Bauer, Dominik; Wolff, Irenaeus

**Conference Paper**

# Biases in Beliefs

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2019: 30 Jahre Mauerfall - Demokratie und Marktwirtschaft - Session: Experimental Economics VI, No. E06-V1

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

# Biases in Beliefs: Experimental Evidence[§]

Dominik Bauer Irenaeus Wolff

*Graduate School of Decision Sciences & Thurgau Institute of Economics*

*University of Konstanz, Germany*

dominik.3.bauer@uni.kn wolff@twi-kreuzlingen.ch

This version: 20th February, 2019

**Abstract:** Many papers have reported behavioral biases in belief formation that come on top of standard game-theoretic reasoning. We show that the processes involved depend on the way participants reason about their beliefs. When they think about what everybody else or another 'unspecified' individual is doing, they exhibit a consensus bias (believing that others are similar to themselves). In contrast, when they think about what their situation-specific counterpart is doing, they show *ex-post* rationalization, under which the reported belief is fitted to the action and not vice versa. Our findings suggest that there may not be an 'innocent' belief-elicitation method that yields unbiased beliefs. However, if we 'debias' the reported beliefs using our estimates of the different effects, we find no more treatment effect of how we ask for the belief. The 'debiasing' exercise shows that not accounting for the biases will typically bias estimates of game-theoretic thinking upwards.

*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 different 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 often is extremely helpful to test economic models. The list of examples for this approach is long. Game theorists have tested whether non-equilibrium beliefs can explain non-equilibrium behavior (*e.g.*, Bellemare et al., 2008; Costa-Gomes & Weizsäcker, 2008; Rey-Biel, 2009). Macroeconomists have explained saving and investment decisions by people's beliefs about future income, demand,

[§]We would like to thank Ariel Rubinstein, Yuval Salant, Robin Cubitt, Marie Claire Villeval, Dirk Sliwka, and Roberto Weber, as well as participants at the ESA 2016 European Meeting in Bergen, the ESA 2018 World Meeting in Berlin, a seminar audience at the University of Nottingham, the research group at the Thurgau Institute of Economics and the members of the Graduate School of Decision Sciences of the University of Konstanz for their helpful comments. Contact: Chair of Applied Research in Economics, University of Konstanz, Universitätsstraße 10, D-78464 Konstanz, Germany.

and inflation (*e.g.*, Guiso & Parigi, 1999; Engelberg et al., 2011). Further examples are development economists studying the adoption of new agricultural technologies (*e.g.*, Delavande *et al.*, 2011b) and health economists studying the reasons for why people engage in activities that put their own health at risk (such as smoking, *e.g.*, Khwaja *et al.*, 2006).[1]

Given belief elicitation has such a broad field of applications, it is crucial that we know how people come up with their beliefs under different circumstances. This is important for interpreting belief-elicitation data from experiments, questionnaires, and surveys, and ultimately for understanding behavior. Therefore, we need to know what biases we bring about by our commonly used elicitation methods. If we trigger specific processes by our elicitation method, we are likely to misinterpret the data systematically. In this paper, we show that the specific way of asking experimental participants for their beliefs triggers different biases. Moreover, by prompting our participants to think about their beliefs in different ways and by exposing them to a specific decision environment, we are able to identify the involved psychological processes and hence, the determinants of beliefs.

When studying beliefs, we face a typical conundrum: what theory assumes to be 'the belief' is an unobserved construct in people's heads. If we want to learn anything about it, we have to make beliefs observable. The classical approach in economics to this problem is to observe choices only and recover the unobservables afterwards, for example by invoking the revealed-preference assumption (Samuelson, 1938). However, reconstructing beliefs from choices can sometimes be very difficult. For example, in numerous games, the same choice can be rationalized by many different beliefs (*e.g.* Manski 2002). For these reasons, an alternative and popular way of making beliefs observable is simply to ask for them in a belief-elicitation procedure.

Now assume we find some systematic bias in the elicited beliefs. What is the origin of the bias? Was the construct in people's heads biased? Or was it the very act of asking, that squeezed 'the belief' into numbers, thereby biasing (only) the report we observe? In our opinion, these questions can be answered only very partially. Trying to tease them apart is beyond the scope of this paper. Having said that, we do present experimental data that shows that 'more than only a report' is biased: we can induce differences in game behavior in the same way as we induce differences in belief reports. However, for the reasons outlined above, in the remainder of the paper we will no longer distinguish whether people's true beliefs (which we do not observe) or 'only' the belief reports are biased when we talk about biased beliefs.

To analyze the interaction between the method we use and the involved psychological processes, we look at three different ways of asking for beliefs which we call 'frames'. The opponent frame asks for beliefs about the participant's matching partner. The random-other frame asks about some

[1]For these and further examples, see, *e.g.*, Trautmann & van de Kuilen (2015).

other individual who is not the matching partner and the population frame asks for beliefs about the whole population of players. The three ways of asking might trigger different ways of thinking about the belief.

Along with the frames, we consider five different processes as potential determinants of beliefs. *Ex-post* rationalization, wishful thinking, hindsight bias, and consensus bias (projection) are four biases that potentially shape beliefs on top of standard game-theoretic reasoning. In standard game-theoretic models, players first form a belief and then act upon this belief. Under *ex-post* rationalization, this process is reversed: agents first make a choice (by whatever process), and then, they form a belief that justifies their taken action. Wishful thinking has people have more faith in events—including actions taken by others—that would lead to a favorable outcome. Under a hindsight bias, people fail to abstract from their knowledge about the realization of an uncertain event (*e.g.*, their own action, viewed from their opponent's perspective) when evaluating an action that was taken without this knowledge (in this case, their opponent's action). And under a consensus bias, people project onto others what they themselves would do, feel, or think.

The five processes just described will point in the same direction under many circumstances, and in different directions, under others. In this paper, we systematically vary the decision environment as well as the frame of belief elicitation to disentangle the processes and to test which of them play a role under what circumstances.

Our paper has two main parts. In each part, we present data from two experiments. The first part establishes the influence of our framing manipulations on belief reports and on behavior, while the second part disentangles the different processes to explain why the observed framing differences occur. In Experiment 1.A, we show that elicited beliefs display considerable framing differences that also influence observed belief-action consistency. In particular, we replicate Rubinstein & Salant's (2015) finding that beliefs are closer to participants' own actions under a population frame than under an opponent frame. We conduct an additional treatment, eliciting beliefs under the random-other frame. The results from the random-other treatment shed light on the framing difference between population and opponent frame. The framing difference is caused by the difference 'interaction partner *vs.* another person' and not by whether the question is about 'one person *vs.* several people'. Experiment 1.B provides evidence of participants behaving differently in the same game, depending on whether the game is presented in an 'opponent frame' or in a 'population frame'. This suggests that the frames affect also the underlying beliefs, not 'just' the *ex-post* belief reports.

In Experiment 2.A, we disentangle the different processes to explain why the observed framing differences occur. We separate consensus bias, hindsight bias, wishful thinking, and game-theoretic reasoning or *ex-post* rationalization. Experiment 2.A provides evidence for a consensus effect in the

random-other frame. Game-theoretic reasoning and *ex-post* rationalization cannot be distinguished fully, but we find some suggestive evidence for *ex-post* rationalization in the opponent frame.

