

Wolff, Irenaeus

Conference Paper

(Un-)Stable Preferences, Beliefs, and the Predictability of Behaviour

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Experiments - Games III, No. D18-V3

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Wolff, Irenaeus (2017) : (Un-)Stable Preferences, Beliefs, and the Predictability of Behaviour, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Experiments - Games III, No. D18-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/168231>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

(Un-)Stable Preferences, Beliefs, and the Predictability of Behaviour[§]

Irenaeus Wolff

*Thurgau Institute of Economics (TWI) / University of Konstanz
Hauptstrasse 90, 8280 Kreuzlingen, Switzerland
wolff@twi-kreuzlingen.ch*

Abstract:

Standard economic theory assumes that agents have stable preferences, form beliefs about others on the basis of the available information, and best respond to these beliefs. This paper supports recent findings that we may have to relax the assumption of stable preferences and points to an additional element that is missing in the above account: whether people best-respond to their beliefs depends on the degree to which they trust the available information. In games in which the experiment participants can observe the revealed preferences of their interaction partners, I show the importance for behaviour of whether participants generally believe in others' preference stability. Whether a participant's behaviour can be predicted by her best-response or a Nash-equilibrium depends heavily both on the participant having stable preferences and on her believing in others' preference stability. Generally believing in others' preference stability typically fosters predictability because such a belief is associated with less dispersed beliefs also in specific situations.

Keywords: Public good, social dilemma, Nash-equilibrium, rational beliefs, conditional cooperation, social preferences, preference stability.

JEL: C72, C92, D83, H41

[§]I am grateful to Urs Fischbacher, Alexander K. Wagner, and the lively research group at the Thurgau Institute of Economics (TWI) for helpful comments all along the way. I thank also Sebastian Fehrler, Simon Gächter, Botond Köszegi, Heinrich Nax, Louis Putterman, Jean-Robert Tyran, and Ro'i Zultan for their comments on earlier drafts, as well as seminar audiences at the Universities of Vienna, Innsbruck, and Nottingham, the participants of the ESA European Meetings 2012 and 2015, and the 15th International Conference on Social Dilemmas for the many fruitful discussions. Funding by SNF grant 100018_152788 is gratefully acknowledged.

1 Introduction

Standard economic theory assumes preferences are stable. Recent studies have called this assumption into question for a variety of domains.¹ The first contribution of this paper is to add the question: do people think that others' preferences are stable? This question is important from a theory-perspective because signalling models, reputation formation, and concepts like the intuitive criterion in game theory can only work if an action reveals something about the signalling agent. If observing agents believe preferences are instable, how much will they infer from a signal? The question also is important from a more applied perspective, because sometimes, people get to know the behaviour of others in a situation where there were no strategic concerns involved, potentially because those others thought they were unobserved. These behaviours sometimes constitute useful information for a different situation, but only when we can rely on the underlying preferences being stable. So, the question arises: will people act on these prior observations? To isolate this question, I study experimentally an environment that minimises strategic concerns on the preference-revelation stage.

The main contribution of this paper is to document that whether people generally believe in others' preference stability is a crucial determinant of their behaviour. This general belief determines whether people's choices can be predicted by standard concepts like best-responses to beliefs or Nash-equilibrium as much as people's preference stability itself. *A priori*, we would not expect beliefs about preference stability to affect whether people best-respond to their beliefs about actions: best-responding to the beliefs about actions is the best people can do, irrespective of how those beliefs are formed. However, generally believing in others' preference stability reduces participants' uncertainty about the other player's action also in a specific interaction, which is generally associated with higher best-response-to-belief rates. Hence, I show that both typologies—whether people have stable preferences or not, and whether they generally believe in others' preference stability or not—are critical for whether we can predict their behaviour by standard game-theoretic concepts like best-replies and Nash-equilibrium.

In the main part of the paper, I use a linear public-good situation as a workhorse, because it is a well-studied situation and a prime example for a situation in which we know the material structure but we do not know people's preferences. I use three typologies in this paper. I classify participants into (i) preference types according to the well-established conditional-contribution preferences, I categorise participants (ii) by whether they have stable preferences and (iii) by whether they believe other participants have stable preferences. I predict each par-

¹E.g., Baucells and Villasís (2010) for risk preferences, Stahl (2014) for ambiguity preferences, Meier and Sprenger (2015) for time preferences, or Brosig et al. (2007) for pro-social preferences.

1 INTRODUCTION

participant's choice in a simultaneous interaction with yet another player, based on the participant's elicited preferences and her elicited belief about the other player's action. Best-responses to beliefs about the other player's action predict about three quarters of the choices of 'stability believers' with stable preferences (who have stable preferences and believe in others' preference stability), compared to only one quarter of the choices of 'instability believers' with instable preferences. These findings do not hinge on stable-preference types or 'stability believers' being of a particular conditional-cooperation type.

I can go one step further. The two 'stability typologies' determine whether I can predict a participant's choice in the simultaneous interaction even without resorting to the participant's belief about her interaction partner's action. Roughly two thirds of the choices of 'stability believers' with stable preferences are predicted by looking at a prespecified equilibrium of the game defined by the participants' elicited preferences. This figure compares to only one third of predicted choices amongst 'instability believers' with instable preferences and short of fifty percent of predicted choices for the 'intermediate' categories. Note how strong this finding is: I elicit participants' preferences in a (sequential) task, repeat the measurement twice, and elicit participants' beliefs over whether others show stability over the first two preference measurements. What I learn from the repeat-measurement and from the stability-belief elicitation then tells me whether I can predict participants' choices in a new (simultaneous) interaction, using only their elicited preferences.

Looking at the population as a whole, slightly less than half display stable preferences, and slightly more than half believe in others having stable preferences. The two typologies are not independent: having stable preferences is associated with a clearly higher probability of believing in preference stability of others. This is in line with a false-consensus effect. On the other hand, stability and beliefs in stability are both linked to preference types: mainly complete free-riders and perfect conditional cooperators have stable preferences and people expect the correct preference types to be stable.

In a second part, I examine whether the findings generalise to a different context, using an analogous experimental setup. Here, I use the investment task introduced by Gneezy and Potters (1997) to elicit risk preferences, and add a market context where it pays to invest into a risky option security only if you believe others have sufficiently stable risk preferences to buy similar amounts of market shares as on a first (observed) occasion. The main findings carry over to participants with stable risk preferences: their behaviour is a best-response to their beliefs in 79% of the cases if they generally believe in others' preference stability, but only in 54% of the cases if they do not. On the other hand, the results for participants with instable risk preferences differ markedly: here, participants believing in others' preference stability best-respond 64% of the time, while those who do not

1 INTRODUCTION

best-respond only 44% of the time.² To understand this second effect, we also have to look at participants' general belief in others' stable preferences: 'instability-believers with instable preferences' strongly increase their investments from the first to the second occasion and expect others to do the same; so it is precisely the disbelief in others' preference stability that makes these participants optimistic and often lets them choose the option security—which by their optimistic beliefs is also their optimal choice most of the time. At the same time, a substantial fraction of the (few) 'stability-believers with instable preferences' invest the most in the first investment task; in other words, judged by their average investment, the predicted choice rests on too optimistic a preference measurement (namely, on the first investment decision). This, together with their less optimistic beliefs means they choose the option security less often, although their beliefs would still justify choosing the security.

This paper is the first to go beyond documenting preference (in-)stability but to look at what this instability means for strategic behaviour. Perhaps more importantly still, the paper is the first to provide evidence that a general belief in others' stability is an important determinant of strategic behaviour. Up to today, most research has been based on the idea that people have stable, inalienable preferences, and that when they interact, these preferences will transform the situation into a game. In this framework, there is no room for beliefs about others' preference stability. The only reasons to think that a person's earlier behaviour may not be informative about what she currently is up to has been that the person was using some signalling strategy—or making an error.³ However, once preferences are seen as malleable and unstable, beliefs over preference stability become relevant when we want to predict strategic behaviour. In this perspective, the present paper opens the door to a completely different world, in which we have to account for issues of instability. The good news is that there are regularities also in this world, a few of which I document in this paper.

