Dohmen, Thomas; Quercia, Simone; Willrodt, Jana

Working Paper

A note on salience of own preferences and the consensus effect

ECONtribute Discussion Paper, No. 219

Provided in Cooperation with:
Reinhard Selten Institute (RSI), University of Bonn and University of Cologne

Suggested Citation: Dohmen, Thomas; Quercia, Simone; Willrodt, Jana (2023) : A note on salience of own preferences and the consensus effect, ECONtribute Discussion Paper, No. 219, University of Bonn and University of Cologne, Reinhard Selten Institute (RSI), Bonn and Cologne

This Version is available at:
https://hdl.handle.net/10419/278395

ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 219

A Note on Salience of Own Preferences and the Consensus Effect

Thomas Dohmen Simone Quercia Jana Willrodt

February 2023 www.econtribute.de

UNIVERSITÄT BONN Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866 is gratefully acknowledged.

A note on salience of own preferences and the consensus effect*

Thomas Dohmen, Simone Quercia, and Jana Willrodt

February 17, 2023

Abstract

In this paper, we hypothesize that the strength of the consensus effect, i.e., the tendency for people to overweight the prevalence of their own values and preferences when forming beliefs about others' values and preferences, depends on the salience of own preferences. We manipulate salience by varying the order of elicitation of preferences and beliefs. Although our results confirm the existence of the consensus effect, we find no evidence of a difference between the two orders of elicitation. While our results highlight the robustness of the consensus effect, they also indicate that salience does not mediate the strength of this phenomenon.

Keywords: Consensus effect, social preferences, trust game, beliefs

JEL Classification Numbers: C91, D01, D83, D91.

*Dohmen: University of Bonn, IZA Institute of Labor Economics, Maastricht University (Email: *tdohmen@uni-bonn.de*); Quercia: University of Verona (Email: *simone.quercia@univr.it*); Willrodt (corresponding author): Düsseldorf Institute for Competition Economics (DICE) (Email: *willrodt@dice.hhu.de*). We thank Armin Falk, Dirk Engelmann, and Hans-Theo Normann as well as participants of the Thurgau Experimental Economics Meeting 2017, M-BEES 2018, ESA World Meeting 2018, and SABE-IAREP Conference 2018 for helpful comments and discussion. Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project A01) and under Germany's Excellence Strategy – EXC 2126/1 – 390838866 is gratefully acknowledged.

1 Introduction

The consensus effect, which refers to an egocentric tendency in assessing and predicting others' actions, values or preferences, is a widely documented cognitive bias. It is robustly found – typically identified empirically as a correlation between an individual's own values and preferences and the belief about the corresponding values and preferences in others – in psychology (see, e.g., Mullen et al., 1985) and in economics (see, e.g., Blanco et al., 2014). While the idea that individuals project their own attributes onto others had been at the core of influential theories in psychology (Cattell, 1944; Heider, 1958; Jones and Nisbett, 1987), Ross et al. (1977) attributed this phenomenon to a systematic distortion in the processing of information.[1] Although researchers are aiming ever since to identify factors that influence the consensus effect, the effect of salience and focus of attention on own attributes, which is an obvious driver of systematic distortion in information processing according to recent models in economics, has not been fully scrutinized. These recent economics models postulate that people tend to focus only on portions of the environment that are salient to them and tend to overweight those salient portions compared to others (see Bordalo et al., 2022, for a review). In the context of the consensus effect, this suggests that the salience of own preferences and values determines how much people focus on their own values and preferences and consequently project them onto others. Nevertheless, no study has tested whether the consensus effect depends on the salience of own preferences.

In this paper, we test the hypothesis that changes in the salience of one's own preference type affect the strength of the consensus effect in a laboratory experiment, in which we manipulate salience by exogenously varying the order of elicitation of preferences and beliefs in a binary trust game. In this game, first movers decide whether to transfer money to a second mover, thereby exposing themselves to a socially risky situation, or not to transfer money and keep the money as a safe payoff. If a first mover sends money, a second mover can decide to reciprocate by splitting the efficiency gains from trust or to return an amount that leaves the first mover with less than they would have obtained if they had not sent money. As second-mover actions do not involve strategic uncertainty, we interpret them as a measure of preferences. Additionally, we measure beliefs about other second movers' strategies. Varying the order of elicitation of preferences and beliefs allows

[1] Ross et al. (1977) labelled the evidence that people "see one's own behavioral choices and judgments as relatively common and appropriate to existing circumstances while viewing alternative responses as uncommon, deviant, or inappropriate" (p. 280) the false consensus effect. Whether the consensus effect is "false" can be argued because it may be rational for an individual to take information about themselves into account when making inferences about a population they are part of (Dawes, 1989).

us to assess whether the consensus effect is stronger when preferences are elicited before beliefs and hence more salient for participants.

