Bauer, Dominik; Wolff, Irenaeus

Conference Paper

Biases in Belief Reports

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Bauer, Dominik; Wolff, Irenaeus (2021) : Biases in Belief Reports, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2021: Climate Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:
https://hdl.handle.net/10419/242458

Biases in Belief Reports[§]

Dominik Bauer Irenaeus Wolff
Thurgau Institute of Economics/University of Konstanz
dominik.3.bauer@uni.kn wolff@twi-kreuzlingen.ch

This version: 11th January, 2021

Abstract: Belief elicitation is important in many different fields of economic research. We show that how a researcher elicits such beliefs—in particular, whether the belief is about the participant's opponent, an unrelated other, or the population of others—affects the processes involved in the formation of belief reports. We find a clear consensus effect. Yet, when matching the opponent's action would lead to a low payoff and the researcher asks for the belief about this opponent, *ex-post* rationalization kicks in and beliefs are re-adjusted again. Hence, we recommend to ask about unrelated others or about the population in such cases, as 'opponent beliefs' are even more detached from the beliefs participants had when deciding about their actions in the corresponding game. We find no evidence of a hindsight bias or wishful thinking in any of the treatments.

JEL classification: C72, C91, D84

Keywords: Belief Elicitation, Belief Formation, Belief-Action Consistency, Framing Effects, Projection, Consensus Effect, Wishful Thinking, Hindsight Bias, *Ex-Post* Rationalization

1 Introduction

Subjective beliefs play a central role in economic theory. When facing a decision, people often do not know the true probabilities of the relevant states of the world. Standard economic theory assumes that in such situations, people form subjective beliefs and act on those subjective beliefs as if they were the true probabilities (Savage, 1954). Because of this assumption, eliciting subjective beliefs

[§]We would like to thank Ariel Rubinstein, Yuval Salant, Robin Cubitt, Marie Claire Villeval, Bård Harstad, Dirk Sliwka, and Roberto Weber, the research group at the Thurgau Institute of Economics and the members of the Graduate School of Decision Sciences of the University of Konstanz, as well as participants of several conferences and seminars for their helpful comments.

often is extremely helpful to test economic models, as well as for understanding behaviour more generally. The list of examples for this approach is long.[1]

Given belief elicitation has such a broad field of applications, it is crucial that we know how people come up with their belief reports depending on the circumstances. And, indeed, there is a sizeable literature on belief elicitation (for recent reviews, see Schotter & Trevino, 2014, or Schlag *et al.*, 2015). However, this literature has focused mainly on two questions: how to incentivize belief reports,[2] and whether to ask for beliefs before or after actions are chosen.[3]

We will focus on a different aspect: whether participants are asked about their situation-specific opponent or about unrelated others. Virtually all studies in the literature use a population treatment (asking about all other participants in the session) or an opponent treatment (asking about the participants' direct interaction partner), but the specific choice is rarely motivated. Importantly, this choice correlates with the results of a study. All major studies in economics documenting a consensus effect (forming beliefs about others using oneself as a model) use a population treatment.[4] In contrast, studies on belief-action consistency typically use opponent treatments and do not find a consensus effect.[5]

Therefore, our first contribution is to answer the question of why a consensus effect seems to be linked to asking about the population. We show that asking about the opponent's behaviour does not eliminate the consensus effect *per se*. Rather, an opponent treatment will make *ex-post* rationalization (fitting one's belief to a prior action in order to appear consistent) override the consensus effect when actions are strategic substitutes. However, this is exactly the type of

[1] For a list of examples from numerous domains, see, *e.g.*, Trautmann & van de Kuilen (2015).

[2] *E.g.*, Armantier & Treich, 2013; Harrison *et al.*, 2014; Hollard *et al.* 2016; Holt & Smith, 2016; Hossain & Okui, 2013; Karni, 2009; Palfrey & Wang, 2009; Trautmann & van de Kuilen, 2015.

[3] *E.g.*, Costa-Gomes & Weizsäcker (2008); out of the 20 studies mentioned in footnotes 4 and 5, 15 ask for beliefs only after choices at least in some treatments, nine do so exlusively, and six use different treatments to control for the timing of the belief question (one study does not give information about the order of elicitation). Additional topics are hedging (Blanco *et al.*, 2010), the usefulness of second order beliefs (Manski & Neri, 2013), and the precision with which probabilities can be expressed or how the support of the belief distribution is determined (Delavande *et al.*, 2011a).

[4] Selten & Ockenfels (1998), Charness & Grosskopf (2001), Van Der Heijden, Nelissen & Potters (2007), Engelmann & Strobel (2012), Iriberri & Rey-Biel (2013), Blanco *et al.* (2014), Danz, Madarász & Wang (2014), Molnár & Heintz (2016), Rubinstein & Salant (2016), Proto & Sgroi (2017).

[5] Costa-Gomes & Weizsäcker (2008), Danz *et al.* (2012), Hyndman *et al.* (2012), Hyndman *et al.* (2013), Manski & Neri (2013), Nyarko & Schotter (2002), Rey-Biel (2009), Sutter *et al.* (2013), Trautmann & van de Kuilen (2015), Wolff (2015).

situation that allows to distinguish a consensus effect from other effects.[6]

Our second contribution is to provide some evidence on whether '*ex-ante* rationalization' (choosing the optimal alternative given a belief, as posited by game theory), the consensus effect, and *ex-post* rationalization are the only processes that matter for belief reports. In light of the huge number of biases that people have been found to exhibit, it is not obvious that no other process would play a role for reported beliefs. And yet, the literature that uses (as opposed to: "studies") belief elicitation discusses exactly the above-mentioned three processes when making sense of empirical observations.

We went through a long list of potential biases to see which of the biases would conceptually fit the setup we had in mind for answering our first research question. For this study, we focus on processes that happen after choices have been made (which will become clearer conceptually in Section 2).[7] From that list, we found 12 biases that seem applicable to our setting, 10 of which happen after choices are made.[8] However, four of them are possible root-causes of *ex-post* rationalization, and two others cannot be isolated from the consensus effect, which is why we group them into two 'bias groups'. In the end, we will be able to distinguish two processes in addition to what has been discussed in the literature on belief reports: hindsight bias and wishful thinking. Reassuringly for the interpretation of existing studies, we do not find evidence for either a hindsight bias or wishful thinking to affect belief reports.

Our paper has two main parts, comprising three experiments. In Experiment 1, we use a pure discoordination game and elicit beliefs in the two standard treatments, the opponent treatment and the population treatment.[9] As pointed out,

[6]Our data are consistent with the hypothesis that *ex-post* rationalization is triggered by cognitive dissonance (beliefs that are inconsistent with one's actions lead to cognitive distress, which motivates belief re-adjustments). When a consensus effect leads to a belief that others are likely to do the same as the player himself, cognitive dissonance then will arise only when actions are strategic substitutes.

[7]The biases that fit our setup if conceptualized appropriately were bias blind spot, cognitive dissonance, confirmation bias, conservatism in updating, correlation neglect, hindsight bias, illusion of control, salience bias, social-desirability bias, and wishful thinking. Biases that did not make sense within our setup were: base-rate fallacy, belief bias, conjunction fallacy, contrast effect, fundamental attribution error, gambler's fallacy, hot-hand fallacy, and status-quo bias.

[8]For an overview, see Table 2 at the beginning of Section 2.

[9]Note that the opponent and population treatments differ in several dimensions, including the incentivization (see Table 1). Therefore, we refrain from calling them "frames" in most of the paper. When we do talk of "frames", we refer to the mental representation of the question in participants' heads.

we replicate the systematic differences from the literature: a consensus effect in the population treatment, and higher observed best-response rates in the opponent treatment.

The population and opponent treatments differ in four ways: whether the participant interacts with (most of) the participant(s) who form the belief's 'target', how many people are the belief's target, the exact incentivization, and whether we ask about a percentage or a probability. To find out which features of the main treatments are responsible for the differences between them, we add a third treatment which we call 'random-other treatment'.[10] A random-other treatment asks for participants' beliefs about the behaviour of some other individual who is not the matching partner, and allows for *ceteris-paribus* comparisons with the corresponding opponent treatment.

Table 1 contrasts the three types of treatments, show-casing all four differences betweeen opponent and population treatments. The data shows that the relevant difference is between the random-other and opponent treatments, and not between the random-other and the population treatments. Thus, it is the *interaction* with the belief's target that makes the difference.

The second part of our paper varies the environment in which we elicit beliefs (*i.a.*, the game people play). We use two experiments to disentangle the processes that lead to biased belief reports, using a rock-paper-scissors-type of game in Experiment 2.A and a sequential battle-of-the-sexes game in Experiment 2.B.

Experiment 2.A rules out a number of potentially active biases that might have affected belief reports in Experiment 1. To test for a consensus effect in the opponent treatment, Experiment 2.B eliminates the 'cognitive need' for *ex-post* rationalization. We find as much of a consensus effect in the opponent treatment as in the population treatment. We thus conclude that initially, opponent treatments lead to as much consensus effect as the other treatments. However, opponent treatments trigger subsequent *ex-post* rationalization whenever the beliefs that result from the consensus effect would lead to cognitive dissonance.

[10] Critcher and Dunning (2013 and 2014) use an "individual" frame to study judgments of morally relevant behaviours. The individual frame is similar to the random other frame in that it asks for the belief about "a randomly selected student… [whose] initials are LB".

Opponent treatment

Object: Single person, the matching partner

"With what probability did your matching partner choose each of the respective boxes of the current set-up?"

Incentivization: $Pr(\text{win}) = 1 - \frac{1}{2}\Big(\big[1 - r(a_{\text{true}})\big]^2 + \sum_{a_j \neq a_{\text{true}}} r(a_j)^2\Big)$, where $r(a_j)$ is the reported probability of the 'Object' playing action a_j and a_{true} is the 'Object's' true choice.

Random-other treatment

Object: Single person, not the matching partner

"With what probability did a person who is not your matching partner choose each of the respective boxes of the current set-up?"

Incentivization: $Pr(\text{win}) = 1 - \frac{1}{2}\Big(\big[1 - r(a_{\text{true}})\big]^2 + \sum_{a_j \neq a_{\text{true}}} r(a_j)^2\Big)$.

Population treatment

Object: Many people, almost all of them not the matching partner

"What is the percentage of other participants of today's experiment choosing each of the respective boxes of the current set-up?"

Incentivization: $Pr(\text{win}) = 1 - \frac{1}{2}\sum_j \big[r(a_j) - f(a_j)\big]^2$, where $f(a_j)$ denotes action a_j's relative frequency in the population.

Table 1: The three types of treatments we use (differences underlined).

2 The applicable processes and our treatments

Table 2 gives a short description of processes known to affect probability judgments and indicates whether a process is applicable within our setting(s). The three bold-faced processes are the processes that have been discussed prominently as determinants of belief reports. We next describe the applicable processes and identify in which treatment(s) they could matter. We summarize our predictions in Table 3 at the end of this Section.