A second contribution of the paper is a precondition to the main contribution: to study whether people believe in the stability of preferences in others. Studying beliefs about preference stability is important because having seen how others react on different occasions is a pervasive feature of everyday life. Therefore, we need to understand how what people observe others do translates into what people expect these others to do on the next occasion.⁴ In this paper, I abstract from repu-

²Note that this substantial effect does not reach statistical significance due to a lack of observations in the 'stability-believers with instable preferences' category.

³This discussion leaves out random-preference models (Loomes and Sugden, 1995) and models of deliberate randomisation (Machina, 1985), both of which have been proposed to account for instable behaviour in risky decisions.

⁴This relates to the literature on social learning. However, in this literature, the focus usually is on learning about a common state of nature, rather than about the future behaviour of others.

1 INTRODUCTION

tation concerns and rule out both learning about the game and strategic reasoning when eliciting participants' preferences. This approach enables me to isolate the effects of knowing about the other person's preferences on a person's choices and beliefs. Thereby, I add to a huge literature that has studied reactions to observations in repeated interactions, using situations in which the behaviour participants observe is determined by strategic reasoning and reputation concerns.⁵

Finally, this is only the second paper to study whether preferences for positive reciprocity are stable over the short time span of an experimental session. Studying preference stability is important because our view of the world as economists takes preferences as a basic input, and we find this concept useful primarily because we have relied on preferences being stable over at least some period of time.

The main part of this paper uses a linear public-good protocol as a test bed. Following Weibull (2004), I use the term public-good *protocol* to denote a situation in which the material consequences have a public-good structure. Whether this translates into a public-good *game* then depends on participants' preferences. In the public-good protocol I use, two players have to divide their money between a common and a private account. Money in the common account is multiplied by $4/3$ and divided evenly amongst the two players, money in the private account is paid out 1:1. Hence, the amount that ends up in the common account increases the sum of payoffs, but a rational payoff-maximising player has a dominant strategy to put everything in the private account. Yet, if people have social preferences, the set of Nash-equilibria defined by their preferences may not be the same as the full-defection equilibrium in our textbooks.⁶

As many other papers before, I focus on conditional-contribution preferences: how much people want to contribute to a public good depending on what others contribute.⁷ To assess preference stability, I elicit these preferences three times throughout the experimental session. To measure participants' beliefs about others' preference stability, I show participants other players' choices from the first preference-elicitation experiment and elicit their beliefs about how these others will act in the second run of the preference-elicitation experiment. Finally, I want to relate participants' preference stability and their beliefs about others' preference stability to their own strategic behaviour. To do so, I let participants play the simultaneous public-good protocol, showing them the other player's choices from the first run of the preference-elicitation experiment, and ask them for their beliefs on the other player's action.

⁵E.g., Bohnet and Huck (2004), Dal Bó (2005), or Duffy and Ochs (2009).

⁶See Wolff (forthc.) for a systematic elicitation of equilibrium sets for public-good protocols.

⁷Cf. the many references provided in Fischbacher and Gächter (2010), or Fischbacher et al. (2012). Conditional-cooperation preferences may be a type of social preferences in their own right, or a manifestation of underlying preferences, e.g., for reciprocity.

1 INTRODUCTION

Related literature. This paper elicits conditional-contribution preferences in a sequential public-good experiment to study strategic interactions in a simultaneous public-good experiment played afterwards. This rests on the assumption that the sequential experiment measures the preferences that are relevant for the simultaneous experiment. Fischbacher et al. (2012) establish the behavioural validity of elicited conditional-contribution preferences for the simultaneous protocol.

This paper brings together the behavioural-validity aspect with the instability aspect. In recent years, there have been a handful of studies measuring preference stability over weeks, months, and even years. For the case of conditional-contribution preferences, Volk et al. (2012) find stable aggregate behaviour over five months, whereas Brosig et al. (2007) and Sass and Weimann (2012) find only selfish participants to exhibit preference stability.⁸ Carlsson et al. (2014) find temporal and contextual stability of behaviour in public-good situations over six years in a non-student sample in rural Vietnam. Given the mixed findings, it could be that the time intervals in the above studies are too large. So far, Muller et al. (2008) is the only paper on short-term instability of conditional-contribution preferences. They find substantial instability, which they attribute to confusion. The present paper wants to measure instability that is not due to confusion. To this end, I limit the scope for confusion by inviting only experienced participants, by choosing a simplified version of the game, and by using an interface geared to making the choices-payoff relation as transparent as possible.

Last but not least, there is a huge literature in psychology on attribution theories, relating to the inferences people draw about others from observing their behaviour.⁹ This research seems to focus predominantly on the correspondence bias (which is closely related to the well-known fundamental attribution error), according to which people tend to underweight situational influences and overweight personality dispositions when evaluating others' behaviour. Interestingly, researchers do not seem to have identified proneness to a correspondence bias as a personality trait. This may be explained by the idea that—taken to the extreme—believing in true character traits as a researcher might be giving evidence of being prone to the fundamental attribution error oneself.¹⁰ At the same time, the fact that people from different cultures display the bias to different degrees in a predictable way would suggest there may also be rather stable individual differences between people with the same cultural background.¹¹ This paper differs from the literature on the correspondence bias in three ways: in this paper, (i) there are no situational influences that could explain differences in beliefs in stability, (ii) the

⁸Closely related, Bruhin et al. (2016) find stability of preferences for reciprocity and distributional concerns over three months.

⁹Cf., e.g., Gilbert (2002) or Gawronski (2004).

¹⁰Indeed, there is such a debate, in particular with respect to virtues, cf. Harman (2009).

¹¹E.g., Choi and Nisbett (1998) or Miyamoto and Kitayama (2002).

paper identifies the belief in others' preference stability as a stable trait, and (iii) the paper examines also the consequences this trait has, in this case for strategic behaviour.

2 Experimental Design

The focus of this paper is on predictability of behaviour depends on whether people have stable preferences in the short run and on whether they generally believe in stable preferences in others. I use a public-good context because it is a well-studied environment, and because it is rich enough to allow for heterogeneous behavioural patterns. To measure preference stability in this context, I elicit participants' preferences for conditional cooperation three times within a session: twice at the beginning, and a third time as the final part of the session (PREFS1, PREFS2, and PREFS3). For each of the PREFS-experiments, I use the method introduced by Fischbacher et al. (2001, see Section 2.1 for details). For none of the experimental parts do participants get any feedback before the end of the session.

To examine what potential instability means for strategic behaviour, I look at a simultaneous public-good protocol (SIMPG, played after PREFS2). As a benchmark for strategic behaviour, I chose the Nash-equilibria that result when interpreting the elicited preferences from the PREFS-experiments as participants' best-reply correspondences in the SIMPG-experiment. I call the set of all mutual best-replies resulting from the elicited preferences the set of *revealed-preference Nash equilibria* (RPNE; cf. Wolff, 2015). I show participants their interaction partner's behaviour from the PREFS1-experiment prior to the SIMPG-experiment, for two reasons (see below for a discussion of the potential signalling issue). First, any Nash-equilibrium concept requires common knowledge about the game being played. Showing participants the PREFS1-behaviour of their interaction partner approximates at least mutual knowledge of the game. And second, I argue that having seen how others react on different occasions is a pervasive feature of everyday life. Revealing others' preferences is an interesting benchmark case to study.

If we want to study how knowing others' reactions to past situations affects their strategic behaviour, we need to elicit their beliefs about the others' actions as well as about others' preference stability. In the part after the SIMPG, I elicit probabilistic beliefs about their interaction partner's SIMPG action. In the penultimate part of each session, I then show participants the PREFS1-behaviour of four other participants and ask them to report probabilistic beliefs on the participants' PREFS2-behaviour. One of the four other participants is the interaction partner from the SIMPG-experiment, the three others are randomly chosen others. Here is a full overview of all seven parts of a session:

FILLERTASK. A social-value orientation task similar to the one presented in Murphy et al.

2 EXPERIMENTAL DESIGN

(2011). Irrelevant for this paper but for the potential signalling issue.

PREFS1 The PREFS-experiment as detailed in Section 2.1.

PREFS2 + BELIEFS. Repetition of the PREFS-experiment with a new interaction partner. Then, I elicited beliefs on the expected first-mover contribution, to train participants in the elicitation method used in SIMPGBELIEFS: probabilistic beliefs elicited by a binarised scoring rule (Hossain and Okui, 2013, probability of receiving an additional payment of 2 Euros determined by a quadratic scoring rule).