Experiment 2.B corroborates the suggestive evidence from Experiment 2.A for *ex-post* rationalization in the opponent frame. All of our belief-elicitation experiments provide evidence of game-theoretic reasoning, but none of them provides evidence for a hindsight bias or wishful thinking. Using our estimates on the biases, we can 'recover' participants' hypothetic unbiased beliefs, and provide an estimate for the best-response rate to those 'underlying' beliefs. This exercise suggests that the amount of game-theoretic reasoning typically will be overestimated in many papers in the literature.

In summary, the three belief-elicitation experiments provide evidence that different ways of asking for beliefs trigger different specific processes. The population and the random other frame influence beliefs by a consensus bias, and under the opponent frame, participants tend to *ex-post* rationalize their actions via beliefs.

Among other things, this result is important for our understanding of the literature. For example, the consensus bias seems to be closely related to the belief-elicitation method employed by the researchers: it seems that all major studies in economics documenting a consensus bias use the population frame.[2] On the other hand, the opponent frame seems like a natural choice in studies that investigate belief-action consistency and best-response rates.[3] Our systematic investigation shows that the correspondence between population frame and consensus bias is more than a mere coincidence.

A consensus bias can be documented only when it stands in contrast to the predictions of standard theory (*i.e.*, when a player wants to choose a different option than her opponent). In such situations, the consensus and *ex-post* rationalization point in opposite directions. However, by the results of this study, a consensus bias is present only under a population or random-other frame. Now consider, as a thought experiment, that the authors documenting the consensus bias had used the opponent frame to elicit beliefs. Not only had they not measured a belief biased towards participants' own actions (because of a consensus bias), they would have measured beliefs biased *away* from participants' actions (because of *ex-post* rationalization under the opponent frame). In other words, had the authors used the opponent frame, they most likely would not have seen a consensus bias but an extraordinarily high proportion of consistent belief-action pairs. This does not mean

[2] Selten & Ockenfels (1998), Charness & Grosskopf (2001), Van Der Heijden, Nelissen & Potters (2007), Ellingsen *et al.* (2010, who also use the random-other frame), Engelmann & Strobel (2012), Iriberri & Rey-Biel (2013), Blanco et al. (2014), Danz, Madarász & Wang (2014), Molnár & Heintz (2016), Rubinstein & Salant (2016), Proto & Sgroi (2017).

[3] 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).

the consensus bias is not a real phenomenon. However, it may not be as general a phenomenon as the widespread references to it in the literature may suggest.

By the results of our experiment, we recommend to take the substantial framing differences into account in the analysis of existing data or the design of new surveys and experiments. In particular, in designing new experiments, we propose to elicit beliefs before (or potentially at the same time as) the corresponding action in the opponent frame. This way, beliefs will not be affected by any of the biases discussed in this paper. At the same time, the typical concern that this order will induce much more game-theoretic behavior seems unwarranted: also under this setup, the measured best-response rate is only 57%, which is well within the range of 50%-72% under the biased estimates from the biased *ex-post*-elicitation treatments. Given that the 50% result from a bias that leads to a lower observed best-response rate, and the 72% result from a bias that leads to an over-estimation of the best-response rate, the 57% are a plausible measurement for the true best-response rate.[4]

# 2 Related Literature

### Methods for belief elicitation

The literature has proposed numerous variants for incentives, procedures and mechanisms of belief elicitation (see Schotter & Trevino, 2014, or Schlag *et al.*, 2015, for recent reviews). The large variety of methods and applications also brings about high variation in explanatory power of beliefs for behavior within and across experimental studies.[5] Most of the literature on belief-elicitation methods concentrates on designing different incentive schemes (that is, payoff-rules) and evaluating their performance with respect to belief-action consistency or properness.[6] Additional topics are hedging (Blanco *et al.*, 2010) or the usefulness of second order beliefs (Manski & Neri, 2013). However, systematic investigations of elicitation procedures that are not related to the incentives and their influences on properness and belief-action consistency are rare. There are two noteworthy exceptions.

Costa-Gomes & Weizsäcker (2008) study belief-action consistency in different generic 3x3 normal-form games. They vary the timing and ordering of belief and action tasks and find no substantial

[4] Note that in the literature, a prominent quality criterion for belief-elicitation procedures has been whether beliefs match the empirical distribution (see, *e.g.,* the list provided in Schlag *et al.*, 2015). However, if people's choices are correlated—which they are—this criterion will favor elicitation procedures that induce a consensus bias. We therefore add (yet) another cautionary remark on the use of this criterion (see Schlag *et al.*, 2015, for a similar argument).

[5] *E.g.*, in some of the 3×3 games in Costa-Gomez & Weizsäcker (2008), best-response rates are around 51%. On the other end, Manski & Neri (2013) find a best-response rate of 89% in a 2-action Hide&Seek game.

[6] *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.

treatment differences. Belief-action consistency is generally low in their study, at around 50%.[7] In a field study on fishermen in India, Delavande *et al.* (2011a) vary procedural details like the precision with which probabilities can be expressed or how the support of the belief distribution is determined. The authors find that their elicitation results are robust with respect to the methodology.
In the literature, different belief-elicitation treatments usually perform the role of a robustness check. Some papers also pursue a methodologic question, searching for a treatment that truthfully elicits participants' beliefs. Our approach is somewhat different. In this paper, we use different belief-elicitation frames as a treatment to induce different mental representations. These treatments will enable us to learn something about the underlying belief-formation process. Having said that, the results will inevitably speak to methodologic questions as well.

**Framing of belief elicitation**

Virtually all studies in the literature use the population or the opponent frame, but the specific choice is rarely motivated. As already mentioned, all major studies in economics documenting a consensus effect use the population frame while the opponent frame is the common choice in studies that investigate belief-action consistency and best-response rates.[8] This again underlines the relevance of studying whether observing a consensus bias or particular consistency levels are specific to the belief-elicitation format.
In a completely different context, Critcher and Dunning (2013 and 2014) use the population and the "individual" frame to elicit behavioral forecasts. The individual frame is similar to the opponent and the random other frame in that they ask for the belief about "a randomly selected student… [whose] initials are LB". They find framing differences in judgments of morally relevant behaviors. However, they only elicit beliefs and report a lack of evidence for a consensus effect.

# 3 Determinants of beliefs

We propose that the specific way of asking for beliefs will trigger different processes that shape the belief reports. Hence, the different ways of asking will lead lead to systematic variation in reported beliefs. In this section, we will first outline the three different ways of asking for beliefs which will also form one of our treatment dimensions in the experiment. Afterwards, we will describe the processes in detail and predict under which of the ways of asking they should be active.

[7] Rubinstein & Salant (2016) also report a "beliefs-first" treatment. Their main effects show up also in this treatment, but less strongly.
[8] See footnotes 3 and 4.

| |
|---|
| **Opponent frame** |
| Object: Single person, the matching partner |
| *"With what probability did your matching partner chose each of the respective boxes of the current set-up?"* |
| **Random-other frame** |
| Object: Single person, not the matching partner |
| *"With what probability did a person who is not your matching partner chose each of the respective boxes of the current set-up?"* |
| **Population frame** |
| Object: All persons in the session, including 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?"* |

**Table 1:** The three different frames of the belief-elicitation question.

## 3.1 Elicitation frames

The three different questions we use for asking for beliefs are spelled out in Table 1. Note that from a standard theory perspective, all three questions are equivalent under a random partner matching protocol.