Before we discuss the processes in detail, however, Figure 1 shows when we expect each process to play a role. Salience bias (being attracted by salient items) and bias blind spot (assuming that only others are affected by a bias, in this case, salience bias) will happen when players form their belief. '*Ex-ante* rationalization' (forming a belief and best-responding to it) then leads to the chosen action.

Process	Short description	Applicability	Focus of experiment…
'ex-ante rationalization'	forming a belief, then best-responding to the belief	✓	2.A
bias blind spot	everybody else is falling for a fallacy, but not me	(✓)	
consensus effect	belief that others are like me ⇒ they will act as I do/would	✓	2.A/2.B
conservatism in updating	(partially) ignoring new information	✓	
correlation neglect	ignoring that two events are correlated	✓	
***ex-post* rationalization**	fitting a belief to an action after that action has been taken	✓	2.A/2.B
cognitive dissonance	when my action is inconsistent with my belief, I adjust the belief as to correct the inconsistency	✓	
illusion of control	belief that I can influence pure-chance moves	✓	
social-desirability bias	reporting behaviour/opinions that conforms untruthfully closely to social norms	✓	
confirmation bias	when I have a theory, I only search for confirming evidence	✓‘	
hindsight bias	not being able to abstract from knowledge acquired after a choice was made, when assessing that choice	✓	2.A
salience bias	being attracted by salient choices/labels	(✓)	
wishful thinking	when people assign a higher probability to favourable outcomes just because they are favourable	✓	2.A
base-rate fallacy	ignoring prior probabilities	✗	
belief bias	if the conclusion is right, the argument must be right, too	✗	
conjunction fallacy	ignoring that the conjunction of two events can never be more likely than either event separately	✗	
contrast effect	draws more attention to items/characteristics that change strongly	✗	
fundamental attribution error	attributing too much to the characteristics of a person and too little to the characteristics of the situation	✗	
gambler's fallacy	belief that prior realisations of an i.i.d. process change future probabilities, to move the observed mean closer to its expected value	✗	
hot-hand fallacy	belief that s.b. who has been lucky several times in a row is more likely to be lucky the next time, too	✗	
status-quo bias	a preference for the current state relative to any changes, irrespective of what the current state is	✗	

Table 2: Overview of all processes considered. Processes that have been prominent in the literature as affecting belief reports are marked in bold face. Processes that are considered jointly with or as underlying causes of other processes are indented and directly follow the corresponding 'process category'. Further note that some of the "non-applicable" ones are non-applicable because they would have required feedback about others' choices.

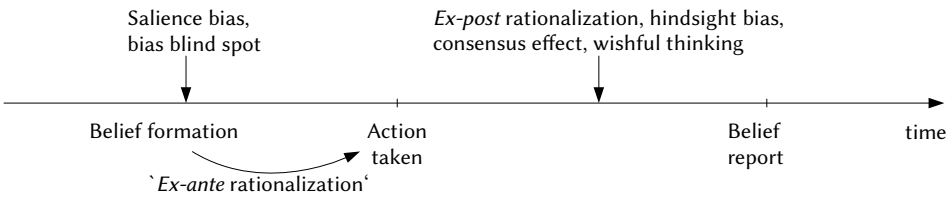


Figure 1: Timing of when and which processes are expected to be active in our setting.

After the players have chosen their action, we (as the researchers) ask them for their belief. At this point in time, biases like *ex-post* rationalization, hindsight bias, the consensus effect, or wishful thinking (and the corresponding underlying processes) play out and re-shape the latent belief into the final belief report. As pointed out in the introduction, we will have to re-adjust Figure 1 at the end of our study, eliminating hindsight bias and wishful thinking, and placing *ex-post* rationalization after the consensus effect.

As a general notion to organize our treatment-specific predictions, our treatments focus on the existence or absence of a direct interaction with the 'belief target', and on the number of people in the 'belief target'. This pushes participants into thinking about equivalent strategic problems from different perspectives and to focus their thinking on different aspects of the problem.

The opponent treatment prompts people to think about their specific interaction partner, even if this player is merely the one they are randomly matched to. In this treatment, it seems more natural to think about the individual incentives of both players and about their common strategic interaction. The random-other treatment also focuses on an individual person, but since there is no interaction between the players, the strategic aspect is much weaker. At the same time, in our experiments the random other person is facing exactly the same situation as the participant reporting a belief. The population treatment invokes a picture of many other people in the same situation. Arguably, the strategic aspects may play least of a role when thinking about the problem on such a 'gobal' scale.

'*Ex-Ante* Rationalization'

What beliefs would we expect in the absence of any biases? Beliefs depend crucially on the strategic situation. Put differently, a given game and its payoffs will influence a participant's beliefs and actions. In particular, we would expect beliefs and actions that are consistent with each other, because otherwise

the player would be making a mistake in at least one of the two decisions. So, what do we learn when action and belief are consistent? Likely, the agent went through one of two processes: making up a belief and best-responding to it ('*ex-ante* rationalization'), or first choosing an action by any process whatsoever and only then making up a belief consistent with the action. This reversed process (action-then-belief) may be due to the agent's wish to appear consistent (*ex-post* rationalization, Eyster, 2002; Yariv, 2005; Charness & Levin, 2005; Falk & Zimmermann, 2013) or to wishful thinking. We discuss both biases in the following.

We expect *ex-ante* rationalization to be present under all of the treatments, as we are not aware of any study documenting that participants' actions would be overall inconsistent with beliefs.

Ex-Post Rationalization

Humans are extremely good at rationalizing whatever they do (so much that certain psychologists even think that beliefs virtually always go second, Chater, 2018). The specific reasons for such *ex-post* rationalization may vary. In the context of our setup, they include cognitive dissonance (Festinger, 1957); social-desirability bias (Edwards, 1953; if participants believe that experimenters expect or like to see consistent behaviour); illusion of control (Langer, 1975; if participants have the perception that they can magically influence the matching); and confirmation bias (Wason, 1960; conceptually more of a stretch). For the purpose of this paper, we subsume all of the above processes under the header of *ex-post* rationalization.

To derive our predictions, we assume cognitive dissonance to be the driving force behind *ex-post* rationalization, noting that the other processes will lead to similar predictions. According to cognitive dissonance, an agent will adjust her beliefs when she cannot come up easily with an 'excuse'.

In a random-other frame, there is no need to align a belief report with an action: while the player's action may have been suboptimal had she played the random other participant, it need not be suboptimal because the actual opponent will/might have done something completely different. Even in a population frame, it is possible to come up with such an excuse for an action being at odds with the belief: *most* others will have acted the way I indicated in my belief report, but *my opponent* is different (and therefore, my action is still 'ok'). It is only in an opponent frame that such an excuse is no longer available, given the belief

is exactly about the player whose choice is potentially relevant for the participant's action-related payoff. Hence, *ex-post* rationalization will affect beliefs in the opponent treatments but not in the random-other or population treatments.

Hindsight Bias

Under a hindsight bias (Fischhoff, 1975), agents overestimate the *ex-ante* probability of an event after learning that the event has materialized. Thus, the hindsight bias is a specific form of information projection (Madarász, 2012). According to information projection, agents cannot abstract from their own information when assessing what other players know. In the special case of the hindsight bias, agents cannot abstract from information that became available only later on when assessing what they or others did before the information became available. Meta-analyses such as Christensen-Szalanski & Willham (1991) and Guilbault *et al.* (2004) underline the robustness of this effect.

Applied to our setting, a hindsight bias means that players are unable to abstract from the information they have (about their own action) when reporting a belief about others' behaviour. Players with a hindsight bias hence form their belief (as if they were) assuming the other players should have anticipated the hindsight-biased player's own choice. Therefore, a hindsight bias increases the probability mass placed on the other player(s) playing a best-response to the hindsight-biased player's chosen action.

We expect that a hindsight bias will exclusively occur in the opponent treatments, because the hindsight relates to the event that my *matching partner* chooses a best response to my own action. In a random-other treatment, the object of belief elicitation does not interact with me. So, this other person will be best-responding to somebody else, which means that the information about my choice should not affect his behaviour. Similarly, the population of other players will mostly best-respond to other people, which means the information about my choice will hardly influence their choices.

Wishful Thinking

A large body of literature studies *unrealistic optimism*, which is described as a tendency to hold overoptimistic beliefs about future events (e.g. Camerer & Lovallo 1999, Larwood & Whittaker 1977, Svenson 1981 or Weinstein 1980, 1989).

Wishful thinking has been brought forward as a possible cause of unrealistic optimism and has been described as a desirability bias (Babad & Katz 1991, Bar-Hillel & Budescu, 1995). Wishful thinking hence means a subjective overestimation of the probability of favorable events (*cf.* also the closely related idea of *affect* influencing beliefs, Charness & Levin, 2005). Despite the large body of evidence on human optimism (Helweg-Larsen & Shepperd, 2001), there is some doubt about whether a genuine wishful-thinking effect truly exists (Krizan & Windschitl, 2007, Bar-Hillel *et al.* 2008, Harris & Hahn, 2011, Shah *et al.,* 2016). In the context of this study, a player whose belief is influenced by wishful thinking places an unduly high subjective probability on the event that others act such that the player receives a (high) payoff.

We expect wishful thinking to be strong when the matching partner is involved, because the desirable outcome depends on this specific person, and negligible, otherwise. Hence, wishful thinking should be prevalent in the opponent treatment, but much less so in the population and random-other treatments.

Consensus Effect

The *consensus effect* is a phenomenon studied by psychologists and economists. Tversky & Kahneman (1973, 1974) link it to the *availability heuristic* and the *anchoring-and-adjustment heuristic*. Joachim Krueger describes the consensus effect in a general but simple way: *"People by and large expect that others are similar to them"* (Krueger, 2007, p. 1). The basic idea has been studied in many different contexts under many different names: [false-]consensus effect (Ross, Greene & House, 1977; Mullen *et al.*, 1985; Marks & Miller, 1987; Dawes & Mulford, 1996), perspective taking (Epley *et al.,* 2004), social projection (Krueger, 2007; 2013), type projection (Breitmoser, 2015), evidential reasoning (al-Nowaihi & Dhami, 2015) or self-similarity bias (Rubinstein & Salant, 2016).

For this study, we define the consensus effect as a psychological mechanism that distorts (reported) beliefs towards a participant's own action. A participant with a belief distorted by a consensus bias reports too high a subjective probability that others choose the same action as herself, relative to the participant's (counterfactual) unbiased belief (which the participant presumably held at the time of making her choice in the game).

