chat × Choice	5.78	5.65
	(4.36)	(4.35)
Older than 32 Years		-4.05**
		(2.02)
Female		-2.84
		(2.03)
Constant	51.03***	54.72***
	(2.31)	(2.65)
Number of Observations	829	829
adjusted R^2	0.03	0.03

Notes: The table reports the results of OLS regressions of treatment differences in the constructed continuous beliefs of Players B in the Quotes conditions. We report in parenthesis the standard errors clustered at the chat level for those who chat. The stars indicate significance at the 1%, 5%, and 10% levels, respectively.

Appendix D Details on Code-Based Chat Analysis

We use two code-based approaches to analyze the chat content. In Appendix D.1, we construct word lists to capture the content of the chats. In Appendix D.2, we analyze the frequency of the most common bigrams and trigrams across treatments.

Appendix D.1 Results on Word Lists

In the first code-based approach, we defined word lists indicating that a subject expresses a favourable or an unfavorable belief about the behavior of Player A. The word list to quantify favourable beliefs contains the following words: "fair", "fairly", "equal", "equally", "even", "evenly", "generous", "nice", "half", "kind", "split", "good", "hope". In analogy, the word list indicating unfavorable beliefs contains the words: "unfair", "unfairly", "greedy", "selfish", "keep", "kept", "take", "himself", "herself", "themselves", "bad". Table 9 reports the absolute and relative frequencies of both topics in the chats, conditional on whether or not participants observe the Quotes prior to the chat. There are no significant differences in these frequencies across the conditions with and without quotes ($p = 0.48$ for words expressing a favorable belief, $p = 0.44$ for an unfavorable belief, $p = 1.00$ for both, Chi-squared tests).

Table 9: Distribution of Manual Topics in Chats

	Topic(s)			
Treatment	Fair	Unfair	Both	Number of Chats
NoQuotes	265 [89.23%]	189 [63.64%]	169 [56.90%]	297
Quotes	212 [86.89%]	164 [67.21%]	139 [56.97%]	244
p-value	0.48	0.44	1.00	

Notes: The table reports the distribution of manually defined topics conditional on Quotes. Relative frequencies per treatment reported in brackets. p-values from Chi-squared tests.

Appendix D.2 Results on Bigram and Trigram Analysis

In order to quantify the tone and content of the chats, we also ran bigram and trigram analyses. These analyses capture the most frequently mentioned pairs and triples of words in the chats (after deleting punctuation and stop words such as articles, prepositions, etc.). The top 11 (due to a tied 10th place) combinations are reported in Table 10 in descending order of frequency. Because the single trigram "think participant chose" contains the frequent bigram "think participant", we exclude this trigram from our analyses. After merging similar bi-/trigrams such as "50 50" and "even split" or "think participant" and "think people", we ended up with three topics – the fair split, thinking about others, and wishing good luck (see the last column of Table 10). Table 11 reports the absolute and relative frequencies of the topics in the chats, conditional on the availability of external plural opinions. Although subjects seem slightly more likely to wish their chat partner good luck at the end of the chat in the NoQuotes treatments, we do not find substantial differences in the frequency with which the three topics are mentioned across the conditions with and without quotes ($p = 0.16$ for fair split, $p = 0.75$ for thinking about others, $p = 0.10$ for wishing good luck, Chi-squared tests).

Table 10: Bigram and Trigram Analyis of Chat Content – Topics

Bigram / Trigram	Count	Topic
50 50	142	Fair split
think participant	137	Thinking about others
participant chose	90	Thinking about others
think people	60	Thinking about others
good luck	56	Wishing good luck
split evenly	55	Fair split
even split	46	Fair split
think chose	46	Thinking about others
think participant chose	44	Thinking about others
chose 50	40	Fair split
think would	40	Thinking about others

Notes: The table shows the ten most frequent bigrams and trigrams that occur in the chats across all treatments. Column "Count" refers to the number of chats in which the bi-/trigram occurs. Column "Topic" refers to the topic assignment that was done manually after the identification of the bi-/trigrams.