We will call our different ways of asking for beliefs the elicitation "frames". However, the different ways of asking are more than just frames. It is easy to frame a payoff of 10€ as a gain (*"You receive 10 €"*) or a loss (*"You have 20€, now you loose 10€"*). What we call a "frame" is more than just describing an equivalent outcome in two different ways. Rather, our frames focus on different identities and numbers of people. 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 frame prompts people to think about their specific interaction partner, although this player is only the one they are randomly matched to out of many other players. In this frame, it seems much more natural to think about the individual incentives of both players and about the strategic interaction they face. In contrast to that, the random-other frame also focuses on an individual person, but since there is no interaction between the players, the strategic aspect is much weaker. The population frame invokes a picture of many other people, most of whom a participant will not interact with. Individual incentives and the strategic aspects may not play as important a role when thinking about the problem on such a "gobal" scale. Rather it seems important what will influence the behavior of the population as a whole.

| | Population | Random Other | Opponent |
|---|---|---|---|
| Game-Theoretic Reasoning | ✓ | ✓ | ✓ |
| *Ex-Post* Rationalization (Cognitive Dissonance) | (✓) | - | ✓ |
| Wishful Thinking | (✓) | - | ✓ |
| Consensus Bias | ✓ | ✓ | ✓ |
| Hindsight Bias | - | - | ✓ |

**Table 2:** Predictions of which processes are active under which frame.

## 3.2 Processes that shape beliefs

We now describe the different processes and identify under which frame they would be active, if at all. All predictions are summarized in Table 2.

### Game-Theoretic Reasoning

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 ('game-theoretic behavior'), or first choosing an action by some process and only then making up a belief consistent with the action. This reversed process (action-then-belief) may either 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 of these biases in the following.
We expect game-theoretic reasoning to be present under all of the frames, as we are not aware of any study documenting that participants' actions are not positively related to their beliefs.

### *Ex-Post* Rationalization

Agents may want to appear consistent both for external reasons (because they do not want to look like a fool in the eyes of the experimenter) and internal reasons. A prime example of an internal reason is cognitive dissonance (Festinger 1957). In the remainder of this paper, we use "*ex-post* rationalization" as a shorthand for "*ex-post* rationalization due to cognitive dissonance". Under cognitive dissonance, making two inconsistent choices—an action and a belief—would lead to mental unease in the player's mind. In order to avoid such mental unease, the player would adapt her belief to make it fit her taken action.

In light of the above, *ex-post* rationalization should be strongest in the opponent frame: believing that *some other* player chose an option that would be bad for me need not cause cognitive dissonance, because my opponent still might have chosen something else. In contrast, if my *opponent* chose something that would be bad for me given my action, this should indeed cause cognitive dissonance in me. The population frame should be somewhere in between these two cases: the more concentrated my belief in the population frame, the more cognitive dissonance I should experience because the more the population will be representative of my opponent (we do not test this final hypothesis in this paper, though).

### 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. For example, people believe that things they like are more likely to happen (*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 stronger the more the matching partner is involved, because the desirable outcome depends on this specific person. Hence, wishful thinking should be most prevalent in the opponent frame, followed by the population frame, and it should be absent in the random-other frame.

### Consensus Bias

The *consensus* bias is a prominent phenomenon, intensively studied by psychologists and economists. Tversky & Kahneman (1973, 1974) link it to the *availability heuristic* and the *anchoring-and-adjustment heuristic*. Joachim Krueger gives a very general but engagingly simple description of what the consensus effect means: *"People by and large expect that others are similar to them"* (Krueger, 2007, p. 1). The basic idea of a consensus bias has been studied in many different contexts and it has been given

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 bias broadly, 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.

We hypothesize that a consensus bias can occur in all elicitation frames. The expectation, that others are similar to myself, should not depend on whether my matching partner is involved or not. Further, it should not depend on whether thinking about one or many persons.

**Hindsight Bias**

Under a hindsight bias (Fischhoff, 1975), agents strongly overestimate the probability of an event after 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' behavior. Players with a hindsight bias hence form their belief (as if they were) assuming the other players should have anticipated that the biased player would choose with a very high probability whatever she ended up choosing in actual fact. Therefore, a hindsight bias increases the probability mass placed on the other player(s) playing a best-response to the player's own action.

We expect that a hindsight bias will exclusively occur in the opponent frame, because the hindsight relates to the event that my *matching partner* chooses a best response to my own action. In the random-other frame, 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 behavior. 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.

# 4 Experimental Design

### General setup

This paper presents the data from four experiments. We next describe the general setup which three of the experiments have in common as well as the specific purposes of all four experiments. Experiment 1.A serves three purposes. First, it replicates the earlier finding that beliefs are closer to participants' own actions under a population frame than under an opponent frame. Second, it showcases the substantial differences the elicitation-frame choice has for interpretations regarding participants' belief-action consistency. And third, it singles out the 'interaction partner *vs.* another person' distinction as the crucial difference between the frames. Experiment 1.B shows that the frames are able to change also behavior (as opposed to 'only' belief reports). Experiments 2.A and 2.B disentangle the mental processes underlying the findings from Experiment 1. 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 game-theoretic reasoning and *ex-post* rationalization. Experiment 2.B separates ('*ex-ante*') game-theoretic reasoning from *ex-post* rationalization.

In particular for Experiments 1.A and 2.A, it is crucial to control participants' preferences because we want to interpret belief-action consistency. Abstracting from stochastic choice and stochastic belief reports (see, *e.g.*, Bauer & Wolff, 2017), belief-action inconsistencies can happen for two reasons: (i) the researcher may have mis-specified the participants' utility function, and (ii) participants may have a bias in their belief reporting. This paper focuses on participants' biases in the reporting of beliefs. In contrast to some of the earlier literature, we choose games in which the predictions do not change for any of the well-documented deviations from risk-neutral payoff maximization. We thereby rule out mis-specification of participants' utility function as a reason for belief-action inconsistency. In particular, non-neutral risk and loss attitudes and social preferences do not matter for the predictions in the games we chose.[9]

In Experiments 1.A and 2.B, 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 on different sets of labels such as "1-2-3-4",

[9]More precisely, social preferences do not matter in Experiments 1 and 3 unless participants have so spiteful preferences that they prefer both participants receiving nothing to both receiving the same positive amount of money. This case should happen so rarely that we abstract from it. In Experiment 2, social preferences do not matter as long as people are not ready to burn own money for the sake of equality (a condition that already Fehr & Schmidt, 1999, impose).

"1-x-3-4", or "a-a-a-B", with randomly changing partners, and without any feedback in between.[10] 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.[11]

Along with every choice in the game, we elicit probabilistic beliefs in every period, incentivizing the belief reports via a Binarized-Scoring Rule (McKelvey & Page, 1990, Hossain & Okui, 2013). In the belief-elicitation task, subjects could earn another 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 3.1. At the end of Experiments 1.A, 2.A, and 2.B, 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.