Correlation neglect (of the correlation between others' choices and one's own; "illusion of validity" in Kahneman & Tversky, 1973) and conservatism in

	Population	Random Other	Opponent
Ex-ante rationalization	✓	✓	✓
Ex-Post Rationalization	-	-	✓
Wishful Thinking	-	-	✓
Consensus Effect	✓	✓	-
Hindsight Bias	-	-	✓

Table 3: Predictions of which processes are active under which treatment.

updating (about others' choices after observing one's own, Edwards, 1968) would both run exactly counter to a consensus effect, and thus would show up as a negative consensus effect (which we do not observe).

While it would be conceivable that the consensus effect is active in all treatments, we rest our hypothesis on the working paper of Rubinstein & Salant (2015): "The population frame highlights similarities among players" while "[t]he opponent frame highlights the strategic aspect of the game". Even though we are talking about symmetric games, the question seems to be about 'the other side of the interaction' (reacting to 'me') in the opponent frame, but about 'someone/many others in the same position' in the other treatments. Hence, we expect to find a consensus effect only in the population and random-other treatments.

Salience bias and Bias blind spot

People who follow a salience bias (Taylor & Fiske, 1975) will choose salient items more often. A bias blind spot means they assume that 'everybody else falls for a bias (in our study, most plausibly a salience bias) but not me' (Pronin, Lin, & Ross, 2002). Both biases may be active in our setting. However, they will act primarily *before* a participant decides on an action (and equally so across all treatments). This might seem less clear for the bias blind spot; however, if a participant thinks everybody else's choices are going to be shaped by salience, then, the participant will have held this belief already at the time of chosing an action (which in that case will be a best-reply to the belief that everybody else chooses the salient item). We are focusing on changes in a belief that happen after an action is chosen, and therefore, we leave bias blind spot and salience bias out of the equation.

Exp.	Game/Treatments	Purpose
1	Discoordination (Pop, Opp)	- Replicating that beliefs are closer to participants's actions under a population treatment than under an opponent treatment - Highlighting the consequences for measured belief-action consistency
	(RO)	- Identifying the critical treatment difference by the random-other treatment: interaction with the 'belief target', whether the 'target' is a single person or many, asking about a percentage *vs* a probability, or the exact incentivization
2.A	To-your-left (with implementation errors) (RO, Opp)	- Separating the consensus effect, hindsight bias, and wishful thinking from *ex-ante* rationalization and *ex-post* rationalization
2.B	Sequential Battle-of-the-Sexes (Pop, Opp)	- Disentangling whether in opponent treatments, ... (i) a consensus effect is overridden by *ex-post* rationalization, or (ii) whether there is no consensus effect in opponent treatments

Table 4: Overview of the experiments and their purpose. Pop stands for the population treatment, Opp for the opponent treatment, and RO for the random-other treatment.

3 Experimental Design

Rationale behind the experiments

We start this Section by describing the specific purposes of the three experiments of this paper. Experiment 1 serves three purposes. First, it replicates Rubinstein & Salant's (2015) finding that beliefs are closer to participants' own actions under a population treatment than under an opponent treatment.

Second, Experiment 1 highlights the consequences the elicitation treatment has for conclusions about participants' belief-action consistency. Third, and most importantly, it shows that the difference in behaviour between the population treatment and the opponent treatment stems from the 'interaction partner *vs.* another person' difference and not from any of the other differences.

Experiments 2.A and 2.B disentangle different mental processes that may underlie Experiment 1's findings. They provide evidence on which of the known biases and processes are important, and when. Experiment 2.A separates the consensus bias, hindsight bias, and wishful thinking from *ex-ante* and *ex-post* rationalization. In addition, we need Experiment 2.B to differentiate between two possible explanations of the data: under an opponent treatment, (i) the consensus effect is overridden by *ex-post* rationalization, and (ii) there is no consensus

effect to begin with. Table 4 summarizes the experiments and their purposes.

Experimental setup

In Experiment 1, participants face a series of 24 one-shot, two-player, four-action pure discoordination games. Players get a prize of 7€ if they choose different actions and nothing, otherwise. Participants play the discoordination games with randomly changing partners, and without any feedback in between.

Participants play the discoordination games on different sets of labels such as "1-2-3-4", "1-x-3-4", or "a-a-a-B".[11] In Experiment 2.A, we use the same general setup. However, participants play one-shot "to-your-left games" (Wolff, 2017), in which a player gets a prize of 12€ if he chooses the action immediately to the left of his opponent. The game works in a circular fashion, so that choosing "4" against a choice of "1" by your opponent would make you win the 12€ in a "1-2-3-4" setting.[12]

To separate wishful thinking from *ex-ante* rationalization and *ex-post* rationalization, we add random implementation errors to Experiment 2.A. There is a 50% probability that the computer changes a participant's decision. If the computer alters the decision, the computer chooses each box with equal probability (including the participant's chosen box). We then inform participants about whether their decision has been altered, and if so, which box the computer has chosen. If the computer changes the decision, the computer's choice is used to determine the game payoff of the participant and of her interaction partner. However, the belief elicitation still targets the other participants' original choices, not the implemented ones. Hence, *ex-ante* and *ex-post* rationalization still mean a higher probability mass on the option to the right of the participant's originally chosen option even when the computer changes the decision. In contrast, wishful thinking implies a higher probability mass on the option to the right of the implemented decision.

We elicit probabilistic beliefs directly after each choice in the game, incentivizing the belief reports via a Binarized-Scoring Rule (McKelvey & Page, 1990, Hossain & Okui, 2013). In the belief-elicitation task, subjects could earn another

[11] For the full list of label sets, see Table A1 in the appendix. All participants went through the same order of sets. We chose the varying sets to keep up participants' attention.

[12] The difference in payoffs is meant to reduce expected-earnings differences accross experiments. In a discoordination game, (both) participants are likely to win fairly often, while in the "to-your-left game", participants will win at a much lower rate.

7€. The Binarized-Scoring Rule uses a quadratic scoring rule to assign participants lottery tickets for a given prize. The lottery procedure accounts for deviations from risk neutrality and, under a weak monotonicity condition, even for deviations from expected utility maximization (Hossain & Okui, 2013). Hence, we control for participants' risk preferences (also) in the belief task.

The exact framing of the belief-elicitation question is subject to treatment variation as described in Section 1. At the end of the experiments, we randomly select two periods for payment. In one period, we pay the outcome of the game and in the other period, the belief task. In Experiment 2.A, we use an opponent and a random-other treatments since they provide the most conservative treatment comparison by changing only the identity of the target.

In Experiment 2.B, participants face two one-shot battle-of-the-sexes games, depicted in Figure 2. In each of the two games, players move sequentially but the second-mover does not receive any information on the first-mover's choice. Following the design of Blanco *et al.* (2014), there is role-reversal between the games and belief-elicitation before choices (using a binarized scoring rule with a winning prize of 6€ and a losing prize of 3€).[13] Again, if a game was payoff-relevant, the belief payment came from the other game.

This design has the feature that a first-mover in the first game will be asked about his belief about first-mover behaviour (in the second game) directly after making his choice (in the first game). And because we are eliciting a belief about other first-movers (in a new game), cognitive dissonance does not create a need for the elicited belief to be "consistent" with the participant's previous first-mover choice (all of the above applies in exactly the same way to participants who acted as second-movers in the first game).[14]

We use a different game than in Experiment 1 and Experiment 2.A because we need different player roles (*i.e.*, an asymmetric game) to get rid of cognitive dissonance. While, technically, implementing a sequential version of the discoordination or to-your-left games would suffice, in neither of the two games the

[13] In our experiment, there is random re-matching between the two games.

[14] In case this still sounds confusing, it may be easier to think what would happen if we used a different setup. If we asked about one's opponent in the same game, cognitive dissonance would apply. Asking about one's peers in the next game while maintaining roles would not allow for an opponent treatment. If there was only a single role we might re-introduce cognitive dissonance (the case for social-desirability concerns would be less clear). So, in order to make sure cognitive dissonance should not be playing a role, we need an asymmetric game played twice, with role-reversal.

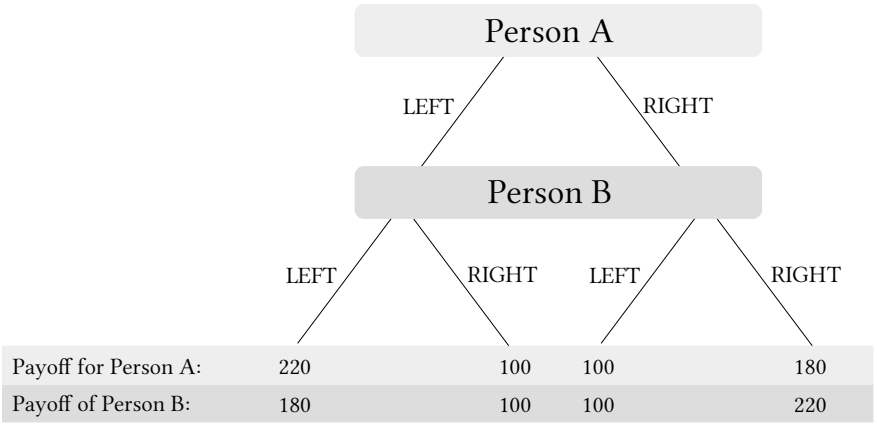


Figure 2: Battle-of-the-sexes game used in Experiment 2.B. The rounded boxes represent information sets: Person B does not learn Person A's choice before the end of the game.

sequentiality would 'make sense' for participants: we conjectured that the asymmetry would not be strong enough. In contrast, in sequential battle-of-the-sexes games like the one we use, the sequentiality has been shown to affect behaviour strongly (Cooper *et al.*, 1993). Finally, we use an opponent and a population treatments in order to induce the largest-possible treatment difference in terms of a consensus effect (judging by the results of Experiment 1).

Procedures

We programmed the experiments using z-Tree (Fischbacher, 2007) and conducted them in the LakeLab at the University of Konstanz. We use the data of 145 participants from Experiment 1, 70 participants from Experiment 2.A, and 222 participants from Experiment 2.B.[15] Experiment 2.B was run as one out of three parts of an experimental session; for 118 participants, this was the first part of the session, for another 104 participants, it was the second part of the session (in the first part, these participants had to bet on the colour of a ball after being shown differing samples of green and blue balls). In both types of sessions, one of the three parts would be paid out, with an exchange rate of 20 experimental currency units per Euro. We used ORSEE (Greiner, 2015) for recruitment. All sessions lasted between 60 and 90 minutes.

[15]For the analysis, we exclude one participant from Experiment 1 who always reported a 100% belief of not having discoordinated. This participant probably tried to hedge, but did not understand that hedging was impossible. We used all data from Experiments 2.A and 2.B.