Table 11: Bigram Analyis of Chat Content – Distribution of Topics

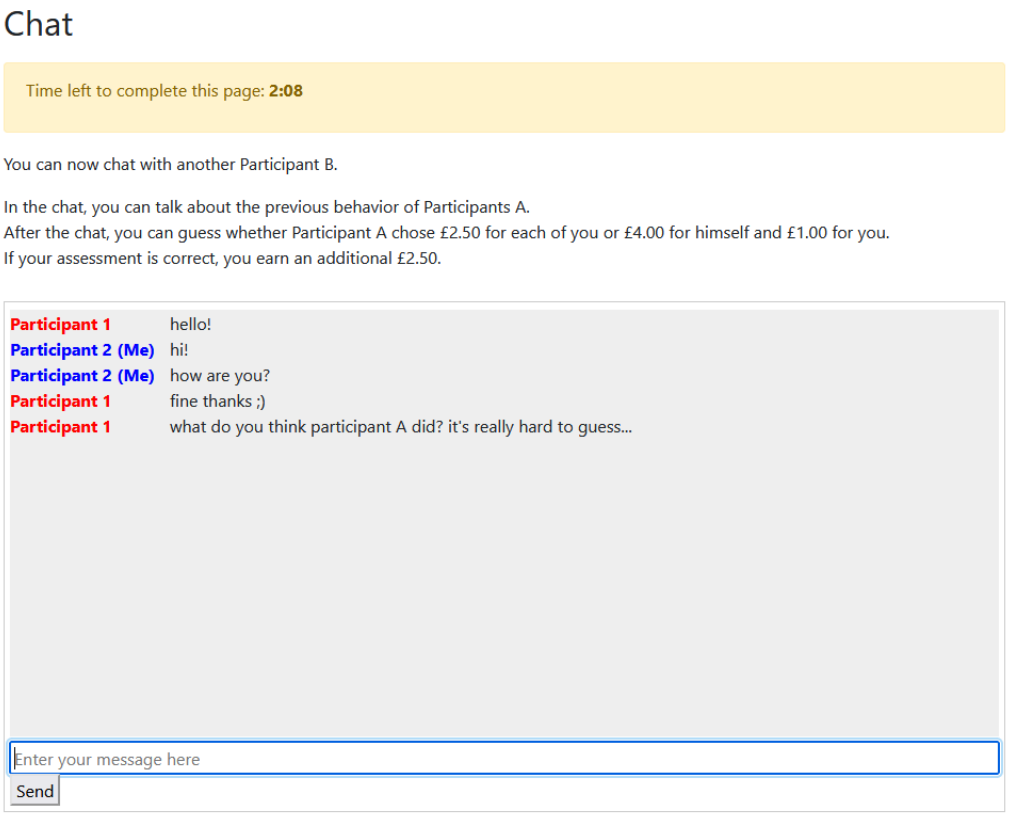
	Topic			
Treatment	Fair split	Thinking about others	Wishing good luck	Number of Chats
NoQuotes	164 [55.22%]	207 [69.70%]	37 [12.46%]	297
Quotes	119 [48.77%]	166 [69.03%]	19 [7.79%]	244
p-value	0.16	0.75	0.10	

Notes: The table reports the distribution of topics from bigrams conditional on Quotes. Relative frequencies per treatment reported in brackets. Topics include all bigrams listed in Table 10. p-values from Chi-squared tests.

Online Appendix

Appendix E Instructions and Screenshots

Figure 3: Screenshot of Chat Window



Notes: Chat screen of Players B with exemplary text. Participants had three minutes to chat with their partner.

[All Participants]

Welcome

Welcome to this online experiment, and thank you for your participation.
The experiment consists of two parts.
In each part, you can earn money.
There is no base payment, but we will ensure that you earn at least £2.50 in the experiment.
You will receive your payment within the next couple of days.

Instructions

This screen contains the instructions for Parts 1 and 2 of the experiment.
Please read them carefully. You will find three comprehension questions on the next screen.
You can only start with the experiment if you answer them correctly.
If at least one answer is wrong, you will drop out of the experiment.

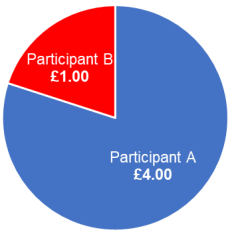
Roles

Participants in this experiment are randomly matched into fixed pairs.
There is one Participant A and one Participant B in each pair.
You will learn whether you are Participant A or B after reading the instructions.

Part 1

In Part 1 of the experiment, **Participant A** decides about the allocation of £5.00 between Participant A and B.
Participant A has two options:

Option 1



Option 2

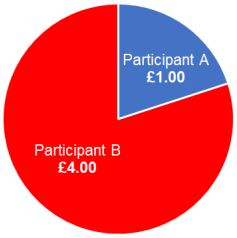
Hence, Participant A can either take **£4.00** for himself and leave **£1.00** for Participant B.
Or he can allocate **£2.50** to each participant in your group.

Part 2

In Part 2 of the experiment, **Participant B** decides about the allocation of another £5.00 between Participant A and B.
However, the options that Participant B will face when he makes this decision are not certain yet.

In 50 % of cases, Participant B will face the same options as Participant A in Part 1.
Hence, he can choose between:

Option 1



Option 2

In this case, Participant B can either take **£4.00** for himself and leave **£1.00** for Participant A.
Or he can allocate **£2.50** to each participant in your group.

In the remaining 50% of cases, Participant B cannot decide between any options.

Certain Allocation

In this case, the allocation will thus be **£2.50** to each participant in your group.

The respective money amount in GBP from both parts will be transferred to the Prolific accounts of Participants A and B.

Information
Which options Participant B faces is determined at the beginning of Part 2.
Therefore, when making his decision, **Participant A does not know the options of Participant B**.

When making his decision, **Participant B does not know the decision of Participant A** in Part 1 of the experiment.

Comprehension Questions *[correct answer marked with *]*

When making his decision, does Participant A know the options of Participant B in Part 2?

- Yes
- No (*)

When making his decision, does Participant B know the decision of Participant A in Part 1?

- Yes
- No (*)

What are the options of Participant A in Part 1 of the experiment?