Experiment 1.B was part of an experimental series of one of the authors (I.W.) and appended to another experiment. At the end of the session, one participant would be randomly selected to get the payoff from this "extra part" of the session. In Experiment 1.B, participants faced a very particular variant of an n-player, three-option, one-shot discoordination game. In particular, participants had the choice between three monetary amounts, 27€, 30€, and 33€. For every other participant who chose a different amount, the randomly selected participant would obtain her chosen amount divided by the number of participants in the respective session (in one case, 24 in one case, 30, and otherwise 26 or 28 participants). This was the way the game was presented to participants in the 'population frame' treatment. In the 'opponent frame' treatment, the game was first presented as a two-player game ("you will receive the amount you stated, but only if the other participant states another amount"). We then announced that they would be playing the game subsequently against all other participants in their session, but that they would be allowed to choose only a single amount for all of the interactions. This single amount would be relevant for each of the interactions, and the randomly selected participant would receive the average payment from all

[10] 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.

[11] 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.

interactions. We thus presented the same game in two different ways. In the 'population frame' treatment, we made them think about the population, whereas in the 'opponent frame' treatment, we focused their attention on a single opponent before pointing out that they would be playing against several individual opponents at the same time (and using the same strategy).

**Procedures**

We programmed the experiments using z-Tree (Fischbacher, 2007) and conducted them in the Lake-Lab at the University of Konstanz. We recruited 301 participants for Experiments 1.A, 2.A, and 2.B, and 214 participants for Experiment 1.B using ORSEE (Greiner, 2015). All sessions lasted between 60 and 90 minutes.

# 5 Framing effects on belief reports, behavior, and the implications for belief-action consistency

## 5.1 Experiment 1.A: Framing effects and belief-action consistency

Rubinstein & Salant (2015) find in a chicken-game experiment that beliefs are closer to participants' own actions under a population frame than under an opponent frame. In Experiment 1.A, we replicate the effect for a pure discoordination game. Note that changing the population frame to an opponent frame changes three things at a time. The first change is that the opponent frame asks for our belief about the person we are currently interacting with, while the population frame is mostly or even fully about "irrelevant" others (interaction partner *vs.* another person). The second change is that the opponent frame is about one person, while the population frame is about several people. Hence, the target is a different statistical object. And finally, because the targets are different, the absolute level of incentives is different.[12]

Following Rubinstein & Salant's (2015) argument and our own intuition, we conjectured that the relevant difference was the difference "interaction partner *vs.* another person". To test this conjecture, we included a third belief-elicitation frame, the random-other frame. This frame asks about the choice probabilities of a randomly drawn 'non-interaction-partner'. This is a *ceteris-paribus*

[12]To see this, think about the case that a participant knows the distribution of others' choices exactly. Then by design, it is optimal for the participant to report the true probabilities under either frame. However, this means that she will obtain the 'belief prize' with certainty under the population frame because the reported distribution is compared to the true distribution. Under the opponent frame, she will obtain the prize with a much lower probability, because her report is compared to *one realization* instead of being compared to the full distribution. In our design, the probability of receiving the prize when (optimally!) reporting the true choice distribution can be as low as 62.5% (under a uniform choice distribution).

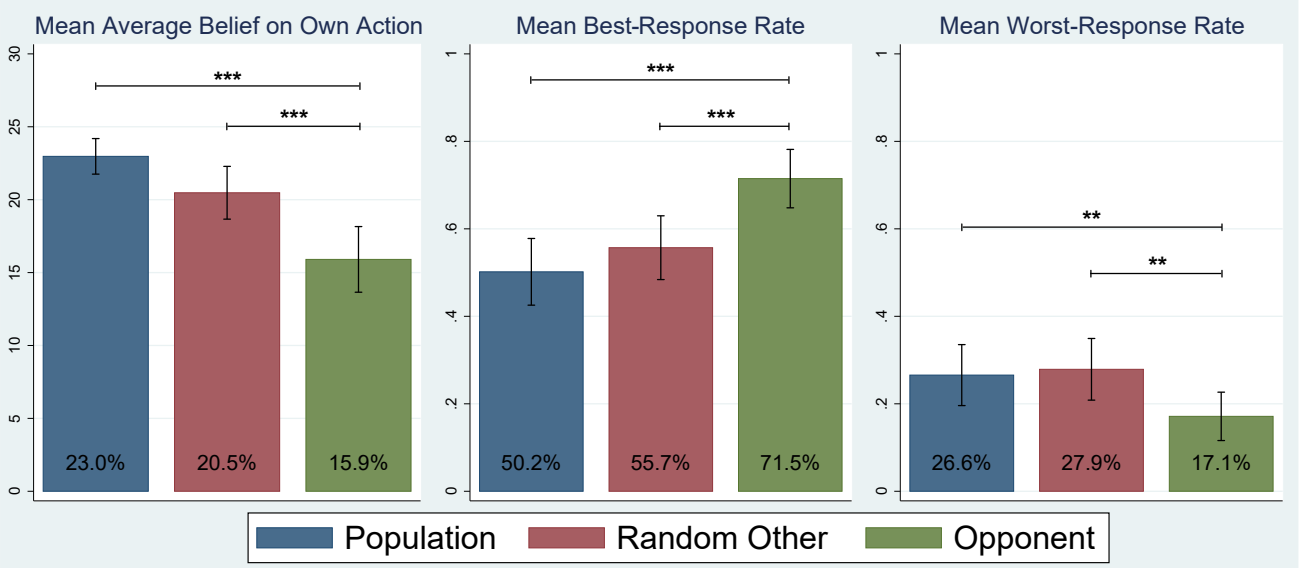



**Figure 1**: Beliefs and consistency in Experiment 1.A. Error bars indicate 95% confidence intervals. Rank-Sum tests: *** $p < 0.01$, ** $p < 0.05$. For all tests, the data is aggregated on the individual level across all periods, yielding one independent observation per participant.

comparison, as both the level of incentives and the number of observations in the target remain unchanged. We analyze the data of 145 participants from Experiment 1.A.[13] We elicited beliefs directly after each action.

### Results of Experiment 1.A

Figure 1 summarizes beliefs and belief-action consistency for the three frames in the discoordination game. For the analysis, we aggregate the data on the individual level across all periods. For each participant, we look at the probability mass in the reported belief on the participant's own action in the corresponding game, averaged across all 24 periods. This is the average subjective probability that the participant did *not* discoordinate. This procedure yields one independent observation per participant. Similarly, we compute the best- and 'worst-response' rate to beliefs for each participant individually. 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.

The mean average belief on the participant's own action (Figure 1, left panel) is significantly higher in the population frame and the random-other frame compared to the opponent frame (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 frame, the average observed best-response rate is lower (mid panel, $p < 0.001$ and $p = 0.004$) and the average worst-response

[13] We exclude one participant from Experiment 1.A who always reported a 100% belief of not having discoordinated. This participant probably tried to hedge, but did not understand that hedging was impossible.