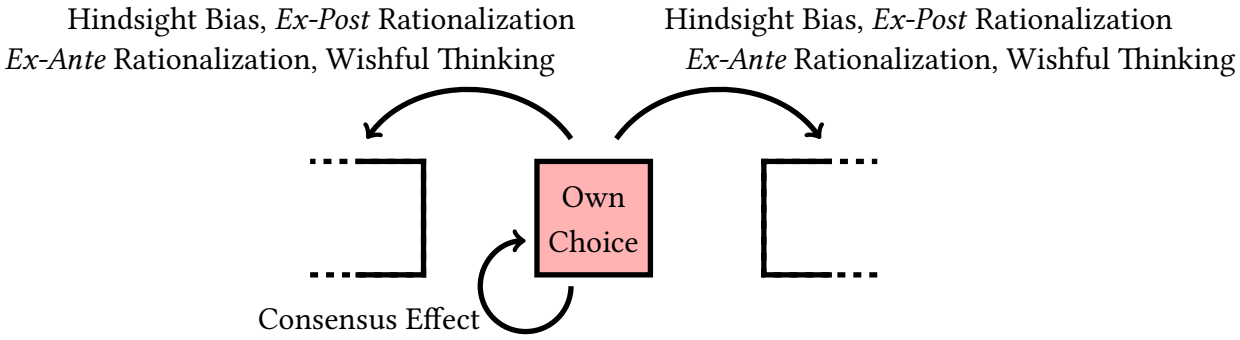


Figure 3: Predictions of the candidate processes in the discoordination game. We indicate the predictions by arrows: The consensus bias will increase the probability mass placed on the other player(s) making the same choice as the observed player, while the other four processes will increase the probability mass placed on the non-chosen options.

4 Framing effects on belief reports, behaviour, and the implications for belief-action consistency

Predictions for Experiment 1

Recall that Experiment 1 had participants play a pure discoordination game with four options. We illustrate which of the psychological processes would load on which options in Figure 3. As summarized in Table 3, we expected to observe *ex-ante* rationalization and a consensus effect in the population and random-other treatments, and *ex-ante* rationalization, *ex-post* rationalization, wishful thinking, and a hindsight bias in the opponent treatment. Consequently, we expected a lower probability mass on participants' own choices in the opponent treatment, leading to higher observed best-response and lower observed 'worst-response' rates. A 'worst-response' means that the participant chooses the action his opponent is most likely to choose, as judged by the participant's reported belief.

Results of Experiment 1

Figure 4 summarizes beliefs and belief-action consistency for the three treatments. For the analysis, we aggregate the data on the individual level across all periods, as we have one independent observation per participant (re-call that

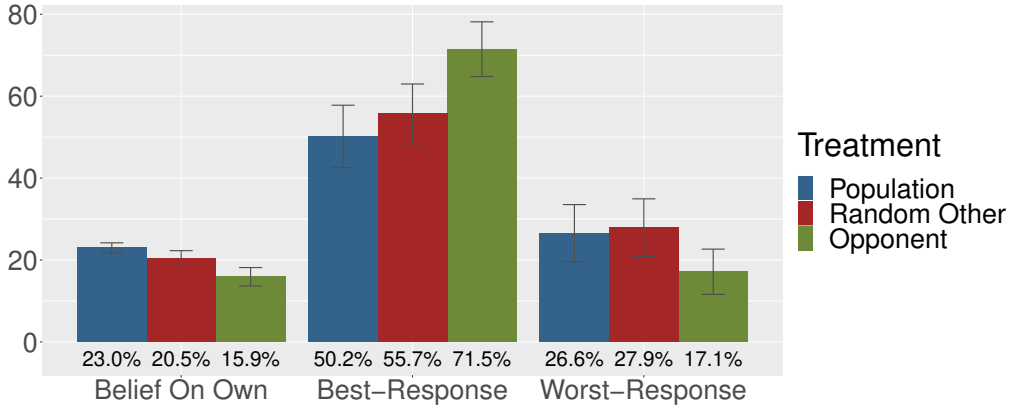


Figure 4: Beliefs and consistency in Experiment 1. Error bars indicate 95% confidence intervals. For all tests, the data is aggregated on the individual level across all periods, yielding one independent observation per participant.

we did not give feedback). For each participant, we look at the probability that the reported belief places on the participant's own action in the corresponding game, averaged across all 24 periods. This is the participant's average subjective probability that (s)he matched the other player's/players' choice, and hence did *not* discoordinate. Similarly, we compute the best- and 'worst-response' rate to beliefs for each participant individually.

The mean average belief on the participant's own action (Figure 4, left panel) is significantly higher in the population treatment and the random-other treatment compared to the opponent treatment (rank-sum tests, population/opponent: $p < 0.001$ and random-other/opponent: $p < 0.001$). The effect is strong enough to impede consistency: compared to the opponent treatment, the average observed best-response rate is lower (mid panel, $p < 0.001$ and $p = 0.004$) and the average worst-response rate is higher (right panel, $p = 0.026$ and $p = 0.019$) in the population treatment and the random-other treatment.[16] The reduction in the observed best-response rate of 16-21 percentage-points and a 9.5 percentage-point increase in the worst-response rate in the population treatment are considerable effect sizes. Note that the observed worst-response rates differ by more

[16]The differences between population and random-other treatment are not significant. Rank-sum tests, beliefs: $p = 0.146$, best-response rates: $p = 0.237$, worst-response rates: $p = 0.822$.

than 50% of the rate in the opponent treatment.

Summary of Part 1

Up to this point, we have documented a considerable framing effect. Most notably, beliefs differ in the *ceteris-paribus* comparison between the opponent and the random-other treatments, where we vary only whether a participant interacts directly with the 'target participant' of the belief. Additionally, the differences in reported beliefs influence observed best- and worst-response rates and hence affect the interpretation of actions and beliefs by the experimenter. What Experiment 1 does not show is whether the differences between the treatments occur because there is (more) consensus under the population and random-other treatments, or because there is (more) hindsight bias, wishful thinking, *ex-ante* or *ex-post* rationalization under the opponent treatment.[17] To disentangle these processes, we need Experiments 2.A and 2.B.

5 Disentangling the Processes

5.1 Experiment 2.A: Isolating Consensus Bias, Hindsight Bias, and Wishful Thinking

Experiment 2.A disentangles the influences of a consensus bias, hindsight bias, and wishful thinking from *ex-ante*/*ex-post* rationalization. For this purpose, we use the "to-your-left game", in which a player wins a prize of 12€ if she chooses the option to the immediate left of the other player's choice (with the right-most option winning against the left-most option).

Predictions for Experiment 2.A

Figure 5 visualizes the predictions of our candidate processes in Experiment 2.A. Because the game is circular, only the relative position of the respective box

[17] The fact that the average probability mass placed on a participants' own choice was below 25% for all treatment could be interpreted as suggesting that there is no consensus effect at all. However, recall that we are talking about a discoordination game in which it makes sense to choose the option that others are least likely to choose. Hence, probability masses of less than 25% are exactly what we should expect *a priori*. The consensus effect simply does not seem to be strong enough to distort beliefs so that the (average) probability mass surpasses 25%.

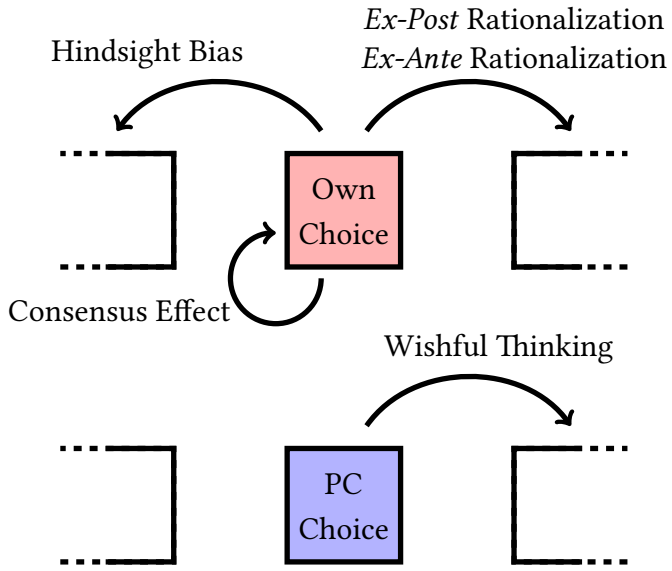


Figure 5: Predictions of the candidate processes in the to-your-left game with implementation errors in case of an implementation error. We color example choices and indicate by arrows the predictions: A consensus bias increases the probability mass placed on the other player(s) making the same choice as the observed player; hindsight bias increases the probability mass on the option to the left of the player's chosen option, while *ex-post* and *ex-ante* rationalization increase the probability mass placed on the option to the right. Wishful thinking increases the probability of the option to the right of the option implemented by the computer.

matters and not the actual position.

In the to-your-left game, a consensus bias still would increase the belief-probability mass participants place on their own actions. A hindsight bias would increase the probability mass on the option immediately to the left of participants' choices, because in hindsight, it would be obvious what the participant's opponent should have chosen in response to the participant's own action. *Ex-ante* and *ex-post* rationalization, and wishful thinking, on the other hand, would increase the probability mass on the option immediately to the right of participants' chosen actions.

To distinguish the effect of wishful thinking, we focus on periods in which the computer changed the selected box. In these periods, wishful thinking should increase the probability mass placed on the option to the right of the computer's

choice. In contrast, *ex-ante* and *ex-post* rationalization yield a higher probability mass on the option to the right of the participant's choice.[18]

Results of Experiment 2.A

We analyze the data from Experiment 2.A with linear dummy regressions reported in Table 5. The dependent variable is the reported belief on a single box. Every participant reports 24 Periods $\times$ 4 Boxes = 96 belief probabilities on single boxes. We regress the beliefs on a set of dummies, indicating whether the particular reported probability would be influenced by an existing consensus bias, wishful thinking (WT), hindsight bias, or *ex-ante*/*ex-post* rationalization (EAR/EPR) according to the predictions indicated above. Further, we use a treatment dummy which is equal to 1 in the random-other treatment and 0 in the opponent treatment. The constant of this regression is a neutral belief where all dummies are zero. Hence such a belief is unaffected by any of the processes we study.

Model 1 uses all observations where the participant made the ultimate decision.[19] Wishful thinking and EAR/EPR cannot be distinguished for the undistorted choices, as both load on the probability to the immediate right of the participant's choice. We hence have to use two separate regressions for the situations with and without implementation error because by design, the interaction EAR/EPR $\times$ WT is perfectly collinear with the implementation error.

Model 1 shows evidence for a consensus bias only in the random-other treatment. There is no evidence for a hindsight bias. Further, probabilities to the right of the chosen option (influenced by EAR/EPR and/or WT) are twice the size of a neutral belief. This huge effect in the opponent treatment is reduced in the random-other treatment. We argue that this reduction is indirect evidence of

[18]Note that depending on which box the computer selected, two different processes may increase the belief-probability mass on the same option. We control for this in the analysis.