Investigating this question does not only contribute to a better understanding of the drivers of the consensus effect, but it also creates insights into the process of belief formation and its determinants. Therefore, our results are informative for economic theory and for policy. Whether beliefs depend on the salience of preferences is relevant for theories of belief formation, not least because salience of own preferences is typically not accounted for in models. One reason might be that agents are assumed to know their preferences so that their type might always be salient to them. Salience of own preferences might become particularly important when people are uncertain about their own preferences, i.e., do not revert to their preferences easily without (costly) introspection or consideration of relevant trade-offs that reveal these preferences. By learning about their own preferences through choice, salience of their own type would increase, with potential repercussions for belief formation. Moreover, from a policy point of view, if the consensus effect is affected by the salience of own preferences, situations in which individuals are frequently primed regarding themselves or their identities may favor egocentric thinking and foster a stronger distortion in beliefs. Finally, from an experimental point of view, knowing whether the salience of preferences affects belief formation is crucial for the design of experiments. If variation in the salience of own type affects belief formation, methods to elicit beliefs have to standardize the degree of salience of an individual's own type.

Our results confirm the existence of the consensus effect but, in contrast to our hypothesis, the findings show that a variation of the degree of salience of own preferences does not affect the strength of the consensus effect. In fact, we document a significant consensus effect for both orders of elicitation and its size is statistically indistinguishable between the two orders. We conclude that the consensus effect is robust to different elicitation orders and that the salience of own type likely plays a minor role in determining the strength of this phenomenon. We discuss the implications of these findings in our concluding section.

Our paper contributes to the extensive literature on the consensus effect (see Mullen et al., 1985, who provide an influential meta-study of 115 studies and Bazinger and Kühberger, 2012, for a more recent overview). The first study that explicitly examines the consensus effect in economics was conducted by Offerman et al. (1996). It provides evidence for the consensus effect in public goods games.[2] Engelmann and Strobel (2000) test the existence of the consensus effect in a wide

[2] Correlations between preferences and beliefs in social dilemmas have been documented first in papers whose main objective was not to investigate the consensus effect (Jacobsen and Sadrieh, 1996; Selten and Ockenfels, 1998; Charness and Grosskopf, 2001).

variety of alternative settings involving different choices and preferences. They explicitly distinguish between a consensus effect and a truly false consensus effect where information about oneself is weighted more heavily than information about a randomly selected other person from the same sample when forming beliefs. Their study attests the presence of a consensus effect, but rejects the presence of a *false* consensus effect. Further evidence for the existence of a consensus effect has been provided using the trust game (Altmann et al., 2008), the sequential prisoner's dilemma (Blanco et al., 2011, 2014; Miettinen et al., 2020) and the leader-follower game (Gächter et al., 2012). The most compelling evidence comes from Blanco et al. (2014) who are the first to explicitly elicit beliefs about second-mover actions and show that these are influenced by subjects' own second-mover actions. However, all mentioned studies rely on elicitation of preferences and beliefs in the same session, with beliefs being elicited *after* preference elicitation, a setup in which subjects' own preferences are extremely salient. The reason is that virtually all these studies are not designed to test the consensus effect as main research question. Typically, these studies are interested in a clean measure of preferences and use beliefs only as a control variable. In this case preferences have to be elicited first so that the measure is clean from potential confounds from previous elicitations. We contribute to this literature by investigating the effect of different orders of elicitation of preferences and beliefs.

Closely related to our approach is the study by Engelmann and Strobel (2012), which shows that individuals are sensitive to the way that information about other people is presented. If information about others is particularly prominent or salient, people overweight it (and underweight information about themselves). But if some cognitive effort is required to retrieve the same information about others, the opposite is true, i.e., people underweight that information. In contrast to their approach, we investigate how the salience of own preferences rather than the salience of others' preferences affects the consensus effect.

The remainder of the paper is structured as follows. Section 2 describes the experimental design and procedures. Section 3 reports our results, Section 4 discusses implications and concludes.