SIMPG. The SIMPG-experiment as detailed in Section 2.1.

SIMPGBELIEFS. Elicitation of beliefs on the interaction partner's choice in the SIMPG-experiment. Payment by a binarised scoring rule with payoffs of 20 Euros (if successful) and 4 Euros (if not successful).

STABILITYBELIEFS. Elicitation of beliefs on the stability of preferences of the interaction partner in the SIMPG-experiment and three randomly-chosen others. Participants were shown the response vector of the other participant from the PREFS1-experiment. Then, they had to state a probabilistic belief on the response of the same other participant as a second-mover in the PREFS2-experiment, for each possible contribution level of the first-mover. For each of the four participants whose response stability participants had to assess, one first-mover contribution was randomly chosen. Then, participants were paid by a binarised scoring rule for their belief accuracy in the four randomly-selected cases, with a payment of 6 Euros per lottery.

PREFS3. Second repetition of the PREFS-experiment with a new interaction partner.¹²

Note that if participants know their behaviour in one experiment may be revealed to others in the next experiment, they may have potential signalling incentives in the first of the two.¹³ My experimental design allows to counter this problem through a number of design choices, discussed in full detail in Wolff (2015) on a very similar earlier design. Most importantly, participants make decisions in seven distinct experimental parts with new interaction partners in each of them, being paid for only one randomly chosen experiment (which should make signalling

¹²In contrast to the first two PREFS tasks, the first-mover in PREFS3 was shown the response-vector of the second-mover from the PREFS1-experiment before deciding on her (unconditional) contribution. However, the situation of the second-mover was exactly the same as in the PREFS1- and PREFS2-experiments. For the purpose of this paper, I therefore regard the PREFS3-experiment simply as a second repeat-measurement of participants' preferences.

¹³To avoid deceiving participants, the instructions included the sentence that "your behaviour from one of the earlier parts will possibly be displayed to other participants in a later part."

2 EXPERIMENTAL DESIGN

prohibitively costly). They do not get any feedback about others' behaviour before the SIMPG-experiment, and each experiment is explained only as soon as it begins. While it is impossible-in-principle to show there have been no signalling attempts by participants, I could not find any evidence of signalling in the data. Also, if I were to base my analysis on the data from the final PREFS3-part in which there could not be any signalling incentives anymore, my main Result 1 would get even stronger, not weaker.

2.1 The SIMPG- and the PREFS-experiments

The SIMPG-experiment consists of a simultaneous two-player linear public-good situation with an $MPCR = \frac{2}{3}$ and an endowment of 15 Euros. Each player has to choose a contribution to the public good from the set $\{0, 3, 6, 9, 12, 15\}$ Euros, which is multiplied by $\frac{4}{3}$ and divided equally among the two players, regardless of each player's own contribution.

In the PREFS-experiment, participants face the same two-player linear public-good payoff structure with an $MPCR = \frac{2}{3}$ and an endowment of 15 Euros as in the SIMPG-experiment. However, the PREFS-experiment differs from the SIMPG in that the former is a sequential game: one participant moves first and the other moves second, being informed of the first participant's choice. Participants have to decide in either role. First, they specify their first-mover contribution to the public good that is implemented if they are not (randomly) chosen to be the second-moving player. Then, I elicit their second-mover choices using the strategy method: they are presented with all possible first-mover contributions and asked to specify their 'conditional' contributions.¹⁴ I hold the conditional-contribution schedules from the PREFS-experiment to be a direct expression of participants' (proximate) preferences. Therefore, I equate schedules and best-response correspondences for the remainder of this article.

To limit the scope for confusion as a major source of (measured) preference instability, I took three measures. First, I restricted the simultaneous game to a two-player six-action game rather than the usual three- or four-player games with 11-21 actions. While the MPCR may look a little complicated, all game payoffs were integer amounts. Second, I always displayed the full payoff matrix in the relevant parts. Moreover, I highlighted the relevant part of the matrix in the preference-elicitation parts of the PREFS-experiments, so that participants would know exactly what payoff profile each of their actions meant. As a third measure, I recruited *experienced* participants.¹⁵ Participants in the experiment had participated in at

¹⁴The order of the combinations was randomised individually for each player. Responses were elicited one-by-one for two reasons: (i) to make each decision as salient as possible, (ii) to elicit 'smooth' response-patterns only in case preferences gave rise to them.

¹⁵I nonetheless asked the usual comprehension questions; participants could only proceed to

3 RESULTS

least one public-good experiment and at least four additional other experiments, with no upper limits.

2.2 Procedures

On the day of the experiment, participants were welcomed and asked to draw lots in order to assign them to a cabin. There, they would find some explanation on the general structure of the experiment and on the selection of the payoff-relevant experiment (and role, if applicable). The instructions for each experiment were displayed directly on their screen during the corresponding part. The (translated) general and on-screen instructions are gathered in the Appendix.

Participants earned on average 19.33 Euros (USD 22) for about 90 minutes; this includes a 2-Euro flat payment for the completion of a post-experimental questionnaire. Altogether, seven sessions with a total of 152 participants were conducted at the LakeLab of the University of Konstanz, between June 2015 and June 2016.

3 Results

3.1 Strategic behaviour and preference instability

In the beginning of Section 2, I referred to the idea of calculating the set of equilibria that results for the preferences elicited in a PREFERENCES-experiment. But what can we expect from a *revealed-preference Nash-equilibrium* (RPNE) prediction if there is preference instability? Before I answer this question, I categorise participants into four groups: *stability-believers with stable preferences* (30%), *stability-believers with instable preferences* (13%), *instability-believers with stable preferences* (26%), and *instability-believers with instable preferences* (31%).

A participant is categorised into one of the *stable-preference* groups if she has *approximately stable* preferences, and into an *instable-preference* group, otherwise. I categorise as *approximately stable* all those who have an average squared difference of at most two from the mean response to each first-mover contribution. This criterion would be fulfilled with equality if a participant replies to each first-mover contribution the same way twice, deviating on the third occasion by one increment of 3 Euros in all contingencies.¹⁶

A participant is categorised into one of the *(in-)stability-believer* groups based on her STABILITYBELIEFS. In the STABILITYBELIEFS-part, each participant sees the

the experiment after answering all questions correctly.

¹⁶Using this criterion, there are 66 *approximately stable* participants. If we were to use a median split instead, the threshold would increase to 11/3. Only eight additional participants have an average squared difference from the mean response of less than 11/3.

3 RESULTS

choices of four other participants from the Prefs1-part and has to state a probabilistic belief what each of the four others would have done in the Prefs2-part. One of the four others was the participant's interaction partner in the SIMPG-part. For the *(in-)stability-believer*-group categorisation, I focus on the participant's beliefs about the three players the participant did *not* interact with in the SIMPG-part. I then categorise a participant into one of the *stability-believer* groups if she places at least 80% probability mass on these three other players responding to all possible first-mover contributions the same way in the Prefs1- and the Prefs2-experiments, and into an *instability-believer* group, otherwise. Not looking at the belief about the SIMPG-interaction partner focuses on 'stability believing' as a characteristic of the person, which promises to be more helpful for prediction than interaction-specific measures.¹⁷

This categorisation is useful: it informs us whether we can predict behaviour in the SIMPG-experiment or not. As a first step, I use the preferences elicited in the Prefs1-experiment together with the SIMPGBELIEFS, to predict choices in the SIMPG-experiment.¹⁸ Then, I go one step further and predict choices in the SIMPG-experiment by participants' (highest) RPNE-action.¹⁹ Figure 1 shows the results.

Result 1. Whether a participant has stable preferences and whether the participant *in general* believes in others' preference stability determines whether we can predict reliably the participant's behaviour *in a specific* new game.

The left-hand part of Figure 1 shows that the SIMPG-choices of *stability-believers with stable preferences* can be predicted using the Prefs1- and SIMPGBELIEFS-data in 72% of the cases. The corresponding figure for *instability-believers with instable preferences* equals 26% which even cannot be distinguished statistically from random sampling from all SIMPG choices (binomial test, $p = 0.191$). The choices of *stability-believers with instable preferences* and *instability-believers with stable preferences* can be predicted in roughly half of the cases (49% and 50%, $p \leq 0.014$).^{20,21}

¹⁷Similar results obtain when focusing on the interaction instead.