[19]The observations where the computer truly altered the decision are analysed in Model 2. All results in Model 1 are robust to adding trials to the sample in which the computer decided but happened to choose the same action as the participant. A regression that controls for trials in which the computer randomly implemented the same option as the participant detects no significant differences between the two situations. The regression has an additional dummy for 'same choice by computer' which we interact with all six exogenous variables from Model 1. We report the regression in Table A1 in the Appendix.

Single Belief	Model 1	Model 2
Consensus	-0.127	0.701
	(2.132)	(1.980)
Consensus × Random-Other Treatment	7.677	-0.043
	(2.802)	(2.165)
Hindsight Bias	-1.729	-1.211
	(1.819)	(1.799)
Hindsight Bias × Random-Other Treatment	1.481	-1.839
	(2.070)	(2.195)
Belief to the right (EAR/EPR & WT)	19.353	
	(3.436)	
Belief to the right (EAR/EPR & WT) × Random-Other Treatment	-6.650	
	(3.924)	
EAR/EPR		8.690
		(2.529)
EAR/EPR × Random-Other Treatment		-2.257
		(2.588)
Wishful thinking (WT)		-0.451
		(1.081)
Wishful thinking (WT) × Random-Other Treatment		2.364
		(2.542)
Neutral Belief (constant)	20.301	23.282
	(1.031)	(0.870)
Implementation error	No	Yes
Number of Observations	3332	2532
Number of Clusters	70	70
R^2	0.1254	0.0389

Table 5: Linear dummy regressions of the belief probability on a given option. Standard errors in parentheses clustered on subject level. EAR stands for *ex-ante*, EPR for *ex-post* rationalization, and WT for wishful thinking.

ex-post rationalization.[20]

Ex-post rationalization should occur exclusively (or at least to a much larger degree) in the opponent treatment: believing that *some other* player chose an option that would be bad for us need not cause cognitive dissonance, because our opponent still might have chosen something else. In contrast, if we state a belief that our *opponent* chose something that would be bad for us given our action, this should indeed cause cognitive dissonance in us. Therefore, the coefficient of

[20] An additional experiment reported in Appendix B provides direct evidence. The setup mirrored that of Experiment 1, except for asking for beliefs before actions. The reversed order should eliminate *ex-post* rationalization as *ex-post* rationalizing a belief by an action is unintuitive: once we form a belief (as in the first stages of the additional experiment), there is no good reason to form yet a different belief that we then contradict out of a taste for consistency. We indeed no longer find find a difference between the treatments, which is due to players placing a higher probability mass on their own action in the opponent treatment, in line with our prediction.

"Belief to the right" (with Frame = 0) should capture the added effects of *ex-ante* and *ex-post* rationalization. In contrast, the "Belief to the right" in the random-other treatment (Frame = 1) should capture *ex-ante* rationalization only. Hence, the interaction effect "Belief to the right × Frame" provides an estimate for the differential effect of *ex-post* rationalization. Like in Experiment 1, the average best-response rate is higher in the opponent treatment than in the random-other treatment when the computer does not change the decision (opponent: 62.1%, random other: 45.2%, rank-sum test $p = 0.006$).[21]

Model 2 includes all decisions where the computer really changed the participant's decision. Hence, Model 2 includes all observations in which the computer decided and did not choose the same action as the participant. There is no more consensus effect in either treatment. Also, there is no evidence for wishful thinking or a hindsight bias. However, EAR/EPR loads on beliefs to the right of the participant's decision also in the randomly altered trials. Further, (neutral) beliefs are closer to uniformity in the random-action trials. The results of Model 2 are robust to including all possible remaining dummy interactions.[22]

Discussion of Experiment 2.A

We interpret the results in the following way: there is a consensus bias in the random-other treatment. There is *ex-ante* rationalization in both treatments, but it is stronger in the opponent treatment. We argue that this difference is due to *ex-post* rationalization, which is less important or absent in the random-other treatment. Finally, a hindsight bias does not seem to play a role. As in Experiment 1, the framing differences in Model 1 affect measured belief-action consistency, with higher observed best-response rates under the opponent treatment compared to the random-other treatment.

When the computer overrides participants' decisions, a certain degree of *ex-ante* rationalization survives in the reported beliefs: also in such cases, participants *on average* seem to report beliefs that make sense given their actions,

[21] The difference in worst-response rates is not significant. Opponent: 20.9%, random other: 22.8%, $p = 0.780$.

[22] The interactions are: (Consensus × Wishful thinking), (Consensus × Wishful thinking × Treatment), (Hindsight Bias × Wishful Thinking) and (Hindsight Bias × Wishful Thinking × Treatment).

despite the fact that beliefs are closer to uniformity.[23] However, there are no more significant framing differences in beliefs or best-response rates with implementation errors. It seems as if the random implementation error detaches participants to a certain degree from the action choice altogether. We also do not see any evidence for wishful thinking, even though wishful thinking does not relate to the chosen action.

We ran Experiment 2.A to disentangle consensus bias, hindsight bias, and—albeit with a caveat—wishful thinking from *ex-ante*/*ex-post* rationalization. Experiment 2.B shows that there is as much of a consensus bias in an opponent treatment as in a population treatment, once we eliminate the cognitive need for *ex-post* rationalization in the opponent treatment.

5.2 Experiment 2.B: Consensus Effect in Opponent Treatments?

In Experiment 2.B, participants play two rounds of the sequential battle-of-the-sexes game depicted in Figure 2, with role-reversal between the games, belief-elicitation before choices, and random rematching between rounds. To study whether a consensus effect exists also in an opponent treatment, we contrast beliefs in such a treatment with beliefs from a population treatment (where we know a strong consensus effect exists).

Predictions for Experiment 2.B

As we outlined above, cognitive dissonance should not affect behaviour in Experiment 2.B, neither in the population nor in the opponent treatment. Hence, *ex-post* rationalization should be eliminated in the opponent treatment. If under an opponent frame, a consensus effect does not exist, we should nevertheless see a treatment difference: in that case, the probability mass placed on a participant's prior action should be higher in the population frame (where we know the consensus effect is at work) than in the opponent frame. If, on the other hand, there is a consensus effect in the opponent frame that is just 'over-written' by *ex-post* rationalization in more standard designs (such as Experiment 1 or Experiment

[23] The reduced average difference to uniformity is only very partially due to a difference in the prevalence of uniform beliefs: under implementation errors, 5% of the reported beliefs are uniform, and without errors, 4%.

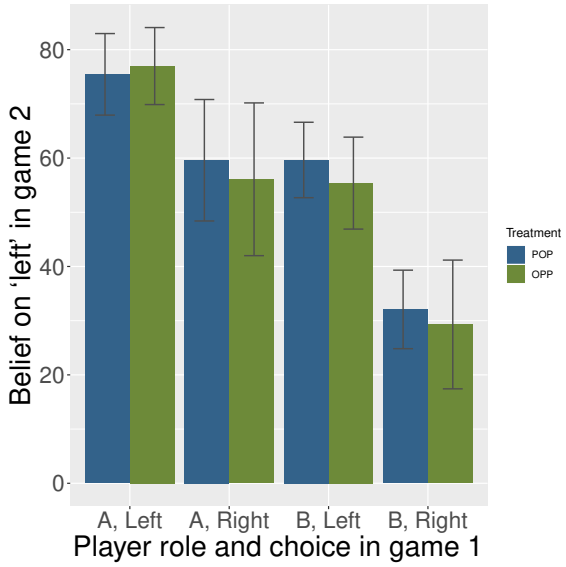


Figure 6: Probability mass placed on "left" in participants' belief reports for game 2, by their role and decision in game 1. Whiskers indicate 95% confidence intervals.

2.A), we should no longer see a difference between the treatments.

Results of Experiment 2.B

The data generally look as expected given the literature.[24] We thus can focus on our research question and look at participants' beliefs for game 2 depending on their choices in game 1. Figure 6 visualizes the results for both player roles and both treatments. First of all, note that we observe a clear consensus effect for either role in both treatments: players who chose "left" in game 1 place more probability mass on others (who 'now'—in game 2—have the role they used to have in game 1) also choosing "left", compared to players who chose "right". This holds for both players A and B. Moreover, there clearly are no more treatment differences between the opponent treatment and the population treatment.

To support the conclusion statistically, we run the linear-probability regression reported in Table 6. As can be seen from the Table, the participant's previ-

[24]Participants in both player roles chose "left" far more often than "right": 74% of As and 70% of Bs in the first game, and 75% of As and 76% of Bs in the second game. These fractions roughly correspond to participants' beliefs: in both games, As expected Bs to play "left" with an average probability of 50-51% (40% would make linear-utility As indifferent), and Bs expected As to play "left" with an average probability of 71%.

	Belief on Left (in %)	$Pr(>\|t\|)$
(Intercept)	34.8 (4.3)	$3 \cdot 10^{-14}$
Person A in Game 1	18.8 (4.4)	$3 \cdot 10^{-05}$
Chose "Left" in Game 1	23.5 (4.8)	$2 \cdot 10^{-06}$
Opponent Frame	−3.1 (6.6)	0.641
Person A in Game 1 × Opponent Frame	4.1 (6.3)	0.509
Chose "Left" in Game 1 × Opponent Frame	−0.6 (6.9)	0.928
Number of Observations	222	
R^2	0.306	

Table 6: Linear dummy regressions of the probability mass placed on "left" for game 2, on the participant's role and decision in game 1. Standard errors in parentheses.

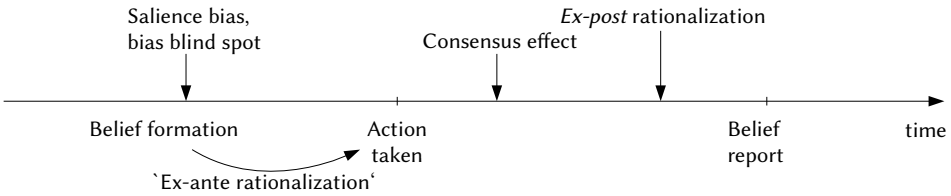


Figure 7: Timing of when and which processes are assumed to be active, taking into account this paper's findings. The Figure is reduced to the three processes we find evidence for, and it implies that the consensus effect and *ex-post* rationalization are serial processes rather than alternative processes that happen at the same time.

ous choice (when the participant was playing in the role that the belief's target is playing now) clearly has an influence on the belief, while the treatment variable (or any of its interactions) does not.

Our results mean that when participants do not have any need to *ex-post* rationalize their actions, they exhibit the same degree of consensus effect under an opponent frame as under a population frame. As a consequence, we have to revise our conceptual picture from Section 2. Figure 7 shows the updated 'model' of participants' belief-report formation. It is reduced to the three processes we find evidence for, and it implies that the consensus effect and *ex-post* rationalization are not two alternative processes that might take effect at a similar point in time. Instead, consensus effect and *ex-post* rationalization seem to be *serial* processes that may be invoked one after the other.