2 Experimental Design and Procedures

As a workhorse to examine our research question, we use the binary trust game depicted in Figure 1. In this game, a first mover chooses between actions "OUT" and "IN". If they choose "OUT", the payoff for both players is 10 €, regardless of the second mover's action. If they choose "IN", there is an efficiency gain and players' payoffs depend on the second mover's choice who can decide whether to distribute

the payoffs equally ("Option B" yielding 15 € for each player), or to keep more for themselves ("Option A" yielding payoffs of 8 € to the first mover and 22 € for the second mover).

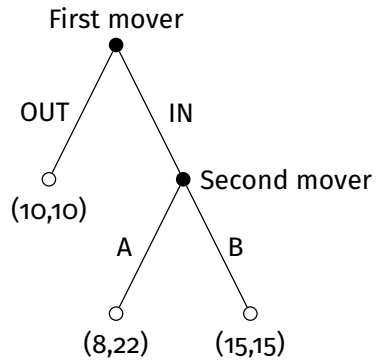

Figure 1. Game tree of the binary trust game.

The subgame-perfect Nash equilibrium for self-interested players is ("OUT", "Option A"). However, joint payoff is maximized if the first mover chooses "IN". In our experiment, subjects play the trust game in both roles. For each participant, the main measures we elicit are: first-mover actions, second-mover actions, and beliefs about other second-movers' behavior. At the end of the experiment, one of these decisions is randomly selected for payment to exclude hedging possibilities (Blanco et al., 2010).

First-mover actions. First movers' actions are elicited by asking players to make a decision between "IN" and "OUT". For self-interested first movers, these decisions reflect only beliefs about second-mover behavior. In particular, if the first mover ranks outcomes $(8, 22) \prec (10, 10) \prec (15, 15)$, they will choose "IN" if and only if their belief about the probability that the second mover chooses "Option B" exceeds some positive threshold. Such a belief is rational in case some second movers are expected to choose "Option B". However, choosing "IN" may also be related to social preferences, preferences for efficiency or other motives such as risk preferences, betrayal aversion or altruism (see, e.g., Bohnet and Zeckhauser, 2004; Cox, 2004). Due to these potential confounds, in our design we also measure beliefs directly rather than inferring them from first-mover choices.

Second-mover actions. Second-mover actions are elicited using the strategy method (Selten, 1965). Participants are asked whether they would choose "Option A" or "Option B" in case their paired first mover chooses "IN". The sequentiality of players' moves ensures the absence of a strategic component in the second-mover choice. Thus, it can be interpreted as a preference measure. Moreover, this measure of (social) preferences is not confounded with efficiency concerns since "Option A"

and "Option B" lead to the same sum of payoffs.[3] Choosing "Option B" can be consistent with several models of social preferences such as inequity aversion (Fehr and Schmidt, 1999), reciprocity (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006), and guilt aversion (Battigalli and Dufwenberg, 2007). As the main objective of the paper is not related to the main motivation behind these choices, it is sufficient to assume that individuals have a psychological cost high enough to make them choose "Option B" instead of "Option A".

Beliefs. Our third measure is the belief that a participant has about other second movers' actions. We ask subjects to state how many out of 20 students playing as "Player 2" (i.e., the second mover) in another session they think will choose "Option B".[4] They answer by choosing between 7 equally-sized intervals from "0 – 2" to "18 – 20". Correct guesses are rewarded with 12€, while there is no payoff for incorrect guesses. Given the choice of a particular interval, they play a lottery in which they win 12€with the probability they estimate for that interval and 0€with the complementary probability. Thus, for any possible distribution of beliefs and any plausible model of risk preferences, individuals have an incentive to select the interval where they put the highest probability mass. This guarantees that risk preferences do not confound belief elicitation.[5] Since first movers' choices may not (only) reflect beliefs about second-mover actions (see above), from now on we focus our attention on second-mover choices and beliefs to identify the consensus effect.

Treatments. We employ two between-subjects treatments, in which we manipulate the salience of individual preferences. In particular, across treatments we vary the order of elicitation of the above mentioned measures. In the *high salience* treatment, we first elicit second-mover choices, then beliefs and finally first-mover choices. To our knowledge, all prior economics experiments on the consensus effect have relied on this particular order, i.e., second-mover decisions directly precede belief elicitation (Jacobsen and Sadrieh, 1996; Selten and Ockenfels, 1998; Charness and Grosskopf, 2001; Blanco et al., 2014). In the *low salience* treatment,

[3] In a sequential prisoner's dilemma, preferences for efficiency may generate a correlation between first- and second-mover actions. In a typical parameterization of the game ($2 * \pi_i(C,C) > \pi_i(C,D) + \pi_j(C,D) > 2 * \pi_i(D,D)$), players who care only about efficiency, i.e., seek to maximize total payoff, should always choose C as first and second movers, thus leading to a perfect correlation of actions even in the absence of a consensus effect. This relationship is less pronounced, but present for players who care about their own payoff as well as efficiency.