¹⁸Strictly speaking, the research design does not allow to make a clear prediction from the conditional-cooperation preferences when a participant's belief is non-degenerate. To be able to make a prediction, I use the best-response to the belief's mode. In case of multiplicity, I take the best-response to the highest other-player action that is modal in the participant's SIMPGBELIEF.

¹⁹In case of multiplicity, I use the highest RPNE-contribution. Using the lowest, the mean, or the median RPNE-contributions (rounded to the nearest possible value) does not change the results in any substantial way. I use the highest RPNE-contribution because this yields the overall highest fraction of predicted choices.

²⁰The above figures also tend to be statistically different: a χ^2 -test for all four figures yields $p < 0.001$, Boschloo tests for pairwise comparisons yield $p \in [10^{-5}, 0.080]$ except for the comparison of the 'intermediate' groups ($p = 1$).

²¹If we define a best-response as the reaction to any of: the belief mode, the average belief rounded to the next-possible value or rounded down to the next-possible value (to allow for some pessimism), best-responses account for 43%, 64%, 50%, and 78% of choices (same order as in Fig. 1).

3 RESULTS

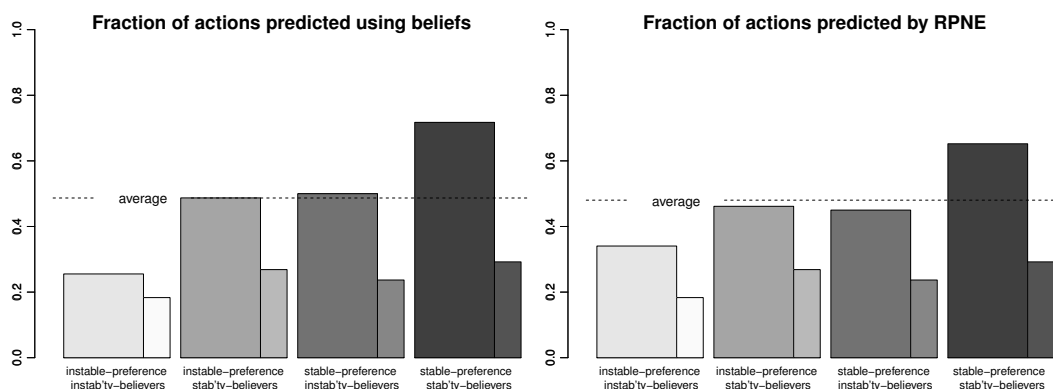


Figure 1: Fraction of SIMPG-choices correctly predicted using either preferences and beliefs (left) or only preferences (right), by whether a participant has *approximately stable* preferences and by whether the participant believes in others having stable preferences. The wide columns represent the data, the narrow columns the corresponding random benchmarks.

Note that Result 1 is not a consequence of easy-to-predict *Defectors* (definition follows in Section 3.3 below) being overrepresented among the *stability-believers with stable preferences*. Not including *Defectors* would yield the following predictability figures (in the same order as in Figure 1): 24%, 45%, 23%, 66%. Result 1 also cannot be explained by *instable-preference* participants learning or revealing their true preferences only by the end: using the preferences elicited in PREFS3 does not decrease the differences, with predictability figures of 26%, 49%, 55%, and 74%. Rather, the unexpected difference in best-response rates between those who generally believe in others' preference stability and those who do not is related to the uncertainty expressed in the participants' beliefs over actions. The standard deviation of the other player's expected action in *instability believers'* beliefs for the simultaneous game is clearly higher compared to that in *stability believers'* beliefs (3.9 vs 2.9, $p = 0.002$, Wilcoxon-Mann-Whitney test). In turn, a higher standard deviation of the other's expected action leads to a lower best-response-to-belief rate.²² There is no such statistical difference with respect to the player's own preference stability.²³

The right-hand part of Figure 1 shows what fraction of SIMPG-choices can be predicted using only data from the PREFS1-experiment. The choices of *stability-believers with stable preferences* are predicted in about two thirds of the cases (65%),

²²Spearman's $\rho = -0.14$, $p = 0.091$; in a probit regression of best-responses on only the expected action's standard deviation (and a constant), the coefficient has $p = 0.045$.

²³A Wilcoxon-Mann-Whitney test for *instability believers'* expectations vs *stability believers'* expectations yields $p = 0.002$ (standard deviations 3.91 vs 2.88). The same test within the *instability believers* yields $p = 0.848$ (3.89 vs 3.95) and $p = 0.670$ within the *stability believers* (2.99 vs 2.79).

3 RESULTS

whereas those of *instability believers with instable preferences* are predicted in only one third (34%). Choices by participants categorised in the ‘intermediate’ groups are predicted in slightly less than half of the cases (46% and 45%; $p \leq 0.034$ for binomial tests against chance for all groups).²⁴ The frequencies of predictable choices go along with the corresponding beliefs: *stability believers with stable preferences* on average put a probability mass of 56% on the event that the other player chooses the (highest) RPNE action, whereas *instability believers* assign only 26% (those in the ‘intermediate’ groups assign 45% and 47%).²⁵ In summary, we can predict *stability believers* and participants with *approximately stable* preferences more easily because they deem it more likely that others will choose a RPNE action, they are more positive about it, and hence, they are more likely to best-respond to it.

3.2 Behaviour of the unpredictable participants

The middle panel of Figure 2 shows the behaviour of those participants whose action could not be predicted by their preferences together with their beliefs, next to the behaviour of ‘predictable’ participants (on the left). The distribution of unpredicted actions looks close to uniform (also by a χ^2 -test, $p = 0.528$).

In the right-most panel of Figure 2, we see the distribution of unpredicted actions split up by the participant’s preference stability and general belief in preference stability. While there are too few observations in each category for a meaningful statistical analysis, eyeballing suggests that behaviour may not be completely random, except for the *instability-believers with instable preferences*. *Stability-believers with instable preferences* seem to prefer to contribute rather little, whereas *instability-believers with stable preferences* seem to prefer to choose some intermediate contribution as a consequence of their uncertainty over others’ behaviour. *Stability-believers with stable preferences*, finally, tend towards the extremes, thereby showing the pattern that is closest to the choices of ‘predictable’ participants.

3.3 Preference-type stability in the PREFS-experiments

Following the literature, I categorise the participants into conditional-contribution types according to their preferences as elicited in the PREFS-experiment. I use

²⁴The given figures also tend to be statistically different: the χ^2 -test for over all groups yields $p = 0.026$, but Boschloo tests indicate differences only for *stability believers with stable preferences* vs. *instability-believers with instable preferences* ($p = 0.003$) and potentially vs. the ‘intermediate’ groups ($p = 0.142$ and $p = 0.080$).

²⁵Wilcoxon Mann-Whitney tests yield $p < 0.01$ for *instability believers with instable preferences* vs each of the other groups, but only $p = 0.171$ for *stability believers with stable preferences* vs the ‘intermediate’ groups.

3 RESULTS

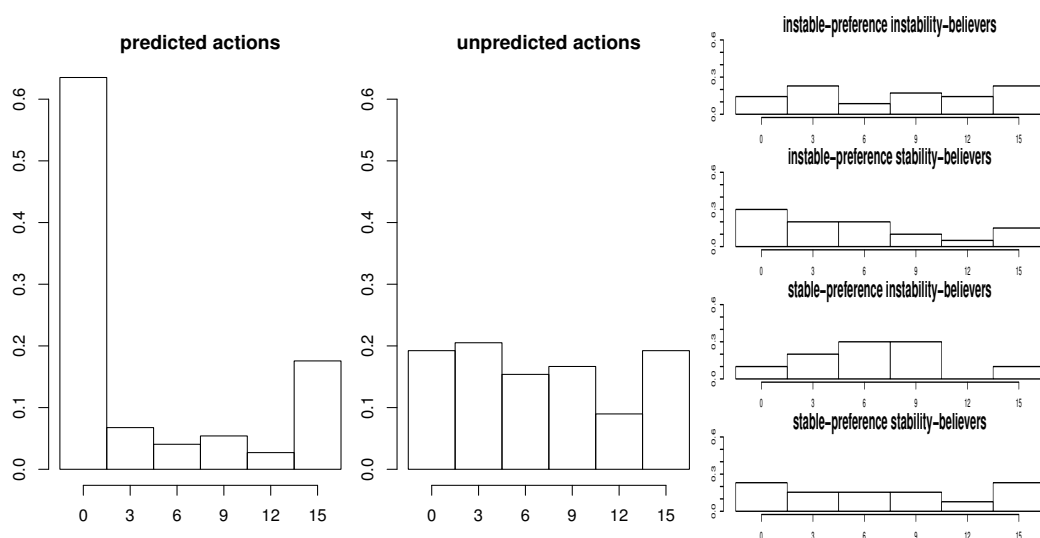


Figure 2: Relative frequencies of successfully predicted actions (left) and unpredicted actions, overall (middle) and by the participant's preference stability and general belief in preference stability (right).

the categories suggested by Fischbacher et al. (2001) but divide the conditional cooperators into three subcategories. To be precise, I use the following categories:

Defectors Always respond by a contribution of 0 Euros.