6 Conclusion

When studying beliefs, researchers have several choices to make, among them, whether to ask participants about the actions of their opponent(s) or about the actions of unrelated others.[25] None of these choices is trivial, and a review of the literature reveals that different researchers make different choices. However, the choices are rarely motivated in the final publication. We claim that the reason is that the exact consequences of each alternative are unknown so far.

In this paper, we show that in particular the choice between an opponent treatment (asking about the opponent's action), a random-other treatment (asking about somebody else's action), and a population treatment (asking about everybody else's action) is by no means innocuous. Asking about others' choices induces belief reports to be affected by a consensus effect in any treatment. However, if the study uses an opponent treatment and actions are strategic substitutes, the latent belief changes (again). In such cases, the reported belief will reflect *ex-post* rationalization.

Our findings thus provide an explanation for the puzzle that, so far, all economics papers documenting a consensus effect have relied on a population treatment: It is only when actions are strategic substitutes that we can discern a consensus effect from *ex-ante* (or *ex-post*) rationalization. However, when actions are substitutes, reporting a belief that is influenced by a consensus effect seems particularly 'bad'. It would mean the participant expects others to make the same choice with a comparatively high probability, in which case the participant should have made a different choice to begin with. This is precisely the type of situation in which an opponent-oriented question leads to cognitive dissonance, and thus, *ex-post* rationalization (random-other and population treatments always offer an excuse for belief-action inconsistencies in that "*my* opponent is different"). In other words, in settings that allow to single out a consensus effect, we will observe the effect only under a belief-elicitation task that does *not* target the participants' opponent.

Our second research question was whether the literature was overlooking

[25]Note that some researchers may avoid asking about a participant's opponent even in a one-shot design because they are afraid of hedging attempts by their participants, which is not an issue in our study. In the discoordination games we study, increased hedging when asking about the opponent would lead to the exact opposite of what we find, so that this is not an issue here. Note further that we preclude rational hedging by never paying both an action and the corresponding belief.

other processes that are relevant for belief reports on top of *ex-ante* rationalization, the consensus effect, and *ex-post* rationalization. This would not have been surprising given the huge number of known biases in the literature. In adding potential biases to the list, we restricted ourselves to biases that we could easily apply given our main interest in understanding the interplay of belief-elicitation treatments with the three 'standard' processes.

Reassuringly for our interpretation of the literature, we find clear evidence only for *ex-ante* rationalization, a consensus effect, and *ex-post* rationalization. And while we cannot identify the exact process behind participants' *ex-post* rationalization, such rationalization shows exactly in those cases when cognitive dissonance or a social-desirability bias (assuming consistent behaviour to be socially desirable) would suggest it should show.

Recommendations. Our results show that we need to take the substantial framing differences into account when analysing existing data or designing new surveys and experiments. In particular, in designing new experiments, we propose to use random-other or population treatments, bearing in mind that the reports will be influenced by social projection. Choosing the alternative—an opponent treatment—means that reported beliefs may lose any connection to the "true beliefs" (the belief at the time of choosing the action) altogether. This danger is present particularly when actions are strategic substitutes.

We also recommend considering to elicit beliefs prior to actions, given that this will prevent consensus effects and *ex-post* rationalization. In our experience, it does not lead to excessively high measured best-response rates (a common concern against such a procedure; see, *e.g.*, the additional experiment reported in Appendix B). However, we already know that under certain circumstances, it will change behaviour (Rutström and Wilcox, 2009).

Our findings show that there may not be an 'innocent' belief-elicitation method. In our study, participants faced a strong monetary incentive to report their true beliefs. Moreover, we incentivized belief reports by a state-of-the-art mechanism that is proper even for people who do not comply with expected-utility maximization (as long as they comply with a weak monotonicity condition, Hossain & Okui, 2013). And still, we have not found a way of asking for a belief that leads to an unbiased belief report without running the risk of changing behaviour.

References

al-Nowaihi, A., & Dhami, S. (2015). Evidential Equilibria: Heuristics and Biases in Static Games of Complete Information. *Games*, 6(4), 637-676.

Armantier, O., & Treich, N. (2013). Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging. *European Economic Review*, 62, 17-40.

Babad, E., & Katz, Y. (1991). Wishful thinking—against all odds. *Journal of Applied Social Psychology*, 21(23), 1921-1938.

Bauer, D., & Wolff, I. (2017). Belief uncertainty and stochastic choice. *Mimeo*.

Bar-Hillel, M., & Budescu, D. V. (1995). The elusive wishful thinking effect. *Thinking & Reasoning*, 1(1), 71-103.

Bar-Hillel, M., Budescu, D. V., & Amar, M. (2008). Predicting World Cup results: Do goals seem more likely when they pay off? *Psychonomic Bulletin & Review*, 15(2), 278-283.

Bellemare, C., Kröger, S., & Van Soest, A. (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica*, 76(4), 815-839.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H. T. (2010). Belief elicitation in experiments: is there a hedging problem?. *Experimental Economics*, 13(4), 412-438.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H. T. (2014). Preferences and beliefs in a sequential social dilemma: a within-subjects analysis. *Games and Economic Behavior*, 87, 122-135.

Breitmoser, Y. (2015). Knowing me, imagining you: Projection and overbidding in auctions. *Working paper*, accessed 2017/09/06, https://mpra.ub.uni-muenchen.de/62052/

Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review*, 89(1), 306-318.

Charness, G., & Grosskopf, B. (2001). Relative payoffs and happiness: an experimental study. *Journal of Economic Behavior & Organization*, 45(3), 301-328.

Charness, G., & Levin, D. (2005). When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. *The American Economic Review*, 95(4), 1300-1309.

Chater, N. (2018). *The mind is flat: the remarkable shallowness of the improvising brain.* Yale University Press, New Haven, USA.

Christensen-Szalanski, J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 48(1), 147-168.

Cooper, R., DeJong, D., Forsythe, R. and Ross, T. (1993). Forward Induction in the Battle-of-the-Sexes Games. *American Economic Review*, 83(5), 1303-1316.

Costa-Gomes, M. A., & Weizsäcker, G. (2008). Stated beliefs and play in normal-form games. *The Review of Economic Studies*, 75(3), 729-762.

Critcher, C. R., & Dunning, D. (2013). Predicting persons' versus a person's goodness: Behavioral forecasts diverge for individuals versus populations. *Journal of Personality and Social Psychology*, 104(1), 28.

Critcher, C. R., & Dunning, D. (2014). Thinking about Others versus Another: Three Reasons Judgments about Collectives and Individuals Differ. *Social and Personality Psychology Compass*, 8(12), 687-698.

Danz, D. N., Fehr, D., & Kübler, D. (2012). Information and beliefs in a repeated normal-form game. *Experimental Economics*, 15(4), 622-640.

Danz, D. N., Madarász, K., & Wang, S. W. (2014). The Biases of Others: Anticipating Informational Projection in an Agency Setting. *Working Paper*, accessed 2017/06/06, http://works.bepress.com/kristof_madarasz/42/ ,

Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment?. *Organizational Behavior and Human Decision Processes*, 65(3), 201-211.

Delavande, A., Giné, X., & McKenzie, D. (2011a). Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design? *Journal of Applied Econometrics*, 26(3), 479-497.

Delavande, A., Giné, X., & McKenzie, D. (2011b). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics*, 94(2), 151-163.

Dhami, S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford University Press, Oxford, UK.

Edwards, A. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 37(2), 90-93.

Edwards, W. (1968). Conservatism in human information processing. In: Kleinmutz, B. (Ed.), Formal Representation of Human Judgement. New York: Wiley, 17-52.

Engelberg, J., Manski, C. F., & Williams, J. (2011). Assessing the temporal variation of macroeconomic forecasts by a panel of changing composition. *Journal of Applied Econometrics*, 26(7), 1059-1078.

Engelmann, D., & Strobel, M. (2012). Deconstruction and reconstruction of an anomaly. *Games and Economic Behavior*, 76(2), 678-689.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87(3), 327.

Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *The Journal of Economic Perspectives*, 30(3), 133-140.

Eyster, E. (2002). Rationalizing the past: A taste for consistency. *Working paper*, accessed 2017/06/06, http://www.lse.ac.uk/economics/people/facultyPersonalPages/facultyFiles/ErikEyster/RationalisingThePastATasteForConsistency.pdf

Falk, A., & Zimmermann, F. (2013). A taste for consistency and survey response behaviour. CESifo Economic Studies, 59(1), 181-193.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3), 817-868.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford: Stanford University Press.

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

Fischhoff, B. (1975). Hindsight $\neq$ foresight: the effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288-299.

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox*. Cambridge: MIT press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York, Cambridge university press.

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

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, 26(2-3), 103-117.

Guiso, L., & Parigi, G. (1999). Investment and demand uncertainty. *The Quarterly Journal of Economics*, 114(1), 185-227.

Harrison, G. W., Martínez-Correa, J., & Swarthout, J. T. (2014). Eliciting subjective probabilities with binary lotteries. *Journal of Economic Behavior & Organization*, 101, 128-140.

Harris, A. J., & Hahn, U. (2011). Unrealistic optimism about future life events: a cautionary note. *Psychological Review*, 118(1), 135.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review*, 5(1), 74-95.

Hossain, T., & Okui, R. (2013). The binarized scoring rule. *The Review of Economic Studies*, 80(3), 984-1001.

Hollard, G., Massoni, S., & Vergnaud, J. C. (2016). In search of good probability assessors: an experimental comparison of elicitation rules for confidence judgments. *Theory and Decision*, 80(3), 363-387.

Holt, C. A., & Smith, A. M. (2016). Belief Elicitation with a Synchronized Lottery Choice Menu That Is Invariant to Risk Attitudes. *American Economic Journal: Microeconomics*, 8(1), 110-139

Hyndman, K. B., Terracol, A., & Vaksmann, J. (2013). Beliefs and (in) stability in normal-form games. *Working paper*, accessed 2017/06/14, http://lemma.u-paris2.fr/sites/default/files/concoursMCF/Vaksman.pdf.

Hyndman, K., Ozbay, E. Y., Schotter, A., & Ehrblatt, W. Z. (2012). Convergence: an experimental study of teaching and learning in repeated games. *Journal of the European Economic Association*, 10(3), 573-604.

Iriberri, N., & Rey-Biel, P. (2013). Elicited beliefs and social information in modified dictator games: What do dictators believe other dictators do? *Quantitative Economics*, 4(3), 515-547.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237-251.

Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica*, 77(2), 603-606.

Khwaja, A., Sloan, F., & Salm, M. (2006). Evidence on preferences and subjective beliefs of risk takers: The case of smokers. *International Journal of Industrial Organization*, 24(4), 667-682.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, 133(1), 95.

Krueger, J. I. (2007). From social projection to social behaviour. *European Review of Social Psychology*, 18, 1-35.