[4] We used the behavior of 20 subjects in another session of the same treatment to assess whether a guess was correct. As is standard in experimental economics, subjects are not informed that there are other treatments.

[5] Giving subjects a choice between 7 intervals rather than all 21 possibilities makes the measurement coarser but at the same time increases subjects' chances of actually guessing correctly, thus increasing the perceived importance of their decision. Throughout the paper we will report beliefs as relative frequencies converted from subjects' answers by taking the mid-point of the chosen interval and dividing by 20.

we first elicit beliefs followed by first-mover choices and finally by second-mover choices.

Sample size. We used ex-ante power analysis to determine our sample size. Apart from setting a desired level of significance (α) and power ($1-\beta$), power analysis requires deciding on the minimal effect size one wants to detect. While conventional values for α (≤ 0.05) and $1-\beta$ (≥ 0.80) are usually employed, the minimal effect size is ultimately an empirical issue. Typically, for replication exercises the effect size found in previous studies is used (see, e.g., OpenScienceCollaboration, 2015; Camerer et al., 2016, 2018). However, for research investigating a novel hypothesis no such guidance exists. In our case, as virtually all previous literature has employed only one order of elicitation, it is difficult to set a minimal effect size for the difference between the two orders *ex ante*. For this reason, we rely on a first experiment conducted in the lab which uses the same design as the current experiment. In this initial experiment, which we report in the Online Appendix, we find support for our hypothesis, that is, a stronger correlation between beliefs and preferences in the *high salience* treatment than in the *low salience* treatment. Moreover, the two correlation coefficients are statistically different in size. Based on the effect size of this first experiment (Cohen's $q = 0.41$), we recruit a total of 286 participants (142 in *high salience* and 144 in *low salience*, 64.7% female, mean age 24.7 years). This sample size allows us to detect the effect size of the original experiment with power slightly higher than 95%.[6]

Procedures. Our experiment was conducted online, recruiting subjects from the BonnEconLab subject pool. Upon accepting to participate, subjects read the instructions on their screens. Participants made decisions in the trust game and filled in a brief socio-demographic questionnaire. At the end of the experiment, participants were asked their bank details to be paid for the selected task (belief elicitation, second-mover or first-mover choices). Payment happened within 24 hours from the end of the experiment. The experiment lasted on average 10 minutes and people earned on average 9.17€.

3 Results

We start our analysis by reporting descriptive statistics on subjects' behavior in the experiment. Considering the two treatments jointly, 40.2% of our subjects choose

[6] Originally, we conducted a first test of our hypothesis gathering data as part of a longitudinal experiment in which participants were invited to take part in three laboratory sessions over three consecutive weeks. In this experiment, we run the exact two treatments that we used in our online experiment plus two additional treatments where some of the measures were elicited in week 1 and some in week 3. The additional treatments were intended to further manipulate salience. The Online Appendix reports the exact design and the results of the first experiment.

"Option B", i.e. reciprocate trust, when playing as second movers and 50.3% choose "IN" when playing as first movers. Subjects' beliefs about the second-mover action of other subjects are quite accurate on average. Pooling all treatments, subjects predict that 42.3% would reciprocate as second movers.

	2^{nd} mover	belief	1^{st} mover
High salience ($n = 142$)	37.3%	41.7%	49.3%
Low salience ($n = 144$)	43.1%	43.0%	51.4%

Notes. "2^{nd} mover" displays the share of participants who chose "Option B", resulting in an equal distribution (15,15), as second movers in the binary trust game. "Belief" describes the average belief subjects hold about the share of second movers in another session choosing Option B. "1^{st} mover" describes the share of participants who chose "IN" as first movers in the binary trust game.

Table 1. Averages of actions and beliefs in trust game by treatment.

Table 1 displays the percentage of trustworthy subjects, their beliefs about others' trustworthiness, and the percentage of trusting participants for the *high salience* and *low salience* treatments. We find no differences between the two treatments in the distributions of the three measures (χ^2-tests for homogeneity; second-mover action: $p = .323$; belief: $p = .472$; first-mover action: $p = .723$).