PerfCCs Perfect conditional cooperators always respond by mirroring exactly the contribution of the first-mover.

ImpCCs Imperfect conditional cooperators have a monotonically-increasing response vector. Respond by less than the first-mover contribution at least some of the time.

NmImpCCs Non-monotonic imperfect conditional cooperators have a non-monotonic response vector, for which the Spearman correlation coefficient with first-mover contributions is positive with $p \leq 0.05$ (one-sided).

Triangles Have a hump-shaped response-vector.

Altruists Always contribute fully with at most one deviation to 12 Euros.

Unclassifiables Cannot be classified into any of the above categories.

Table 1 displays the type distributions, depending on the time of measurement, that is, depending on whether I use the response vector from the PREFS1-, the

3 RESULTS

PREFS2-, or the PREFS3-experiment. Overall, the numbers of conditional cooperators (including all three categories, 41-53%, or roughly half of our sample) and of *Defectors* (roughly one quarter of our sample) are similar to those in Fischbacher et al. (2001).

	PREFS1	PREFS2	PREFS3
<i>Defectors</i>	35	42	41
<i>PerfCCs</i>	29	38	41
<i>ImpCCs</i>	25	24	30
<i>NmImpCCs</i>	8	16	9
<i>Triangles</i>	17	14	14
<i>Altruists</i>	6	2	2
<i>Unclassifiables</i>	32	16	15
Total	152	152	152

Table 1: Distribution of player types, by time of measurement.

Result 2. There is substantial preference instability in our sample. Slightly more than half of the participants show preferences that are not even approximately stable over three measurements within the same session.

As can be seen from Table 1, the frequencies of players classified as specific types fluctuate considerably: the number of participants classified as *Defectors* and *PerfCCs* or *ImpCCs* increases, and the number of participants classified as *Triangles*, *Altruists*, or *Unclassifiable* participants decreases. Despite these fluctuations, there could be a lot of preference stability. At the same time, there may be even more individual-level instability masked by aggregate frequencies. In total, only 44 out of 152 participants (29%) respond exactly the same way in all three PREFS-repetitions. The *approximate stability* criterion (defined in Section 3.1) which allows for some variation is fulfilled by 66 out of 152 participants (43%). Conversely, the elicited preferences of 57% of the participants are not even approximately stable.

Result 3. There is considerable heterogeneity in terms of preference stability: among *Defectors* and perfect conditional cooperators (*PerfCCs*), two thirds have completely stable preferences, but only 1 out of the remaining 88 participants does.

Specifically, 23 out of 35 *Defectors* and 20 out of 29 *PerfCCs* (as classified in the PREFS-experiment) respond exactly the same way in all three PREFS-repetitions. Out of the remaining 88 participants, this holds true only for a single *Triangle*. Approximate preference stability is fulfilled for 24 out of 35 *Defectors*, 21 out of 29 *PerfCCs*, 13 out of 25 *ImpCCs*, half of the *NmImpCCs*, 1 *Triangle*, 1 *Altruist* and

3 RESULTS

	Coefficient	Std. Error	p-value
(Intercept)	71.1	(4.01)	$1 \cdot 10^{-56}$
appr. stable preference	12.9	(3.02)	$2 \cdot 10^{-5}$
same type	7.1	(4.29)	0.0968
appr. stable preference \times same type	-5.5	(5.94)	0.3573
<i>PerfCC</i>	2.7	(3.26)	0.4086
<i>ImpCC</i>	-6.2	(3.62)	0.0895
<i>NmImpCC</i>	-1.3	(5.14)	0.8081
<i>Triangle</i>	1.0	(4.38)	0.8253
<i>Altruist</i>	-15.3	(6.74)	0.0235
<i>Unclassifiable</i>	-15.9	(3.88)	0.0001
other player is a <i>PerfCC</i>	1.4	(3.67)	0.7033
other player is an <i>ImpCC</i>	-12.8	(3.65)	0.0005
other player is a <i>NmImpCC</i>	-23.3	(4.62)	$6 \cdot 10^{-7}$
other player is a <i>Triangle</i>	-19.4	(4.23)	$5 \cdot 10^{-6}$
other player is an <i>Altruist</i>	-26.1	(8.79)	0.0031
other player is <i>Unclassifiable</i>	-28.4	(3.60)	$1 \cdot 10^{-14}$

Table 2: Average probability mass a player’s belief placed on the other player choosing the same reaction in PREFS2 as in PREFS1, regressed on own and other’s preference type and on whether the player has approximately stable preferences. Linear-probabilities model with standard errors clustered on participants.

2 *Unclassifiables*. Again, we see that some 70% of the participants classified as *Defectors* or *PerfCCs* are approximately stable, compared to less than a quarter of the remaining participants.

3.4 Beliefs about preference stability

Do participants themselves believe in preference stability? As a measure for the degree of belief in preference stability, I use the probability mass a participant places on the event that the other participant will respond to the first-mover contribution exactly the same way in the PREFS2- as in the PREFS1-experiment, averaged over all possible first-mover contributions. Table 2 shows the results of a regression of this ‘average belief in stability’ on the participant’s preference type, the other player’s preference type, on whether the two types are the same and on whether the participant has approximately stable preferences.

Result 4. Participants’ beliefs in preference stability on aggregate are ‘rational’: participants correctly believe *Defectors* and *PerfCCs* to display a high degree of

3 RESULTS

response-stability, whereas they correctly expect other types to display more variable responses.

As we see in the lower part of Table 2, only *Defectors* (baseline category) and *PerfCCs* are expected to display a high degree of preference stability. The ‘average belief in stability’ drops sharply when the other player is of other types (by 13-28 percentage points). This corresponds very well to the finding that *Defectors* and *PerfCCs* truly are the most stable types.

Result 5. There seems to be a personal element in believing in others’ preference stability: having stable preferences strongly boosts participants’ beliefs in others’ preference stability, and participants’ own type also seems to have an influence.

As the second row in the main part of Table 2 shows, a participant assigns on average 13 percentage points more probability mass to another participant showing the same preferences again if the participant has *approximately stable* preferences than when she does not. Rows five to ten of Table 2 show that *Unclassifiables*, *Altruists*, and potentially also *ImpCCs* place substantially less probability weight on others showing stable preferences, compared to the baseline category of the *Defectors*. Interestingly, we see only weak evidence that beliefs in others’ preference stability are boosted by both players having the same type of preferences, and no evidence at all if the focal participant has stable preferences.

3.5 Beliefs about others’ SIMPG-choices

How good are people at predicting others’ SIMPG-choices? This question is addressed in Figure 3. Figure 3 plots how much probability mass participants placed on the action chosen by their interaction partners, again depending on whether a participant has *approximately stable* preferences and on whether the participant believes in others having stable preferences.

Result 6. Participants are not able to predict others’ behaviour in a simultaneous public-good situation very well even when they are shown the preferences elicited in a sequential public-good situation like in Fischbacher et al. (2001). (Only) Part of the reason might be that participants underestimate the predictive power of this information.