Krueger, J. I. (2013). Social projection as a source of cooperation. *Current Directions in Psychological Science*, 22(4), 289-294.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32, 311-328.

Larwood, L., & Whittaker, W. (1977). Managerial myopia: Self-serving biases in organizational planning. *Journal of Applied Psychology*, 62(2), 194

Madarász, K. (2012). Information projection: Model and applications. *The Review of Economic Studies*, 79(3), 961–985.

Manski, C. F. (2002). Identification of decision rules in experiments on simple games of proposal and response. *European Economic Review*, 46(4), 880-891.

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

Marks, G., & Miller, N. (1987). Ten years of research on the false consensus effect: An empirical and theoretical review. *Psychological Bulletin*, 102(1), 72.

McKelvey, R. D., & Page, T. (1990). Public and private information: An experimental study of information pooling. *Econometrica*, 58, 1321-1339.

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology*, 21(3), 262-283.

Molnár, A., & Heintz, C. (2016). Beliefs About People's Prosociality Eliciting predictions in dictator games. *Working Paper*, accessed 2017/09/06, http://publications.ceu.edu/sites/default/files/publications/molnar-heintz-beliefs-about-prosociality.pdf

Nyarko, Y., & Schotter, A. (2002). An experimental study of belief learning using elicited beliefs. *Econometrica*, 70(3), 971-1005.

Palfrey, T. R., & Wang, S. W. (2009). On eliciting beliefs in strategic games. *Journal of Economic Behavior & Organization*, 71(2), 98-109.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The Bias Blind Spot: Perceptions of Bias in Self Versus Others. *Personality and Social Psychology Bulletin*, 28(3), 369-381.

Proto, E., & Sgroi, D. (2017). Biased beliefs and imperfect information. *Journal of Economic Behavior & Organization*, 136, 186-202.

Rey-Biel, P. (2009) Equilibrium play and best response to (stated) beliefs in normal form games, *Games and Economic Behavior*, 65(2), 572-585.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3), 279-301.

Rubinstein, A., & Salant, Y. (2015). "Isn't everyone like me?": On the presence of self-similarity in strategic interactions. Working paper version of Rubinstein & Salant (2016), accessed 2017/09/06, https://pdfs.semanticscholar.org/34ee/9a1799fcb4c43207136437e3a1e3c3ef25a6.pdf

Rubinstein, A., & Salant, Y. (2016). "Isn't everyone like me?": On the presence of self-similarity in strategic interactions. *Judgment and Decision Making*, 11(2), 168.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, 5(17), 61-71.

Savage, L. J. (1954) The Foundations of Statistics. New York: John Wiley and Sons. (Second ed., Dover, 1972).

Selten, R., & Ockenfels, A. (1998). An experimental solidarity game. *Journal of Economic Behavior & Organization*, 34(4), 517-539.

Schlag, K. H., Tremewan, J., & Van der Weele, J. J. (2015). A penny for your thoughts: a survey of methods for eliciting beliefs. *Experimental Economics*, 18(3), 457-490.

Schotter, A., & Trevino, I. (2014). Belief elicitation in the laboratory. *Annual Review of Economics*, 6(1), 103-128.

Shah, P., Harris, A. J., Bird, G., Catmur, C., & Hahn, U. (2016). A pessimistic view of optimistic belief updating. *Cognitive Psychology*, 90, 71-127.

Sutter, M., Czermak, S., & Feri, F. (2013). Strategic sophistication of individuals and teams. Experimental evidence. *European Economic Review*, 64, 395-410.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47(2), 143-148.

Taylor, S. E., & Fiske, S. T. (1975). Point of view and perceptions of causality. *Journal of Personality and Social Psychology*, 32(3), 439–445.

Trautmann, S. T., & van de Kuilen, G. (2015). Belief elicitation: A horse race among truth serums. *The Economic Journal*, 125(589), 2116-2135.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207-232.

Tversky, A., & Kahneman, D. (1974). Heuristics and biases: Judgment under uncertainty. *Science*, 185, 1124-1130.

Van Der Heijden, E., Nelissen, J., & Potters, J. (2007). Opinions on the tax deductibility of mortgages and the consensus effect. *De Economist*, 155(2), 141-159.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12(3), 129-140.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5), 806.

Weinstein, N. D. (1989). Effects of personal experience on self-protective behaviour. *Psychological Bulletin*, 105(1), 31.

Wolff, I. (2015). When best-replies are not in equilibrium: understanding cooperative behaviour. *Working paper*,accessed 2017/09/06, http://kops.uni-konstanz.de/handle/123456789/33027

Wolff, I. (2017). Lucky Numbers in Simple Games. *Mimeo*.

Yariv, L. (2005). I'll See It When I Believe It–A Simple Model of Cognitive Consistency. *Working Paper*, accessed 2017/06/06, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.207.2893

7 Appendix

A Figures & Tables

Single Belief	Model 1′
Consensus	-0.127 (2.133)
Consensus × Frame	7.677*** (2.804)
Belief to the right (EAR/EPR & WT)	19.353*** (3.439)
Belief to the right (EAR/EPR & WT) × Frame	-6.650* (3.926)
Hindsight Bias	-1.729 (1.820)
Hindsight Bias× Frame	1.481 (2.071)
Same Choice by the Computer	0.610 (1.121)
Consensus × Same Choice by the Computer	2.171 (2.233)
Consensus × Frame× Same Choice by the Computer	-3.127 (2.699)
Belief to the right (EAR/EPR & WT) × Same Choice by the Computer	-3.787 (3.480)
Belief to the right (EAR/EPR & WT) × Frame× Same Choice by the Computer	1.036 (4.077)
Hindsight Bias× Same Choice by the Computer	-0.200 (2.620)
Hindsight Bias× Frame× Same Choice by the Computer	0.983 (3.152)
Constant	20.301*** (1.032)
R^2	0.1190

Table A1: OLS dummy regressions of single belief elements with interactions for trials in which the computer (by chance) selected the same action as the participant. Standard errors in parenthesis clustered on subject level (70 clusters). Asterisks: *** $p < 0.01$, * $p < 0.1$

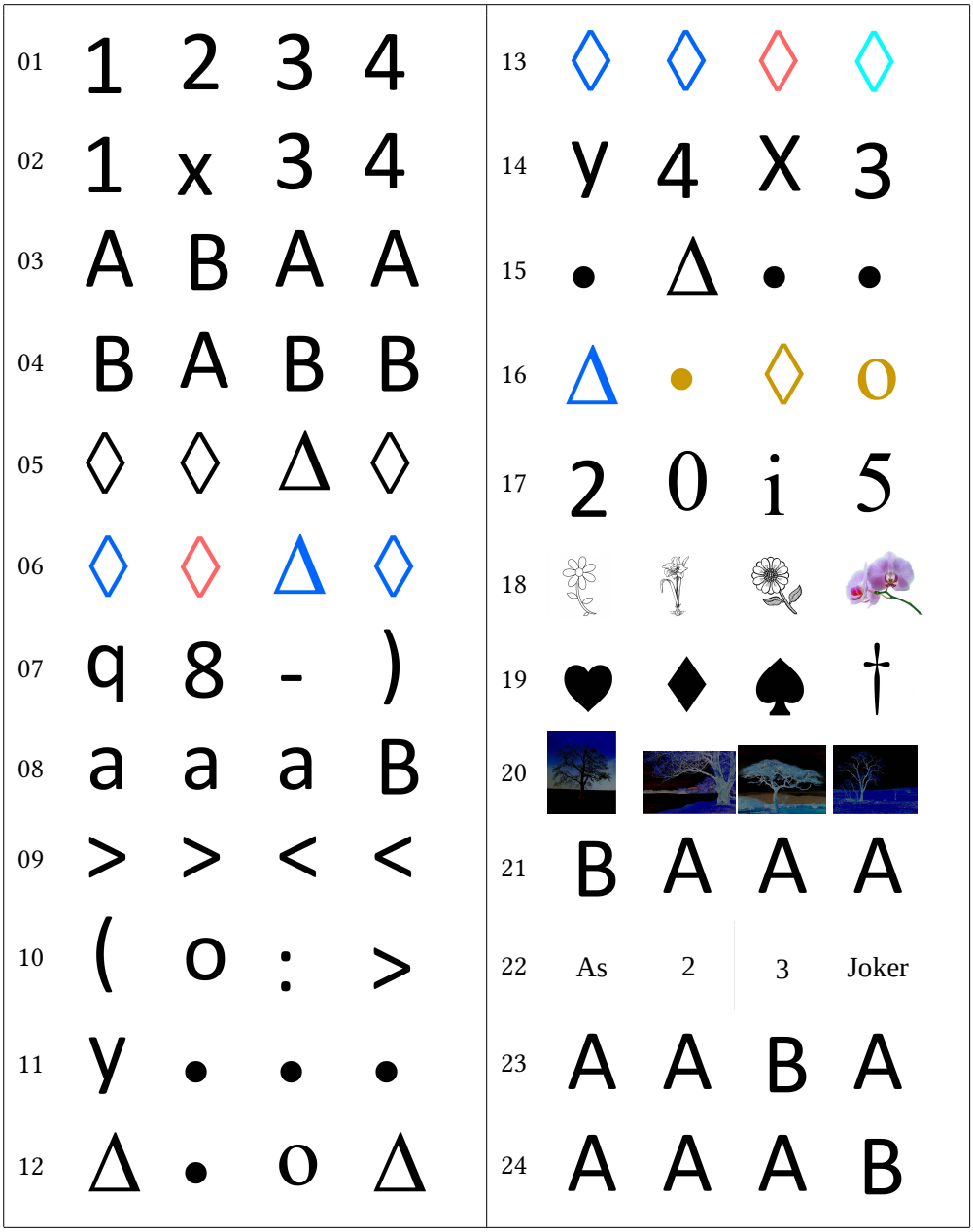


Figure A1: The 24 label sets, used to label the four options of the game. One set for each period.

Single Belief	Model $1''$
Consensus	-0.251 (2.136)
Consensus × Frame	7.330*** (2.389)
Hindsight Bias	-1.810 (2.042)
Hindsight Bias × Frame	0.510 (2.017)
Belief to the right	18.448*** (2.506)
Belief to the right × Frame	-5.433* (3.104)
Constant	20.588*** (0.919)
R^2	0.1445

Table A2: OLS dummy regressions of single belief elements, used to correct beliefs. Standard errors in parenthesis clustered on subject level (70 clusters). Asterisks: *** $p < 0.01$, * $p < 0.1$

B An Additional Experiment on *ex-post* rationalization

In an additional experiment, we eliminated the potential for *ex-post* rationalization in the opponent frame by asking participants about their beliefs (directly) *before* they make their choice in the discoordination games from Experiment 1 (both players obtain 7€ iff they choose different options).[26] Comparing the own-action probabilities from this treatment to the corresponding probabilities from Experiment 1 yields an estimate for the importance of *ex-post* rationalization. We can interpret the probability difference in this way because we already know from Experiment 2.A that both the consensus effect and wishful thinking do not seem to play a role under the opponent frame. As an additional benchmark, we also ran two sessions under the random-other frame. Under this frame, we expect there to be no difference between Experiment 1 and the additional experiment (as stated above, we see little scope for *ex-post* rationalization in the random-other frame). 86 subjects participated in the additional experiment.