Following the previous literature (Mullen et al., 1985; Blanco et al., 2014), we attest the presence of a consensus effect whenever there is a significant positive correlation between second-mover actions and beliefs. In Table 2, we report the Spearman rank-order correlation coefficients for both treatments. We find correlations that are significantly different from zero ($\rho = .450$ and $\rho = .444$ for the *high salience* and *low salience* treatment, respectively; both $p < .001$). This provides strong evidence for the presence of the consensus effect.

Strikingly, the correlation coefficients are almost identical in size which speaks against our hypothesis that increasing the salience of own preferences strengthens the consensus effect. In fact, when comparing the two correlation coefficients, we cannot reject the null hypothesis that the two correlation coefficients are equal

	High salience	Low salience
$\rho_{2^{nd},belief}$	.450***	.444***
	(< .001)	(< .001)
N	142	144

Notes. ρ: Spearman's correlation coefficient (between second-mover actions and beliefs). p-values in parentheses.

Table 2. Consensus effect by treatment

(one-sided z-test: $p = .9483$).[7] Hence, we find no evidence for the order of elicitation to have an effect on the size and the significance of the consensus effect.

The distributions of beliefs conditional on own type displayed in Figure 2 corroborate these conclusions. The top panels of Figure 2 show the average beliefs on others' trustworthiness conditional on own second-mover strategies for each treatment. In both treatment conditions, second movers choosing "Option B" believe on average that a larger fraction of other second movers in another session of the experiment would choose "Option B" than second movers choosing "Option A". In fact, second movers choosing "Option B" in the *high (low) salience* believe that 56.8% (54.6%) of other second movers would choose "Option B" as well, while those who choose "Option A" believe only 32.6% (34.3%) would choose "Option B". These differences in average beliefs are statistically significant in both treatment conditions (Wilcoxon rank-sum tests: $p < .001$), providing evidence for the robustness of the consensus effect to the degree of salience of own type.

The fact that, in contrast to our hypothesis, the difference in beliefs conditional on own type is equally strong in both treatment conditions, is further revealed by the bottom panels of Figure 2 where we depict the empirical distribution of beliefs conditional on second-mover type. Comparing the distributions using Kolmogorov-Smirnov tests reveals a significant difference by type for both treatments (*high salience*, two-sided exact test: $p < .001$; *low salience*, two-sided exact test: $p < .001$).

[7] To test whether correlations coefficients are significantly different from one another we use the following procedure. We apply the approximate Fisher's z transformation (Fisher, 1915) to transform the distribution of the relevant correlation coefficients: $z' = \frac{1}{2}(\ln(1+\rho) - \ln(1-r))$. This generates variables distributed with an approximate normal distribution with standard error $\sigma_z = \frac{1}{\sqrt{N-3}}$ on which a z-test can be performed. Although this procedure is aimed at Pearson correlation coefficients, Myers and Sirois (2006) find it to be the most efficient for Spearman correlation coefficients as well. It was implemented using the CORTESTI package (Caci, 2000) in Stata.