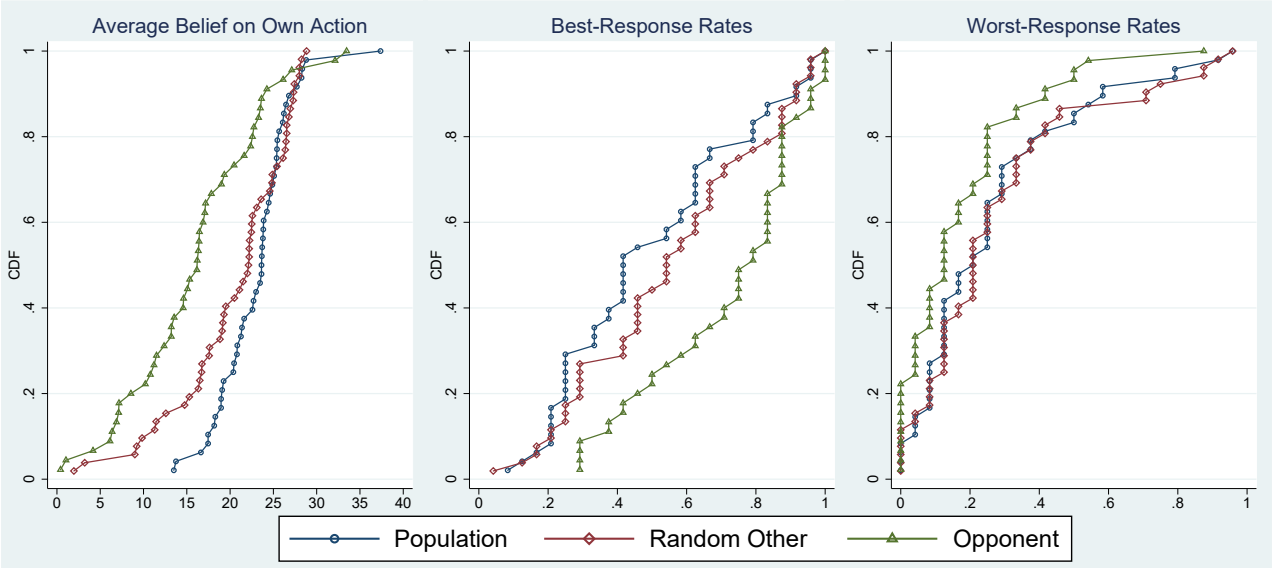



**Figure 2**: Cumulative distributions of individual belief and consistency data in Experiment 1.A across frames.

rate is higher (right panel, $p = 0.026$ and $p = 0.019$) in the population frame and the random-other frame.[14] The reduction in the best-response rate of more than 20 percentage-points and a 9.5 percentage-point increase in the worst-response rate in the population frame are considerable effect sizes. Note that the worst-response rate differs by more than 50% of the rate in the opponent frame.

For a more detailed picture of the results, we depict cumulative distribution functions of the same data in Figure 2. Own-action probabilities in the population frame second-order stochastically dominate those in the opponent frame and the distributions differ significantly according to a Kolmogorov-Smirnov test ($p < 0.001$). This effect again carries over to consistency: The best-response rate distribution in the opponent frame first-order stochastically dominates the respective distribution of the population frame and the distributions differ significantly ($p = 0.001$). Similar results hold when comparing the distributions of the opponent and the random-other frame (beliefs: $p = 0.002$, best-response rates: $p = 0.008$).[15]

## 5.2 Experiment 1.B: Framing effects on game behavior

The two framings we used for the '27-30-33' game yield markedly different patterns of behavior, as shown in Figure 3. In the Opponent frame, far more participants choose the high monetary amount (33€), and the distributions differ significantly by a $\chi^2$-test ($p = 0.019$). As for beliefs, the

[14]There is no significant difference between population and random-other frame. Rank-sum tests, beliefs: $p = 0.146$, best-response rates: $p = 0.237$, worst-response rates: $p = 0.822$.

[15]The distributions of the population and random-other frames do not differ significantly. Kolmogorov-Smirnov tests, beliefs: $p = 0.174$, best-response rates: $p = 0.305$. There is also no significant difference between the distributions of worst-response rates across frames (all $p > 0.122$).

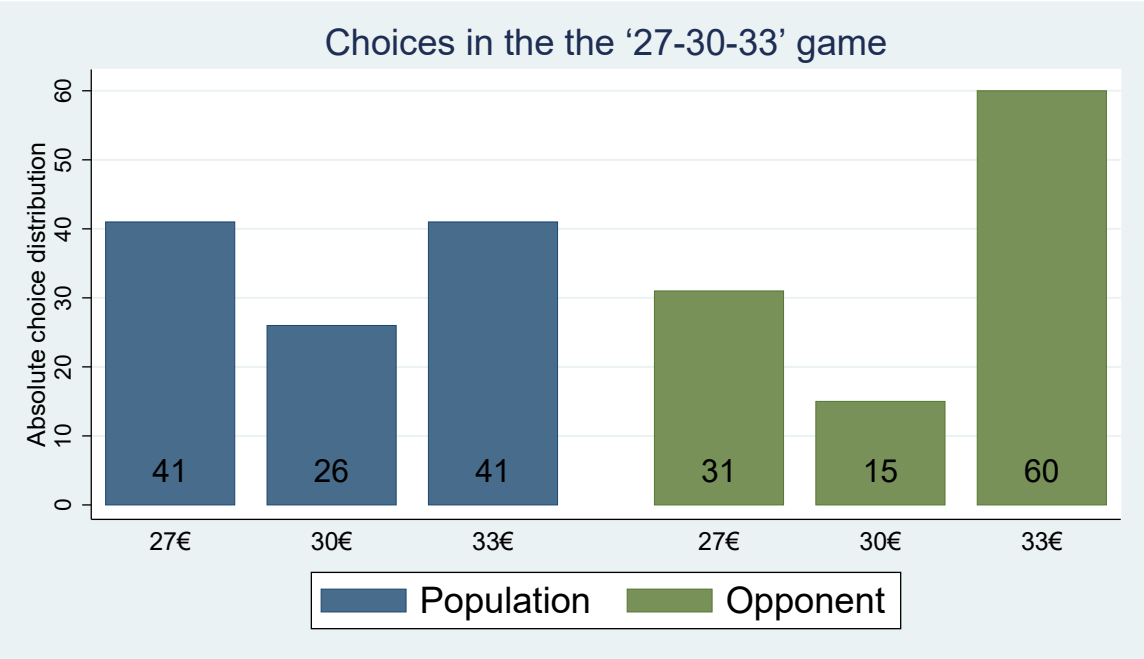



**Figure 3**: Data from Experiment 1.B.

different framing of an otherwise equivalent task makes a considerable difference for participants' choices in the game. Assuming that observed choices follow from participants' true beliefs, this result provides evidence that the frames also affect the underlying true beliefs and not 'just' the *ex-post* belief reports (which possibly could be the case in Experiment 1.A). Having said that, we will no longer distinguish between belief reports and true beliefs in the remainder of the paper, for the reasons outlined in the introduction.

## 5.3 Summary of Part 1

Up to this point, we have documented a considerable framing effect in equivalent tasks, both on the beliefs and on the behavioral level. Experiment 1.A shows the effect in belief elicitation. Although the questions in all frames are theoretically equivalent (up to the absolute level of incentives), reported beliefs differ substantially across frames. Most notably, beliefs differ in the *ceteris-paribus* comparison between the opponent and the random-other frame, where we vary only the identity of the target participant. 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.A does not show is whether the differences between the frames occur because there is (more) consensus under the population and random-other frames, or because there is (more) hindsight bias, wishful thinking, game-theoretic reasoning, or *ex-post* rationalization under the opponent frame. To disentangle these processes, we need Experiments 2.A and 2.B.