As we can see in Figure 3, the overall average probability mass participants place on the actual choice of their interaction partner is only one third. Interestingly, participants who believe in others’ preference stability have the (insignificantly) more accurate beliefs, with a probability mass of up to 40% placed on the actually chosen contribution level. This figure increases to 44% if we define ‘stability believing’ in terms of the belief on the SIMPG-interaction partner herself (in

4 BELIEFS IN PREFERENCE STABILITY IN A RISK CONTEXT

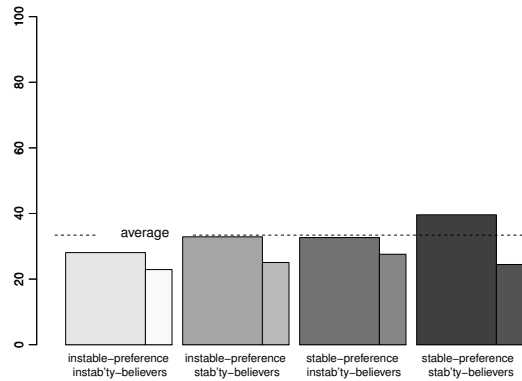


Figure 3: Empirical accuracy of beliefs: probability mass placed on the action chosen by the interaction partner, depending on whether a participant has *approximately stable* preferences and on whether the participant believes in others having stable preferences. The wide columns represent the data, the narrow columns the corresponding random benchmarks.

that case, irrespective of preference stability). The fact that *stability-believers* have the more accurate beliefs on the behaviour of the other player might suggest that *instability-believers* place too little weight on the information they are given. On the other hand, the fact that *stability-believers* also have rather inaccurate beliefs clearly shows that believing in the information can be at most part of the story.²⁶

4 Beliefs in preference stability in a risk context

So far, we have seen that the degree to which a participant believes in others' preference stability in general predicts whether we as researchers can predict the participant's behaviour in a different interaction with a new partner. This section briefly summarises the results of a second experiment that tests whether we find similar results in a completely different context, this time on risky decisions.

4.1 Experimental setup

Experiment 2 had five parts. Again, each part was described only as it started, and the payoff-relevant part was determined randomly. In parts 1, 2, and 5, I measured participant's risk attitudes by the method of Gneezy and Potters (1997): in

²⁶Comparing the distribution of probability masses placed on the other player's SIMPG-action to that of the corresponding probability masses from beliefs randomly sampled from the same participant category (category-mean probability mass depicted in the narrow bars in Figure 3), by means of bootstrapped Kolmogorov-Smirnov tests, yields p-values of 0.016, 0.121, 0.572, and 0.016 (ordered as in Figure 3).

each of these parts, participants had to choose an integer amount x to invest into a risky asset out of an endowment of 10 Euros. With a probability of two thirds, the asset turned worthless, yielding a payoff of $(10 - x)$ Euros, but with a probability of one third, the asset went up in price, yielding a payoff of $(10 + 2.5x)$. In part 3, I showed participants the part-1 behaviour $x_j^{(1)}$ of three other participants and told them they could buy an option security on a market index, whose value was determined by the three other participants' part-2 behaviour $x_j^{(2)}$: if the three participants had invested at least one Euro less than their part-1 sum (i.e., if $\sum x_j^{(2)} \geq \sum x_j^{(1)} - 1$), the option would pay off, otherwise it would be worthless. The price of the option was set to the focal participant's part-1 investment $x_i^{(1)}$ plus 2, the payoff in case of success was $(10 + 1.5[x_i^{(1)} + 2])$ Euros. I chose this payoff to make it a difficult decision for participants, not in order to keep the payoff structure similar to the other parts.²⁷ In part 4, I asked for participants' beliefs by the same method as in the SIMPGBELIEFS- and STABILITYBELIEFS-parts of the main experiment. I asked them about their belief concerning the "market" in part 3 first (about $\sum x_j^{(2)}$), followed by their beliefs about 6 individual other participants' part-2 behaviour $x_j^{(2)}$ (the second, fourth, and fifth being the participants determining the market). For each of the seven beliefs, they got 3 Euros in case their earned lottery tickets were successful. 168 participants participated in 6 sessions in January and February 2017.

4.2 Results

As in the main experiment, I classified participants into those with *approximately stable* vs *instable* preferences and into *stability-believers* vs *instability-believers*. As before, a participant was categorised as having *approximately stable* preferences if her average squared deviation was less than two thirds of the smallest increment (in this case, of 1) from the mean choice in parts 1, 2, and 5, yielding 85/168 participants in the *stable* category. A participant was categorised as a *stability-believer* if she placed an average probability of 60% on the other participants repeating their part-1 investments in part 2, again evaluating only the beliefs on those participants that were irrelevant for the focal individual's part-3 decision.²⁸ Figure 4 shows the

²⁷Note that, had participants known part 3 in advance, they would have had an incentive to change their behaviour in part 1. To avoid any issues of deception, I used the technique introduced by Bardsley (2000): participants knew that 1 out of the 5 experimental parts (part 3) could not become payoff-relevant, without knowing which part this would be. They did not get to know the connection between their part-1 choice and the part-3 situation at any time of the experiment.

²⁸I use a cutoff of 60% rather than of 80% as in the main experiment, for two reasons: first, the smallest increment in Experiment 2 was 1 rather than 3, so erroneous deviations should be expected more often in Experiment 2 (if participants expect others to make Fechner-type errors). And second, this leaves us with 33/168 *stability-believers* rather than with only 16, giving us more statistical

4 BELIEFS IN PREFERENCE STABILITY IN A RISK CONTEXT

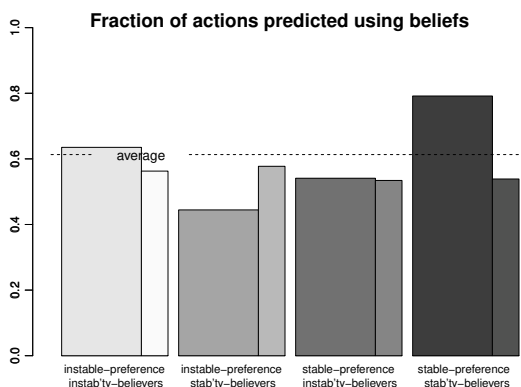


Figure 4: Fraction of part-3 choices correctly predicted using part-1 preferences and market beliefs, by whether a participant has *approximately stable* preferences and by whether the participant believes in others having stable preferences. The wide columns represent the data, the narrow columns the corresponding random benchmarks.

fractions of part-3 choices correctly predicted by participants' risk attitudes (as measured in part 1) together with their market beliefs (on $\sum x_j^{(2)}$).²⁹

Result 7. The general belief in others' risk-preference stability determines how reliably we can predict the participant's behaviour in a market game, in particular for participants with stable preferences.

Result 7 is immediately obvious from Figure 4 for participants with *approximately stable* preferences as well as for participants with *instable* preferences; the only surprise is that this time, the belief or non-belief in others preference stability seems to predict belief-action correspondence *negatively* amongst the participants with *instable* preferences.³⁰ To understand this surprising effect, we have to take

power. Even with the cutoff at 60%, we only have 24 *stability-believers with stable preferences* (and, as in the public-good context, very few—9—*stability-believers with instable preferences*).

²⁹I use the CRRA utility function $u(x) = \frac{1}{1-\rho} x^{(1-\rho)}$, in which part-1 investments $x_i^{(1)}$ directly translate into an estimate of ρ . For choices $x_i^{(1)} = 10$, I postulate $\rho = 0$, whereas for $x_i^{(1)} = 0$, I use an arbitrary $\rho = 1.5$ (which roughly would be the next component of the sequence of ρ s when going from $x_i^{(1)} = 9$ to $x_i^{(1)} = 0$). Both of these extreme cases together make up for 12% of the observations (split roughly evenly). Nothing substantial changes if we exclude these observations from the analysis. Using the estimate of ρ , I then compare the expected utility of buying the option security in part 3 given the probability assigned to the event that the option yields a positive payoff (i.e., that $\sum x_j^{(2)} \geq \sum x_j^{(1)} - 1$) to the utility of 10 Euros, to predict the participant's part-3 choice.

³⁰A χ^2 -test is inconclusive on whether correctly predicted choices differ between the four categories ($p = 0.122$), while Boschloo-tests indicate a significant difference between *stability-* and *instability-believers with stable preferences* ($p = 0.041$) but not between *stability-* and *instability-believers with instable preferences* ($p = 0.261$; note, however, that there are only 9 observations in the former group).