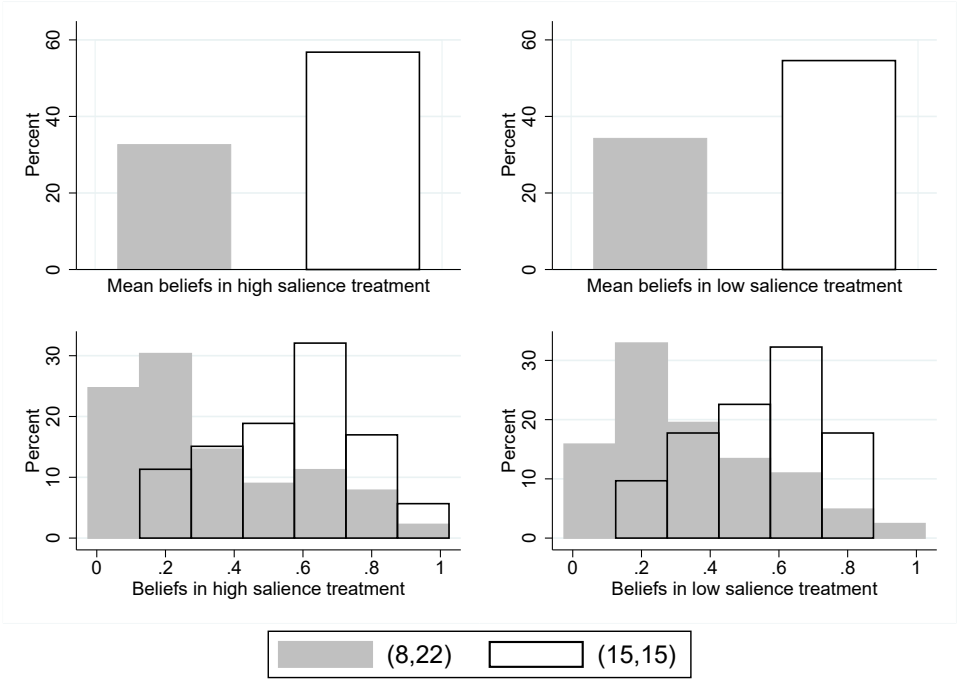

Figure 2. The top (bottom) panels show the average (distribution of) beliefs about the share of trustworthy second movers in another session conditional on own second-mover type.

4 Summary and conclusion

In this paper, we have tested whether the extent to which people project their own preference onto others when forming beliefs about others' preferences depends on the salience of their own preference type. In order to manipulate salience we have conducted two between-subject treatments that vary the order of elicitation between preferences and beliefs in a binary trust game. We interpret second-mover actions as preferences. We measure participants' own preferences and elicit their beliefs about the distribution of second movers' actions in our experiment.

In our *high salience* treatment, preference elicitation, measured by second-mover behavior in the trust game, precedes belief elicitation, while in our *low salience* treatment the opposite order is used. We find strong confirmatory evidence for the existence of a consensus effect, that is, a significant correlation between preferences and beliefs in both treatment conditions. In contrast to our hypothesis, however, the order of elicitation of preferences and beliefs does not affect the strength of the consensus effect as the two correlation coefficients are not statistically different from each other.

The robustness of the consensus effect strongly indicates that people take into account their own preferences when forming beliefs about others' preferences.

Moreover, salience of own preferences does not seem to contribute to additional overweighting of one's own type. Arguably, own preferences become more salient when being confronted with behaviors that reveal them to oneself. This may occur for several reasons: people may not be focused on their preferences despite knowing their preferences and being in the decision situation makes their type salient. Alternatively, people may not fully know their preferences and be uncertain about their type. In this case, being confronted with the decision situation may simply reduce uncertainty. In light of our results, however, individuals seem to put the same weight on own preferences when forming beliefs about others, independent of salience of own preferences or uncertainty about their own type. In this sense, salience of own preferences does not affect the formation of beliefs about others' preferences. Other potential drivers of the consensus effect such as selective exposure and cognitive availability, logical information processing, motivational processes, social support or self-esteem maintenance need to be scrutinized (see Marks and Miller, 1987, for a more detailed account of these theoretical views).

Beyond, our findings have additional implications for theory, policy and the design of experiments. The robustness of the consensus effect in the absence of information about others indicates that beliefs are going to be correlated with preferences and this could lead to polarization in beliefs if preferences are polarized. Hence, in settings with limited information about others and polarized preferences, a more polarized distribution of beliefs is expected. This phenomenon reinforces confirmation bias that is widely observed in echo chambers where groups that have similar preferences interact. From a policy perspective, the finding that exogenously directing participants' focus onto themselves does not strengthen the consensus effect indicates that people's distorted beliefs about others' preferences and values are not fully corrected by reducing their exposure to echo chambers; this is because their belief formation would still be affected by their own preferences.

Likewise, our findings have implications for the design of experiments, in which preferences and beliefs are elicited, and their interpretation. Even though researchers should be aware of the potential distortion of beliefs due to the consensus effect, our results are reassuring in the sense that the order of elicitation of preferences and beliefs does not cause additional distortions. As a result, there is no superior order, and researchers can reliably use only one of the orders (as done by a large fraction of the previous literature) when both preferences and beliefs ought to be elicited in within-subject designs.

References

Altmann, Steffen, Dohmen, Thomas, and Wibral, Matthias (2008). "Do the reciprocal trust less?" *Economics Letters*, 99(3), 454–457.

Battigalli, Pierpaolo and Dufwenberg, Martin (2007). "Guilt in games." *American Economic Review*, 97(2), 170–176.

Bazinger, Claudia and Kühberger, Anton (2012). "Is social projection based on simulation or theory? Why new methods are needed for differentiating." *New Ideas in Psychology*, 30(3), 328–335.

Blanco, Mariana, Engelmann, Dirk, Koch, Alexander K, and Normann, Hans-Theo (2010). "Belief elicitation in experiments: is there a hedging problem?" *Experimental Economics*, 13(4), 412–438.

Blanco, Mariana, Engelmann, Dirk, Koch, Alexander K, and Normann, Hans-Theo (2014). "Preferences and beliefs in a sequential social dilemma: a within-subjects analysis." *Games and Economic Behavior*, 87, 122–135.