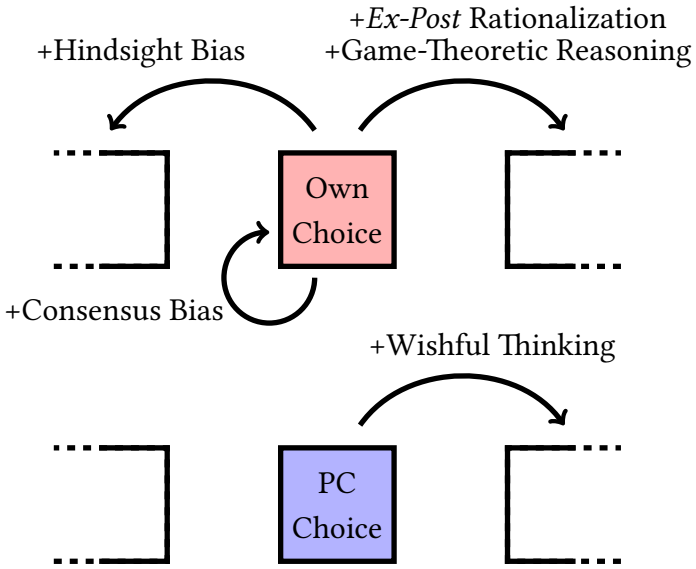



**Figure 4**: Predictions of the candidate processes in the to-your-left game with implementation errors in the case of an implementation error. We color example choices and indicate by arrows the predictions of the four candidate effects that depend on the relative position of the choices.

# 6 Disentangling the Processes

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

Experiments 2.A and 2.B are designed to explain why the framing effects documented in Experiment 1.A occur. Experiment 2.A disentangles the influences of a consensus bias, hindsight bias, and wishful thinking from game-theoretic reasoning/*ex-post* rationalization, which is not possible in the standard discoordination game. 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 4 visualizes the predictions of our candidate processes in Experiment 2.A. Because the game is circular, only the relative position of the respective box 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. Game-theoretic reasoning, *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 separate wishful thinking from game-theoretic reasoning and *ex-post* rationalization, we introduce random implementation errors. In every period, after the participant chooses one of the boxes, there is a 50% probability that the computer changes the participant's decision. If the computer alters the decision, the computer chooses each box with equal probability (including the participant's chosen box). 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 following each action always targets the other participants' original choices, not the implemented ones. This means that when the computer changes the decision, wishful thinking would increase the probability mass of the option to the right of the implemented decision. In contrast, game-theoretic reasoning and *ex-post* rationalization still mean a higher probability mass on the option to the right of the participant's originally chosen option.[16]

We elicit beliefs directly after each action and 70 participants took part in Experiment 2.A. We use only the random-other and opponent frames since they provide the most conservative treatment comparison by changing only the identity of the target.

### Results of Experiment 2.A

We analyze the data from Experiment 2.A with a dummy regression reported in Table 3. The dependent variable is the reported belief on a single box. Every participant reports 24 Periods $\times$ 4 Boxes = 96 beliefs on single boxes. We regress the beliefs on a set of dummies, indicating whether the particular belief can be influenced by a consensus bias, wishful thinking (WT), hindsight bias, or game-theoretic reasoning/*ex-post* rationalization (GT/EPR) according to the predictions above. Further, we use a frame dummy which is equal to 1 in the random-other frame and 0 in the opponent frame. The constant of this regression is a neutral belief where all dummies are zero. Hence such a belief is unaffected by our candidate effects. Model 1 uses all observations where the participant made the ultimate decision.[17] Wishful thinking and GT/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 GT/EPR $\times$ WT is perfectly collinear with the

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

[17]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 |
|---|---|---|
| False consensus | -0.127 | 0.701 |
| | (2.132) | (1.980) |
| False consensus × Random-Other Frame | 7.677*** | -0.043 |
| | (2.802) | (2.165) |
| Hindsight Bias | -1.729 | -1.211 |
| | (1.819) | (1.799) |
| Hindsight Bias × Random-Other Frame | 1.481 | -1.839 |
| | (2.070) | (2.195) |
| Belief to the right (GT/EPR & WT) | 19.353*** | |
| | (3.436) | |
| Belief to the right (GT/EPR & WT) × Random-Other Frame | -6.650* | |
| | (3.924) | |
| GT/EPR | | 8.690*** |
| | | (2.529) |
| GT/EPR × Random-Other Frame | | -2.257 |
| | | (2.588) |
| Wishful thinking (WT) | | -0.451 |
| | | (1.081) |
| Wishful thinking (WT) × Random-Other Frame | | 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 3:** Linear dummy regressions of single belief elements. Standard errors in parentheses clustered on subject level. Asterisks: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

implementation error.

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

*Ex-post* rationalization should occur exclusively (or at least to a much larger degree) in the opponent frame: 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 "Belief to the right" (with Frame = 0) should capture the added effects of game-theoretic reasoning and *ex-post* rationalization. The "Belief to the right" in the random-other frame (Frame = 1) should capture mostly game-theoretic reasoning only and no (or less) *ex-post* rationalization. Hence, the interaction effect "Belief to the right × Frame" provides an estimate for the differential effect of *ex-post* rationalization. Like in Experiment 1.A, the average best-response rate is higher in the opponent

frame than in the random-other frame when the computer does not change the decision (opponent: 62.1%, random other: 45.2%, rank-sum test $p = 0.006$).[18]

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 frame. Also, there is no evidence for wishful thinking or a hindsight bias. However, GT/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.[19]

### Estimates of unbiased beliefs

The results in Table 3 also give evidence on the size of the respective biases. Having quantified the biases, we are able to reconstruct an estimate for participants 'unbiased' beliefs. We do this correction for all observations used in Model 1. To do so, we subtract the estimated coefficients for the biases from participants' reported beliefs whenever indicated by the respective dummy variables. Subsequently, we re-scale the beliefs to 100%. This procedure yields estimates for unbiased beliefs only on the average level because participants might differ, for example, in how strongly they project their own behavior onto others. Further, we exclude beliefs with multiple best responses and extreme beliefs (that place 100% probability mass on one box) from the correction. Uniform or extreme beliefs are likely to be formed by some alternative process, where the biases do not apply.[20] It hence does not make sense to correct for the biases in these cases. For consistency, we re-run Model 1 in Table 3 excluding beliefs with multiple best-responses and extreme beliefs for the correction. The estimation results are similar to the results in Table 3 and reported in Table A2 in the Appendix.[21]

We correct beliefs for the consensus effect and hindsight bias, and depending on the frame. As already mentioned, we interpret the coefficient of (Belief to the right $\times$ Frame) as the effect size of *ex-post* rationalization in the opponent frame. We hence correct beliefs for this coefficient as well, but *not* for our estimate of Game Theoretic thinking (which would be 'Belief to the right' + 'Belief to the right $\times$ Frame'). We then compare actual decisions and corrected beliefs, and compute the best-response rate under the hypothetic 'unbiased' beliefs. We do this for every participant separately.

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

[19] The interactions are: (False consensus $\times$ Wishful thinking), (False consensus $\times$ Wishful thinking $\times$ Frame), (Hindsight Bias $\times$ Wishful Thinking) and (Hindsight Bias $\times$ Wishful Thinking $\times$ Frame).

[20] For example, it seems very unlikely that people hold unbiased beliefs very often that are exactly uniform *after* the biases play out.