- £4.00 for himself and £1.00 for Participant B OR you both get £2.50. (*)
- He has no choice, the allocation is £2.50 for you both.
- It is not certain yet.

[Participant A]

Your Role

You answered all comprehension questions correctly.

We will now start with Part 1 of the experiment.
The computer has determined that you will be in the role of **Participant A** in this experiment.

Your Allocation Decision

You can now decide how to split the £5.00 between you and Participant B of your group.
How do you want to allocate the £5.00?

- £4.00 for myself and £1.00 for Participant B
- £2.50 for myself and £2.50 for Participant B

Further Questions

Next, we would like to ask you two short questions.
We ask you for your opinion. There is no right or wrong answer here.

Question 1
Participant B will, similar to you, decide how to split £5.00 between you and himself.
Suppose that Participant B can select between two options.
The first option is a fair split of £2.50 for each of you two.
The alternative option assigns £4.00 to Participant B and £1.00 to you.
What do you think: What would Participant B in your group choose in this case?

- £2.50 for Participant B himself and £2.50 for me
- £4.00 for Participant B himself and £1.00 for me

Question 2
We will also ask Participant B what he thinks about your choices.
What do you think: What will Participant B think about your choice?

- Participant B thinks that I chose £4.00 for myself and £1.00 for him
- Participant B thinks that I chose £2.50 for myself and £2.50 for him

---

[Participant B]

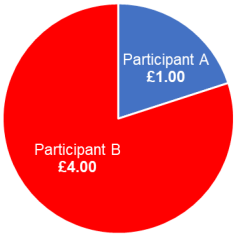
Your Role

You answered all comprehension questions correctly.

The computer has determined that you will be in the role of **Participant B** in this experiment.
Participant A in your pair has already made his decision in Part 1.
We will now start with Part 2 of the experiment.

[Choice] The computer has further determined that you will decide between the following two options:

Option 1



Option 2

Please consider your options carefully.

[NoChoice] The computer has further determined that you cannot choose the allocation.
The allocation will thus be a fair split:

Certain Allocation

Chat

Before the experiment continues, you can chat with another participant of this experiment.
This participant in the chat is also playing in the role of Participant B.

In the chat, you can talk about the previous behavior of Participants A.
After the chat, you can guess whether Participant A chose £2.50 for each of you or £4.00 for himself and £1.00 for you.
If your assessment is correct, you earn an additional £2.50.

Please do not reveal your identity in the chat and do not use offensive language.

[Quotes & Chat] To start the conversation and to give you some food for thought, here are two quotes by famous personalities:

I think we are living in selfish times.
— Javier Bardem, Hollywood actor and Oscar winner

I'm just thankful I'm surrounded by good people.
— Jon Pardi, singer and songwriter

Food for Thought *[Quotes & NoChat]*

On the next screen, you can guess whether Participant A chose £2.50 for each of you or £4.00 for himself and £1.00 for you.
If your assessment is correct, you earn an additional £2.50.

To give you some food for thought, here are two quotes by famous personalities:

I think we are living in selfish times.
— Javier Bardem, Hollywood actor and Oscar winner

I'm just thankful I'm surrounded by good people.
— Jon Pardi, singer and songwriter

Your Allocation Decision

[Choice] You can now decide how to split the £5.00 between you and Participant A of your group.

How do you want to allocate the £5.00?

- £4.00 for myself and £1.00 for Participant A
- £2.50 for myself and £2.50 for Participant A

[NoChoice] The computer has determined that you cannot choose the allocation.
The allocation will thus be **£2.50 for yourself and £2.50 for Participant A**.

Assessment

Please assess the previous behavior of the Participant A in your group.
If your assessment is correct, you receive a bonus of £2.50.

What do you think: What was the choice of Participant A in your group?

- £2.50 for Participant A himself and £2.50 for me
- £4.00 for Participant A himself and £1.00 for me

Please also assess the previous behavior of the last 100 Participants A in this experiment.
If your answer deviates by at most 5 from the truth, you receive a bonus of £2.50.

What do you think: How many of the last 100 Participants A have chosen to take £4.00 for themselves rather than choosing £2.50 for both players?

Note: Please move the slider before continuing

Assessment

On the last screen you stated that you think that Participant A in your group took £2.50 for himself and left £2.50 for you.
How likely do you think that this is true?

Note: Please move the slider before continuing

[All Participants]

Survey

As the final part of this experiment, we would like to ask you a few questions.

Did you understand the rules of the experiment?

- I understood them fully
- I understood them almost fully

- I understood them only partly
- I did not understand them

How old are you?

What is your gender?

- female
- male
- other

What is your highest educational degree?

- No degree
- High school
- Bachelor
- Master
- PhD

If you go/went to university, what is/was your major?

- Not applicable
- Economics
- Law
- Psychology
- Political sciences
- Medicine
- Natural sciences
- Engineering
- Other social sciences
- Other

How often do you use social media (Facebook, Twitter, ...)?

- Not at all
- Rarely
- Regularly
- Daily

Do you share your political views on social media?

- Yes
- No

Do you actively create content on social media?

- Yes
- No

Do you and your closest friends share the same political orientation?

- Yes
- No