Blanco, Mariana, Engelmann, Dirk, and Normann, Hans Theo (2011). "A within-subject analysis of other-regarding preferences." *Games and Economic Behavior*, 72(2), 321–338.

Bock, Olaf, Baetge, Ingmar, and Nicklisch, Andreas (2014). "hroot: Hamburg registration and organization online tool." *European Economic Review*, 71, 117–120.

Bohnet, Iris and Zeckhauser, Richard (2004). "Trust, risk and betrayal." *Journal of Economic Behavior & Organization*, 55(4), 467–484.

Bordalo, Pedro, Gennaioli, Nicola, and Shleifer, Andrei (2022). "Salience." *Annual Review of Economics*, 14, 521–544.

Caci, Herve M. (2000). "CORTESTI: Stata module to test equality of two correlation coefficients."

Camerer, Colin F, Dreber, Anna, Forsell, Eskil, Ho, Teck-Hua, Huber, Jürgen, Johannesson, Magnus, Kirchler, Michael, Almenberg, Johan, Altmejd, Adam, Chan, Taizan, et al. (2016). "Evaluating replicability of laboratory experiments in economics." *Science*, 351(6280), 1433–1436.

Camerer, Colin F, Dreber, Anna, Holzmeister, Felix, Ho, Teck-Hua, Huber, Jürgen, Johannesson, Magnus, Kirchler, Michael, Nave, Gideon, Nosek, Brian A, Pfeiffer, Thomas, et al. (2018). "Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015." *Nature Human Behaviour*, 2(9), 637–644.

Cattell, Raymond B (1944). "Projection and the design of projective tests of personality." *Character & Personality; A Quarterly for Psychodiagnostic & Allied Studies*, 12, 177–194.

Charness, Gary and Grosskopf, Brit (2001). "Relative payoffs and happiness: an experimental study." *Journal of Economic Behavior & Organization*, 45(3), 301–328.

Cox, James C (2004). "How to identify trust and reciprocity." *Games and economic behavior*, 46(2), 260–281.

Dawes, Robyn M (1989). "Statistical criteria for establishing a truly false consensus effect." *Journal of Experimental Social Psychology*, 25(1), 1–17.

Dohmen, Thomas, Quercia, Simone, and Willrodt, Jana (2022). "On the Psychology of the Relation between Optimism and Risk Taking." IZA Discussion Paper No. 15763.

Dufwenberg, Martin and Kirchsteiger, Georg (2004). "A theory of sequential reciprocity." *Games and Economic Behavior*, 47(2), 268–298.

Engelmann, Dirk and Strobel, Martin (2000). "The false consensus effect disappears if representative information and monetary incentives are given." *Experimental Economics*, 3, 241–260.

Engelmann, Dirk and Strobel, Martin (2012). "Deconstruction and reconstruction of an anomaly." *Games and Economic Behavior*, 76(2), 678–689.

Falk, Armin and Fischbacher, Urs (2006). "A theory of reciprocity." *Games and Economic Behavior*, 54(2), 293–315.

Fehr, Ernst and Schmidt, Klaus M (1999). "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, 114(3), 817–868.

Fischbacher, Urs (Feb. 2007). "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10(2), 171–178.

Fisher, Ronald A (1915). “Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population.” *Biometrika*, 10(4), 507–521.

Gächter, Simon, Nosenzo, Daniele, Renner, Elke, and Sefton, Martin (2012). “Who makes a good leader? Cooperativeness, optimism, and leading-by-example.” *Economic Inquiry*, 50(4), 953–967.

Heider, Fritz (1958). “The psychology of interpersonal relations.” *Wiley, New York*.

Jacobsen, Eva and Sadrieh, Abdolkarim (1996). “Experimental proof for the motivational importance of reciprocity.” University of Bonn, Germany.

Jones, Edward and Nisbett, Richard (1987). “The actor and the observer: Divergent perceptions of the causes of behavior.” In *E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), Attribution: Perceiving the causes of behavior (pp. 79–94)*. Lawrence Erlbaum Associates, Inc.

Marks, Gary and Miller, Norman (1987). “Ten years of research on the false-consensus effect: An empirical and theoretical review.” *Psychological bulletin*, 102(1), 72–90.

Miettinen, Topi, Kosfeld, Michael, Fehr, Ernst, and Weibull, Jörgen (2020). “Revealed preferences in a sequential prisoners’ dilemma: A horse-race between six utility functions.” *Journal of Economic Behavior & Organization*, 173, 1–25.