[21] The following results continue to hold when we use the unrestricted estimates in Table 3 to correct beliefs.

As we have shown above, the original best-response rates differ across frames.[22] However, the corrected average best-response rates do no longer differ significantly across frames (opponent: 46.2%, random other: 44.8%, rank-sum test $p = 0.959$). This result suggests that we can 'debias' the reported beliefs to estimate the true amount of game-theoretic thinking in the to-your-left game. In this perspective, the original best-response rates are biased upwards in the opponent frame (signed-rank test, $p < 0.001$) and biased downwards in the random-other frame ($p = 0.013$).

**Discussion of Experiment 2.A**

We interpret the results in the following way: there is a consensus bias in the random-other frame. There is 'game-theoretic reasoning' in both frames, but it is stronger in the opponent frame. We argue that this difference is due to *ex-post* rationalization, which is less important or absent in the random-other frame. Finally, a hindsight bias does not seem to play a role.

As in Experiment 1.A, the framing differences in Model 1 affect measured belief-action consistency, with higher observed best-response rates under the opponent frame compared to the random-other frame. Using the estimates, we can correct for the observed biases to find participants' hypothetic 'true beliefs before reporting' and assess the amount of game-theoretic reasoning in the game. Our results suggest that this is indeed possible. The framing difference vanishes under the corrected beliefs, which suggests that we did not miss any process that would affect beliefs differentially in the two treatments. The estimated 'true' best-response rates of about 45% suggest the degree of game-theoretic reasoning may be over-estimated in many of the existing studies.

When the computer overrides participants' decisions, only a certain degree of game-theoretic reasoning survives in the reported beliefs: also in such cases, participants *on average* seem to report beliefs that make sense given their actions, 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 large 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.

Experiment 2.A was able to disentangle consensus bias, hindsight bias, and—albeit with a caveat—wishful thinking from game-theoretic reasoning/*ex-post* rationalization. The results hint towards overestimated observed best-response rates under the opponent frame, mainly due to *ex-post* rationalization, and underestimated best-response rates in the random-other frame due to a consensus

[22] The original best-response rates differ also when using only observations with unique best-responses and which are not extreme (opponent: 55.1%, random other: 42.3%, rank-sum test $p = 0.071$).

[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%.

effect. However, the evidence with respect to the discrimination between game-theoretic reasoning and *ex-post* rationalization is only suggestive. To disentangle these two aspects, we need Experiment 2.B.

## 6.2 Experiment 2.B: Identifying *ex-post* rationalization

In Experiment 2.B, we eliminate 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.A.[24] Comparing the own-action probabilities from this treatment to the corresponding probabilities from Experiment 1.A 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.A and Experiment 2.B (as stated above, we see little scope for *ex-post* rationalization in the random-other frame). 86 subjects participated in Experiment 2.B.

### Results of Experiment 2.B

The results in Figure 5 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 (2016). We interpret the results as additional evidence for *ex-post* rationalization in the opponent frame.

# 7 Conclusion

This paper uses several experimental manipulations to study under which circumstances game-theoretic thinking, *ex-post* rationalization, hindsight bias, wishful thinking, and a consensus bias influence a person's reported belief. Eliciting beliefs in a question targeting people who are not

[24] *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 Experiment 2.B), 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.

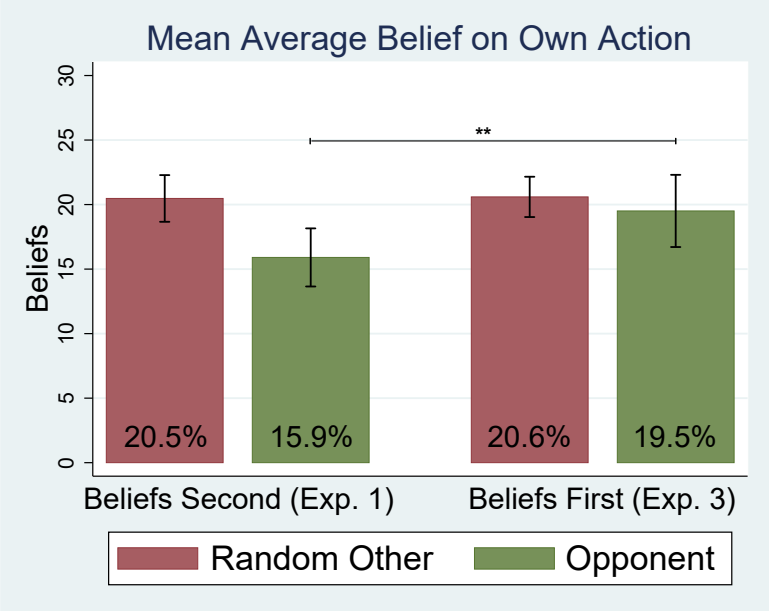



**Figure 5**: Beliefs in the Beliefs-First and the Beliefs-Second treatments. Error bars indicate 95% confidence intervals. Rank-Sum tests: ** $p < 0.05$. For all tests, the data is aggregated on the individual level across all periods yielding one independent observation per participant.

the participant's current interaction partners causes beliefs to be influenced by a consensus bias. A participant with such a belief reports a high subjective probability that others choose the same action as herself. When the question focuses on the participant's current matching partner, there is evidence of *ex-post* rationalization. Under *ex-post* rationalization, the reported belief is fitted to the action and not vice versa. There is no evidence of a hindsight bias or wishful thinking but substantial game-theoretic thinking in all conditions. This means that reported beliefs are consistent with behavior on average. However, the systematic variation in beliefs affects belief-action consistency in predictable ways. Furthermore, we show that the same manipulations can also affect game behavior, which suggests that they also have an influence on participants' underlying beliefs, not 'only' on their reported beliefs.

The findings suggest that there may not be an 'innocent' belief-elicitation method. In this study, participants faced a comparatively strong monetary incentive to report their true beliefs. Moreover, we incentivized the belief reports by a state-of-the-art mechanism that is proper even for people who do not comply with expected-utility maximization. And still, we do not seem to be able to find a way of asking for a belief that leads to an unbiased belief report, unless we ask before participants take their actions. If we were to recommend any method at all, we therefore would recommend to elicit the beliefs before (or potentially, at the same time as) the corresponding actions, using the opponent frame. Of course, this might bias our estimate of strategic thinking upwards; however,

at least in our data set, we find little evidence that it does.

By correcting beliefs for the biases, we are able provide an additional estimate for participants' unbiased beliefs. Using the 'debiased' beliefs, we calculate a 'debiased' best-response rate. The 'debiased' best-response rates suggest that we included all relevant biases and processes as, after the correction, there is no framing difference left to explain. The 'debiased' best-response rate also provides a strong indication that many of the papers in the literature may have over-estimated the degree of game-theoretic reasoning present in economic experiments. This concerns in particular studies in which (a) the opponent frame was used or (b) the population frame was used and—unlike in our study—actions were strategic complements.

On a methodologic note, our findings are important for experimental researchers who wish to elicit beliefs. The choice of method brings about systematic differences in results. For example, our findings are able to shed some light on why studies documenting a consensus bias all seem to use a population frame, while studies that are after consistency use the opponent frame. Moreover, our findings can inform also other applied researchers: in surveys about inflation, future demand, and other important indicators, reported expectations are likely to be biased. First, a manager might *ex-post* rationalize a recent investment decision by reporting favorable expectations. Hence, researchers will have to control for major question-related recent investment decisions (and be it the decision *not* to invest). On the other hand, when asked for the outlook of a typical company of the same branch, the manager might project an unfavorable situation of the manager's own company onto other enterprises, downplaying the importance of other relevant indicators. These considerations provide support for the necessity of taking into account the effects of belief biases in any survey, questionnaire, or experiment that asks people for their beliefs.