Results The results in Figure B1 show that removing the potential for *ex-post* rationalization indeed changes the own-action probabilities in participants' reported beliefs: under the opponent frame—the frame under which we would expect *ex-post* rationalization—average own-action probabilities are roughly four percentage points (or 25%) higher when beliefs are elicited before actions compared to when they are elicited after the action (rank-sum test, $p = 0.028$). In contrast, under the random-other frame (where we argued *ex-post* rationalization should play no role) there is no difference ($p = 0.742$), which is in line with the results of Rubinstein & Salant

[26] *Ex-post* rationalization of a belief by an action would be unintuitive: we may well choose an action without forming a belief in the standard setup, but once we form a belief (as in the first stages of the additional experiment), there does not seem to be a good reason to form yet a different belief that we then contradict out of a taste for consistency.

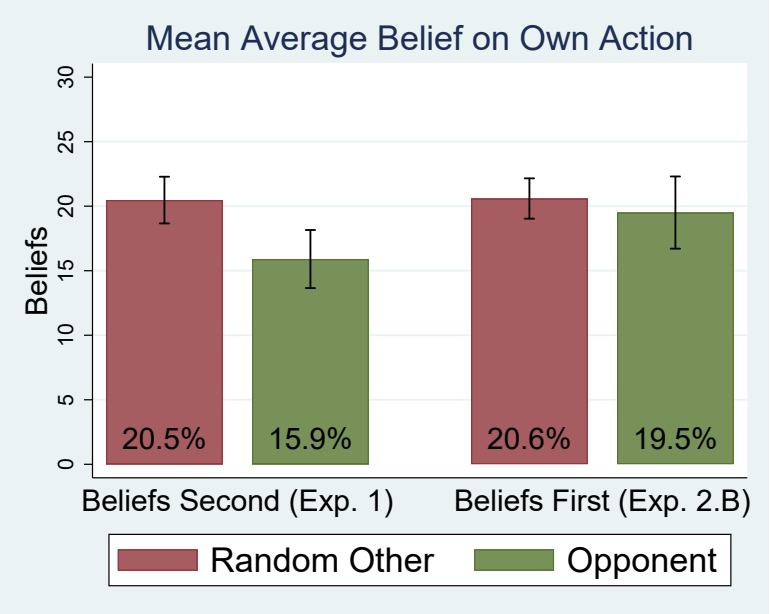


Figure B1: Beliefs in the Beliefs-First and the Beliefs-Second treatments. Error bars indicate 95% confidence intervals. For all tests, the data is aggregated on the individual level across all periods yielding one independent observation per participant.

(2016). We interpret the results as additional evidence for *ex-post* rationalization in the opponent frame.

C Experimental Instructions

The instructions are translated from german and show the opponent frame as example. Boxes indicate consecutive screens showed to participants. The instructions of the additional experiment in Appendix B had the same content, but were slightly more complicated due to the belief elicitation before the action.

Today's Experiment
Today's experiment consists of 24 situations in which you will make two decisions each.

Decision 1 and Decision 2

In the first situation, you will see the instructions for bot decisions directly before the decision. In later situations, you can display the instructions again if you need to.

The payment of the experiment
In every decision you can earn points. At the end of the experiment, 2 situations are randomly drawn and payed. In one of the situations, we pay the point you earned from decision 1 and in the other situation, you earn the points from decision 2. The total amount of points you earned will be converted to EURO with the following exchange rate:

1 Point = 1 Euro

After the experiment is completed, there will be a short questionnaire. For completion of the questionnaire, you additionally receive 7 Euro. You will receive your payment at the end of the experiment in cash and privacy. No other participant will know how much money you earned.

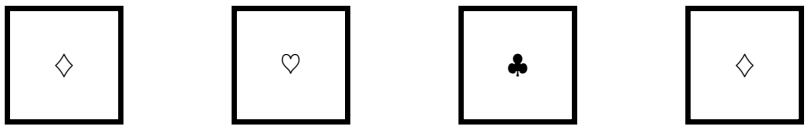
Instructions for decision 1
In today's experiment, you will interact with other participants. **You will be randomly rematched with a new participant of today's experiment in every situation.**
Decision 1 works in the following way: You and your matching partner see th exact same screen. On the screen, you can see an arrangement of four boxes which are marked with symbols. You and the other participant choose one of the boxes, without knowing the decision of the respective other. [One of] You can earn an price of X Euro.

Experiment 1 & 3
[You only receive the X euro only if you choose **another** box than your matching partner. If both of you choose the same box, bot do not receive points in this decision]

Experiment 2
[The relative position of your chosen boxes determines who wins the price. The participant wins, whose box lies to the immediate left of the other participant's box. If one participant chooses the most left box, then the other participant wins, if he chooses the most right box. If you don't win, you receive a price of 0 euro. It is of course possible, that neither you, nor the other participant wins.]

You will only learn at the end of the experiment, which box was chosen by the other participant and which payoff you receive in a certain situation.
The arrangement of symbols on the boxes is different in every situation. Below, you can see an example of how such an arrangement could look like.

Example: The four boxes are marked from left to right by Diamond, Heart, Spade, Diamond.



In this example, there are two boxes which are marked with the same symbol. However, the boxes on the most left and most right count as are different boxes.

Only Experiment 2

Instructions for decision 1
Although you choose a box in every situation, in some situations a box which was randomly chosen by the computer will be payoff relevant for you. This works in the following way:
After your decision, the computer draws one ball from the following urn in each situation:

If the blue ball that says "You" is drawn your own choice in decision 1 is relevant in this situation.
If the green ball that says "Computer" is drawn, the computer chooses one of the four boxes randomly (with equal probability of $\frac{1}{4}$) for you. This box will then be payoff relevant for you.
Your own decision is hence relevant with probability $\frac{1}{2}$ (=50%). The decision of the computer is relevant with probability $\frac{1}{2}$ (=50%).

The decision of your matching partner

To determine whether you won the price, we always use the original decision of your matching partner. This also holds if the computer decides for you or the other participant.
To determine whether you won the price, we hence always use the original choice of your matching partner and, depending on the drawn ball, your decision or the decision by the computer.

Text in squared brackets is frame dependent. We show the opponent frame as example.

Instructions for decision 2
In decision 2, your payoff also depends on your own decision and [on the decision of your matching partner. It will be the same matching parter, you already interacted with in decision 1.] We now explain decision 2 in detail.

Decision 2
Decision 2 refers always to a situation in which you already made decision 1. You will hence see the arrangement of boxes from the respective situation again. Again, the decision 1 [of your matching partner is relevant for you.] Decision 2 is about your assessment, [how your matching partner decided. We are interests in your assessment of the following question:]

[*See description of frames above*]

Only Experiment 2
[Please note that decision 2 is about the **actual** (human) decision of your matching partner and **not** about a possible computer decision.]

For every box, you can report your assessment [with what probability your matching partner chose the respective box]. You can enter the percentage numbers in a bar diagram. By clicking into the diagram, you can adjust the height of the bars. You can adjust as many times as you like, until you confirm. Since your assessments are percentage numbers, the bars have to add up to 100%. The sum of your assessment is displayed on the right. You can adjust this value to 100% by clicking. Or you enter the relative sizes of your assessments only roughly and then press the "scale" button. Please note, that because of rounding, the displayed sum ma deviate from 100% in some cases.
On the next page, we explain the payoff of decision 2.

Text in squared brackets is frame dependent. We show the opponent frame as example.

The payoff in decision 2
In this decision, you can either earn 0 or 7 points. Your chance of earning 7 points increases with the precision of your assessment. Your assessment is more precise, the more it is in line with [the decision behaviour of your matching partner. For example, if you reported a high assessment on the actually selected box, your chance increases. If your assessment on the selected box was low, your chance decreases.]
You may now look at a detailed explanation of the computation of your payment, which rewards the precision of your assessment.

It is important for you to know, that the chance of receiving a high payoff is maximal in expectation, if you assess the behaviour of your matching partner correctly. It is our intention, that you have an incentive to think carefully about the behaviour of your matching partner. We want, that you are rewarded if you have assessed the behaviour well and made a respective report.

Your chance will be computed by the computer-program and displayed to you later. At the end of the experiment, one participant of today's experiment will roll a number between 1 and 100 with dies. If the rolled number is smaller or equal to your chance, you receive 7 points. If the number is larger than your chance, you receive 0 points.

Text in squared brackets is frame dependent. We show the opponent frame as example.

Payment of the assessments
At the end of your assessment, you will receive the 7 points with a certain chance (p) and with $(1-p)$, you receive 3 points. You can influence your chance p with your assessment in the following way:

As described above, you will report an assessment for each box, on how likely [your matching partner is to select that box. One of boxes is the actually selected. At the end, your assessments are compared to the actual decision of your matching partner.] Your deviation is computed in percent.

Your chance p is initially set to 1 (hence 100%). However, there will be deductions, if your assessments are wrong. The deductions in percent are first squared and then divided by two.

For example, if you place 50% on a specific box, but [your matching partner selects another box,] your deviation is equal to 50%. Hence, we deduct $0.50 * 0.50 * \frac{1}{2} = 0.125$ (12.5%) from p.

[For the box, which is actually selected by your matching partner, it is bad if your assessment is far away from 100%. Again, your deviation from that is squared, halved and deducted. For example if you only place 60% probability on the actually selected box, we will deduct $0.40 * 0.40 * \frac{1}{2} = 0.08$ (8%) from p.]

With this procedure, we compute your deviations and deductions for all boxes.
At the end, all deductions are summed up and the smaller the sum of squared deviations is, the better was your assessment. For those who are interested, we show the mathematical formula according to which we compute the quality of your assessment and hence your chance p of receiving 7 points.

$$p = 1 - \frac{1}{2}\left[\sum_i (q_{box_i,estimate} - q_{box_i,true})^2\right]$$

The value of p of your assessment will be computed and displayed to you at the end of the experiment. The higher p is, the better your assessment was and the higher your chance to receive 7 points (instead of 0) in this part. At the end of the experiment, the computer will roll a random number between 0 and 100 with dies. If this number is smaller or equal to p, you receive 7 points. If the number is larger than p you receive 0 points.

Summary
In order to have a high chance to receive the large payment, it is your aim to achieve as few deductions from p as possible. This works best, if you have an good assessment of the behaviour of participant B and report that assessment truthfully.