Mullen, Brian, Atkins, Jennifer L, Champion, Debbie S, Edwards, Cecelia, Hardy, Dana, Story, John E, and Vanderklok, Mary (1985). “The false consensus effect: A meta-analysis of 115 hypothesis tests.” *Journal of Experimental Social Psychology*, 21(3), 262–283.

Myers, Leann and Sirois, Maria J. (2006). “Differences between Spearman correlation coefficients.” In. *Encyclopedia of Statistical Sciences*. American Cancer Society.

Offerman, Theo, Sonnemans, Joep, and Schram, Arthur (1996). “Value orientations, expectations and voluntary contributions in public goods.” *The Economic Journal*, 817–845.

OpenScienceCollaboration (2015). “Estimating the reproducibility of psychological science.” *Science*, 349(6251).

Ross, Lee, Greene, David, and House, Pamela (1977). “The ”False Consensus Effect”: An Egocentric Bias in Social Perception and Attribution Processes.” *Journal of Experimental Social Psychology*, 13, 279–301.

Selten, Reinhard (1965). *Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes*.

Selten, Reinhard and Ockenfels, Axel (1998). “An experimental solidarity game.” *Journal of Economic Behavior & Organization*, 34(4), 517–539.

Online Appendix

In this section, we describe our first lab experiment which helped us determine the sample size for the online experiment reported in the main text. Our original lab experiment was a longitudinal experiment in which participants were invited to take part in three laboratory sessions over three consecutive weeks. Table A1 reports the structure of the longitudinal experiment with all the task subjects went through. In bold font, we report the parts where subjects did tasks related to the binary trust game.

Week 1	**Week 2**	**Week 3**
Mood Question	Mood Question	Mood Question
General Risk Question	General Risk Question	General Risk Question
Big Five	Big Five	Big Five
	Trust Question	Locus of Control
Binary Trust Game: Treatments		**Binary Trust Game: Treatments**
Risk Premia (Choice Lists)	"Will you win?" task	Risk Premia (Choice Lists)
Risk Scenarios	Urns Task	Common Ratio Effect
	Bet: Heads or Tails?	BRET
	Ambiguity preferences and Beliefs	
Sociodemographics	Optimism: LOT and SOP	Optimism: LOT and SOP
		IQ
Mood Question	Mood Question	Mood Question

Notes. For detailed information on the tasks not described in this paper refer to Dohmen et al. (2022)

Table A1. Overview of all tasks participants completed.

Each session lasted about one hour and contained several distinct parts. The experiment took place in the summer and fall of 2016. It was computerized using z-Tree (Fischbacher, 2007) and participants were recruited from the BonnEconLab subject pool using the software h-root (Bock et al., 2014). We only invited subjects who had not played the trust game at the BonnEconLab before. One part per week was randomly selected for payoff, with each part being equally likely to be selected. This was clearly communicated to subjects before choices were made. This payment scheme precludes hedging, while keeping the stakes within one part sizable enough for subjects to exert effort.

The treatments related to our research question in our first experiment were as follows. Treatments *high salience* (n = 34) and *low salience* (n = 44) were identical to the respective treatments of our main experiment. In treatment *low salience – time*, the measurement of preferences and beliefs took place two weeks apart. Subjects played the trust game in the role of second mover in week 1, but beliefs (and first-mover actions) were not elicited until week 3. The idea behind this manipulation is that subjects' own preferences are less salient if beliefs are elicited two weeks later. In this treatment we had 54 participants. A fourth treatment *low salience – time & order* combines our two manipulations of salience. That is, beliefs (and first-mover actions) were elicited in week 1, while second-mover actions are elicited in week 3. In this treatment we had 34 participants.

In Table A2, we report Spearman rank correlation coefficients between preferences and beliefs for each treatment.

	High salience	**Low salience**	**Low salience - time**	**Low salience - time & order**
$\rho_{2^{nd},belief}$	.559***	.218	.234*	.268
	(.001)	(.155)	(.089)	(.126)
N	34	44	54	34

Notes. ρ: Spearman's correlation coefficient (between second-mover actions and beliefs). p-values in parentheses.

Table A2. Consensus effect by treatment.

As can be seen from the table, the correlation in the *high salience* treatment is significantly different from zero and substantially higher compared to all other treatments. The correlation in the *high salience* treatment is significantly higher both compared to *low salience* and *low salience - order* (one-sided z-test: $p < .042$)