# 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.

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

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.

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.

Ellingsen, T., Johannesson, M., Tjøtta, S., & Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, 68(1), 95-107.

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 behavior. 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, CA: 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.

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 behavior. *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.

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.

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 behavior. *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.

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.

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 behavior. *Psychological Bulletin*, 105(1), 31.

Wolff, I. (2015). When best-replies are not in equilibrium: understanding cooperative behavior. *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

# 8 Appendix

## A Figures & Tables

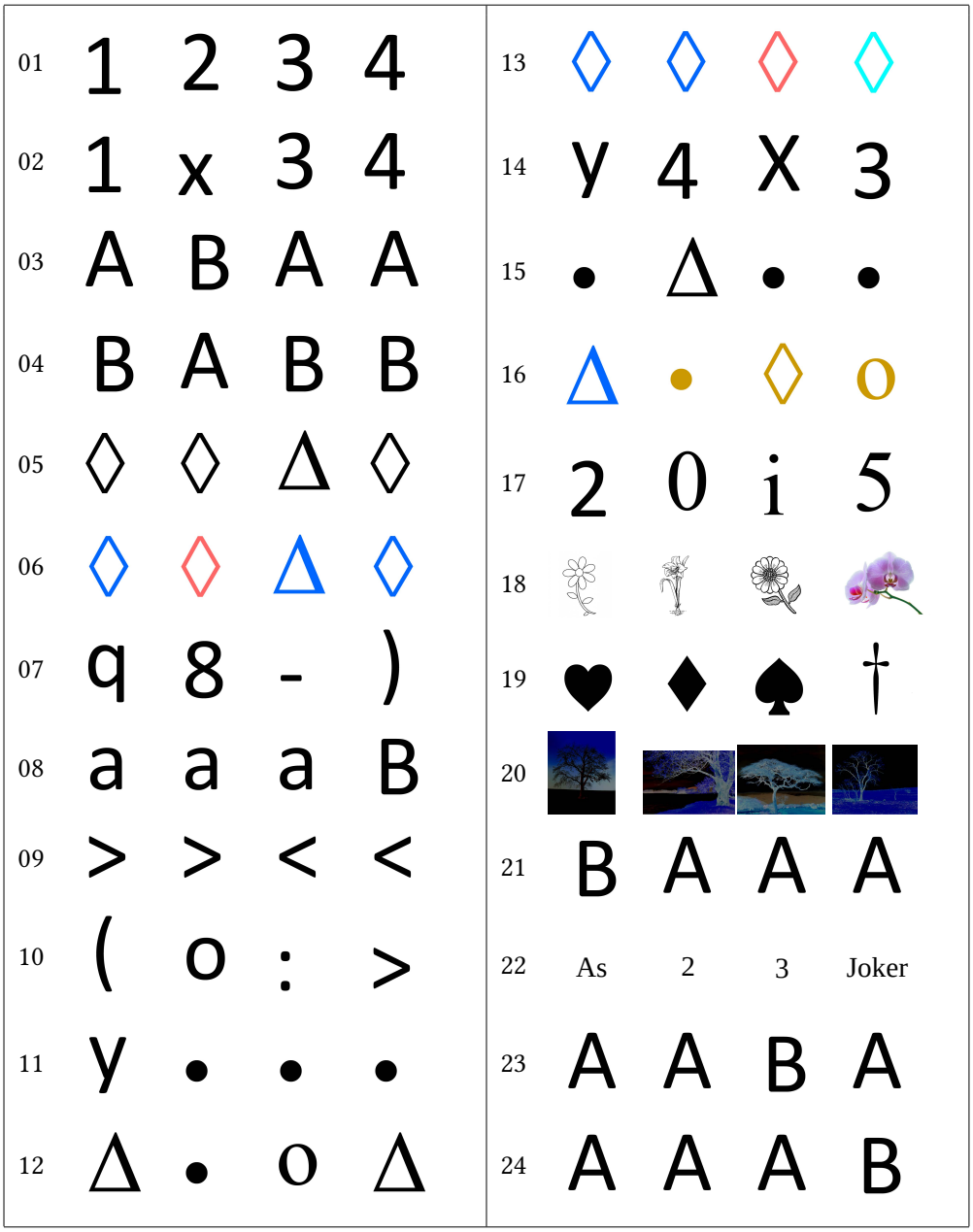



**Figure A1**: The 24 label sets, used to label the four options of the game. One set for each period.

| **Single Belief** | Model $1'$ |
|---|---|
| False consensus | -0.127<br>(2.133) |
| False consensus × Frame | 7.677***<br>(2.804) |
| Belief to the right (GT/EPR & WT) | 19.353***<br>(3.439) |
| Belief to the right (GT/EPR & WT) × Frame | -6.650*<br>(3.926) |
| Hindsight Bias | -1.729<br>(1.820) |
| Hindsight Bias× Frame | 1.481<br>(2.071) |
| Same Choice by the Computer | 0.610<br>(1.121) |
| False consensus × Same Choice by the Computer | 2.171<br>(2.233) |
| False consensus × Frame× Same Choice by the Computer | -3.127<br>(2.699) |
| Belief to the right (GT/EPR & WT) × Same Choice by the Computer | -3.787<br>(3.480) |
| Belief to the right (GT/EPR & WT) × Frame× Same Choice by the Computer | 1.036<br>(4.077) |
| Hindsight Bias× Same Choice by the Computer | -0.200<br>(2.620) |
| Hindsight Bias× Frame× Same Choice by the Computer | 0.983<br>(3.152) |
| Constant | 20.301***<br>(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: *** $p < 0.01$, * $p < 0.1$

| **Single Belief** | Model $1''$ |
|---|---|
| False consensus | -0.251<br>(2.136) |
| False consensus × Frame | 7.330***<br>(2.389) |
| Hindsight Bias | -1.810<br>(2.042) |
| Hindsight Bias × Frame | 0.510<br>(2.017) |
| Belief to the right | 18.448***<br>(2.506) |
| Belief to the right × Frame | -5.433*<br>(3.104) |
| Constant | 20.588***<br>(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 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 experiment 3 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.

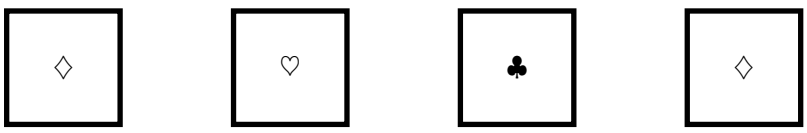
**Instructions for decision 1**
In today's experiment, you will interact with other participants. **You will be randomly rematched with a new participant of today's experiment in every situation.**
Decision 1 works in the following way: You and your matching partner see th exact same screen. On the screen, 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 behavior 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 behavior of your matching partner correctly. It is our intention, that you have an incentive to think carefully about the behavior of your matching partner. We want, that you are rewarded if you have assessed the behavior 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 behavior of participant B and report that assessment truthfully.**