5 CONCLUSIONS

a closer look at the participants in the respective categories. For the participants with *instable preferences*, 91-100% of the predicted actions are to invest in the option security in part 3 (compared to 72-75% for the groups with *stable preferences*). Moreover, the *instability-believers with instable preferences* are the only group who clearly increase their investments from investment decision 1 to investment decision 2 (by 0.70 on average, compared to negative average changes of up to -0.13 in the other groups). If the *instability-believers with instable preferences* then project their own increase in investments from decision 1 to decision 2 onto others (which they do, predicting an average increase of 0.42 amongst those others who do not affect the outcome of the market decision and of 0.34 amongst those who do), their disbelief in others' preference stability will make many of them optimistic enough to invest in the option security. Which means they will tend to choose the predicted option relatively often. Consider now the *stability-believers with instable preferences*. First of all, note that the share of participants for whom the investment in part 1 (which is used to predict the part-3 decision) is larger than their average investment from parts 1, 2, and 5 is largest in this group (44%, compared to 20-25% in the other groups). This means that the prediction will tend to overestimate these participants' willingness to take risks. Moreover, their belief in stable preferences makes them much less optimistic than their *instability-believing* counterparts, both of which leads them to choose the option security much less often—which in turn leads to a lower share of correct predictions.

Let us turn to the *stable-preference* groups. Here, there is no systematic shift in risk taking over the different investment decisions, and no systematic expectation that others' behaviour will shift in a certain direction. This means that the story from the public-good game carries over: *stability-believers* have a clearer idea of what they think others will do, and thus act accordingly, while *instability-believers* are much more uncertain about what will happen.³¹ This difference then translates into more noise in the *instability-believers'* decision-making.

5 Conclusions

In Fischbacher and Gächter's (2010) important study on cooperation in public-good experiments, only 50% of the participants choose an action that by their elicited preferences would be a best-response to their belief, in the first as well as in the last period.³² Fischbacher et al.'s (2012) analysis seems to suggest that a lot depends on the participants' preference types, with defectors playing a best-

³¹Standard deviations in the market beliefs are 1.3 vs 2.7, $p = 0.002$, Wilcoxon Mann-Whitney test. Note that the same holds true for the *instable-preference* groups, with 1.0 vs 3.0, $p < 0.001$. However, in this case, the effects described above dominate the uncertainty effect.

³²I am grateful to the authors for sharing their data with me.

5 CONCLUSIONS

response much more often than other types. This study shifts the analysis to a different level. What matters is whether participants have stable preferences and whether they in general believe in others' preference stability. If they are 'stability believers' with stable preferences, we can predict their behaviour much better than the 50% above suggests, irrespective of their preference type. While a substantial fraction of these participants are defectors, another substantial fraction are not, and the finding holds up for the non-defectors just as well. Whether my conjectures about the uniform-on-aggregate behaviour of the large fraction of players whose behaviour cannot be predicted are correct is an important question for future research.

In economics, we always thought of preferences as of something stable. In fact, the usefulness of the concept of preferences hinges on them being sufficiently stable. Studies suggesting that the temporal stability of preferences may be limited have therefore started to open the door to a completely new world. This paper is one of the first to look at what is behind this door. However, I have not addressed the question of what it means when I classify somebody as playing a best-response to her beliefs when her elicited preferences are instable. I have not addressed how we should think about a 'best-response', an 'equilibrium', or a 'game' without preference stability, either. For this paper, I have defined all the concepts empirically, to analyse whether these benchmarks help in understanding behaviour. And I argue they do, at least in the public-good experiment: we can predict about half of all choices even of those participants with instable preferences provided they believe in others' preference stability, and even if we use no more than their elicited (instable!) preferences.³³

With one exception, participants who generally believe in others' preference stability are 25 percentage points more likely to play a best-response to their belief about their interaction partner's action, compared to participants who do not believe in others' preference stability. This result might seem odd from a game-theoretic perspective: I should do the best I can do given my belief about the other's action, no matter where that belief comes from. However, if I have a general tendency to be uncertain about what others will do, my belief may be highly volatile also in a specific situation. Then, by the time I am asked to report it, my belief may have changed enough to make my action no longer a best-response to the reported belief. At the same time, my expectation of the other player's contribution will have a higher variance, which is exactly what we observe.³⁴

³³For the risk-context experiment, the benchmark of about 50% best-responses under random play unfortunately prevents us from seeing similar effects.

³⁴The only exception are participants in the risk-context experiment who have instable preferences; for the instability-believers among them, the initial preference measurement tends to be biased towards too much risk-aversion, so that they have a greater tendency to buying the option security in part 3 of the experiment than their initial preference measurement suggests. Given the

5 CONCLUSIONS

Intuitively, we can make sense of most of the findings if we accept the idea that some people still have to discover their true preferences.³⁵ If the discovery process is noisy, many of these people will display ‘inconsistent’ response vectors in the first preference elicitation experiment(s). If another participant sees one of those ‘inconsistent’ response vectors, her belief about the ‘inconsistent’ participant’s action in the simultaneous game will be very uncertain. The same may apply to a participant who has not discovered his true preferences yet and who projects his being in the preference-discovery process onto others by a false-consensus bias. An uncertain belief, in turn, may be less likely to be best-responded to, which makes the corresponding participant’s action harder to predict. The only result that remains surprising once we accept the idea of the preference-discovery process is that a participant’s general belief in others’ preference stability can disrupt the above chain of reasoning. A participant who generally believes in others’ stability will have little uncertainty in her belief about the other group member’s action even if she has not discovered her own preferences yet or if that other group member’s observed response vector is ‘inconsistent’. Documenting the importance of this neglected personality characteristic probably is the most important contribution of this paper.

Technical acknowledgements

The experiments were computerised using z-Tree (Fischbacher, 2007), participants were recruited using ORSEE (Greiner, 2004) with Mozilla Firefox. The equilibria of the game were calculated during the experiment using R (R Development Core Team, 2001, 2012, Ihaka, 1998), which was also used to analyse the data in combination with RKWard (Rödiger et al., 2012) and RStudio (RStudio Team, 2015). For the statistical testing, R packages Exact (Calhoun, 2015, Boschloo test), dgof (Arnold and Emerson, 2011, bootstrapped Kolmogorov-Smirnov test), plm (Croissant and Millo, 2008) and lmtest (Zeileis and Hothorn, 2002, both for the regression with cluster-robust standard errors), texreg (Leifeld, 2013, conversion of regression output to \LaTeX), and doBy (Højsgaard and Halekoh, 2016, calculating groupwise summary statistics) were of particular value. Most of this was done on a computer running on KDE-based (KDE e.V., 2012) Kubuntu, which required the use of wine for the programming of the experiment. The article was written using Kile.

experiment was set up in a way as to make buying the market-security the optimal choice for a majority of the participants, we have a(n insignificantly) higher fraction of best-responses amongst these participants, compared to their stability-believing counterparts.

³⁵Cf., e.g., Plott (1996) or Cubitt et al. (2001). A similar argument can be made for the case that participants have a preference for deliberate randomisation, as, e.g., in Machina (1985).

References

- Arnold, T. A. and Emerson, J. W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal*, 3(2):34–39.
- Bardsley, N. (2000). Control without deception: Individual behaviour in free-riding experiments revisited. *Experimental Economics*, 3(3):215–240.
- Baucells, M. and Villasís, A. (2010). Stability of risk preferences and the reflection effect of prospect theory. *Theory and Decision*, 68(1):193–211.
- Bohnet, I. and Huck, S. (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. *American Economic Review*, 94(2):p362 – 366.
- Brosig, J., Riechmann, T., and Weimann, J. (2007). Selfish in the end? An investigation of consistency and stability of individual behavior. Working paper.
- Bruhin, A., Fehr, E., and Schunk, D. (2016). The many faces of human sociality: Uncovering the distribution and stability of social preferences. Working paper.
- Calhoun, P. (2015). *Exact: Unconditional Exact Test*. R package version 1.6.
- Carlsson, F., Johansson-Stenman, O., and Nam, P. K. (2014). Social preferences are stable over long periods of time. *Journal of Public Economics*, 117:104–114.
- Choi, I. and Nisbett, R. E. (1998). Situational salience and cultural differences in the correspondence bias and the actor-observer bias. *Personality and Social Psychology Bulletin*, 24:949–960.
- Croissant, Y. and Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27(2).
- Cubitt, R. P., Starmer, C., and Sugden, R. (2001). Discovered preferences and the experimental evidence of violations of expected utility theory. *Journal of Economic Methodology*, 8(3):385–414.
- Dal Bó, P. (2005). Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *American Economic Review*, 95(5):1591–1604.
- Duffy, J. and Ochs, J. (2009). Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior*, 66(2):785–812.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

REFERENCES

- Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs and the dynamics of free riding in public good experiments. *American Economic Review*, 100(1):541–556.
- Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404.
- Fischbacher, U., Gächter, S., and Quercia, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology*, 33(4):897 – 913.
- Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology*, 15(1):183–217.
- Gilbert, D. T. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*, chapter Inferential Correction, pages 167–184. Cambridge University Press.
- Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, 112(2):631–645.
- Greiner, B. (2004). An online recruitment system for economic experiments. In Kremer, K. and Macho, V., editors, *Forschung und wissenschaftliches Rechnen 2003*, volume 63, pages 79–93. Ges. für Wiss. Datenverarbeitung, Göttingen.
- Harman, G. (2009). Skepticism about character traits. *Journal of Ethics*, 13:235–242.
- Højsgaard, S. and Halekoh, U. (2016). *doBy: Groupwise Statistics, LSmeans, Linear Contrasts, Utilities*. R package version 4.5-15.
- Hossain, T. and Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, 80(3):984–1001.
- Ihaka, R. (1998). R: Past and future history. In Weisberg, S., editor, *Proceedings of the 30th Symposium on the Interface*, pages 392–396.
- KDE e.V. (2012). About kde. <http://www.kde.org/community/whatiskde/>. Last accessed on 31st January, 2013.
- Leifeld, P. (2013). texreg: Conversion of statistical model output in R to \LaTeX and HTML tables. *Journal of Statistical Software*, 55(8):1–24.
- Loomes, G. and Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, 39(3–4):641–648.

REFERENCES

- Machina, M. J. (1985). Stochastic choice functions generated from deterministic preferences over lotteries. *Economic Journal*, 95(379):575–594.
- Meier, S. and Sprenger, C. D. (2015). Temporal stability of time preferences. *Review of Economics and Statistics*, 97(2):273–286.
- Miyamoto, Y. and Kitayama, S. (2002). Cultural variation in correspondence bias: The critical role of attitude diagnosticity of social constrained behavior. *Journal of Personality and Social Psychology*, 83:1239–1248.
- Muller, L., Sefton, M., Steinberg, R., and Vesterlund, L. (2008). Strategic behavior and learning in repeated voluntary contribution experiments. *Journal of Economic Behavior & Organization*, 67(3–4):782–793.
- Murphy, R. O., Ackermann, K. A., and Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8):771–781.
- Plott, C. R. (1996). *The Rational Foundations of Economic Behaviour*, chapter Rational individual behaviour in markets and social choice processes: the discovered preference hypothesis, pages 225–250. Palgrave MacMillan, Basingstoke.
- R Development Core Team (2001). What is R? *R News*, 1:2–3.
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Rödiger, S., Friedrichsmeier, T., Kapat, P., and Michalke, M. (2012). Rkward: A comprehensive graphical user interface and integrated development environment for statistical analysis with r. *Journal of Statistical Software*, 49(9):1–34.
- RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.
- Sass, M. and Weimann, J. (2012). The dynamics of individual preferences in repeated public good experiments. Working Paper 2/2012, Otto-von-Guericke-Universität Magdeburg.
- Stahl, D. O. (2014). Heterogeneity of ambiguity preferences. *Review of Economics and Statistics*, 96(4):609–617.
- Volk, S., Thöni, C., and Ruigrok, W. (2012). Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior & Organization*, 81(2):664–676.
- Weibull, J. W. (2004). Testing game theory. In Huck, S., editor, *Advances in Understanding Strategic Behaviour: Game Theory, Experiments and Bounded Rationality. Essays in Honour of Werner Güth*, pages 85–104. Palgrave Macmillan.

REFERENCES

- Wolff, I. (2015). When best-replies are not in equilibrium: Understanding cooperative behaviour. Research Paper Series 97, Thurgau Institute of Economics.
- Wolff, I. (forthcoming). What are the equilibria in public-good experiments? *Economics Letters*.
- Zeileis, A. and Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, 2(3):7–10.
-

Appendix

For the Appendix, please refer to
<<http://www.wiwi.uni-konstanz.de/fischbacher/home/staff/dr-irenaeus-wolff/>>.

Appendix

The Appendix is meant for online publication only.

Appendix Instructions (translated)³⁶

I General instructions

General information

You will now participate in an economic experiment. If you read the following explications thoroughly, you can—depending on your choices—earn money. Therefore, it is very important that you read these explications thoroughly.

The instructions you receive from us are for your personal information only. During the experiment, communication is absolutely prohibited. Non-compliance with this rule will lead to exclusion from the experiment and all payments. If you have questions, please raise your hand. We then answer your question at your cubicle.

In this experiment, you will receive money. The amount you receive depends on your decisions and on the decisions of the other experiment participants. Additionally, you receive a compensation of 2 Euros for completing the ensuing questionnaire.

The experiment

The experiment you are participating in today consists of six independent parts. In each of these parts, you will be matched with a different participant. In any case, the participants matched to you will be different people. You will not get to know the identities of the participants you are matched with, neither during nor after the experiment. In the same vein, the participants you are matched with will not get to know your identity.

In some of the parts, there are several participant roles. The role you will take on in actual fact in the different parts will be announced only at the end of the experiment. Therefore, you will make all potentially relevant decisions. Similarly, we will announce only at the very end which of the six parts is relevant for payment. Therefore, you have to determine for all parts what you decide in the

³⁶The German original is available from the author upon request.

according roles. At the end, you will be paid according to the decision you have taken in the relevant role of the randomly-drawn part of the experiment.

Your role and the relevant part are determined by the roll of a die by the participant we have randomly chosen to be the person making the lucky draw at the beginning of the experiment.³⁷ However, we will announce the realisations of the die rolls only at the end of the experiment. Hence you will know only then which of your decisions will be relevant for your payment.

We describe the individual parts directly on the screen. At each point of the experiment, you only receive the description of the according part. We point out to you that your behaviour from one of the earlier parts will possibly be displayed to other participants in a later part. Further, we would like to inform you that the average payoff to be expected from each of the parts is the same.

³⁷The participant making the lucky draw did not take part in the actual experiment and did not get to know anything about it. The participant was merely asked to roll the die three times, record the results on screen as well as on a sheet of paper (the latter was later put up at the wall in the laboratory), and come to the experimenters' room directly afterwards to collect 8 Euros for the faithful completion of the task.

II On-screen instructions (translated)

Part 1 out of 7

Part 1: In this part there are two types of participants, participant A and participant B. Since only participant A has to make a decision, we explain the experiment from participant A's point of view.

In the following, participant A sees 13 situations, one after the other, in which he has to make a decision. In each situation there are 9 options of distributing money between participant A and participant B. Participant A can choose one of the 9 options in each situation.

Please note: Possible payoff distributions are depicted as follows: One line shows the payoff of participant A, the other line shows the payoff of participant B. Which line shows which, is randomly determined and differs between each situation. **So, your own payoffs will sometimes appear in the upper line and sometimes in the bottom line.** By clicking on a distribution option with the left mouse button, you make your choice. After your decision a new screen will appear that you exit with the "next"-button.

At the end of the experiment, the roll of a die will determine which of the 13 situations will be paid out and whether you receive your payment in the role of participant A or participant B.

Situation Number : 1 out of 13

	15	16	18	19	20	21	23	24	25	
The other person receives										The other person receives
You receive	23	21	19	18	16	15	13	12	10	You receive

Part 1: Screenshot of the instruction stage.

Part 2 out of 7

Part 2: In this part of the experiment you form a group of 2 with another randomly assigned participant. Each participant has an endowment of 15 Euros which he can either transfer to his private account or invest in a project. The whole 15 Euros or a fraction of it can be invested in the project. Money that is not invested in the project will automatically be transferred to the private account.

Every Euro invested in the project by a group member yields 1.33 Euros for the group. *Both group members (including the one investing in the project) receive 0.67 Euros of that amount, irrespective of whether and how much a group member invested in the project.* For every Euro that a participant transfers to his private account, he receives 1 Euro. The payoff of the other group member is not affected by that.

Therefore, your payoff is calculated as follows:

Your payoff =
 money transferred to the private account
 + 0.67 * sum of the amounts that you and the other group member invested in the project

This results in the following **situation features:**

- The more is invested in the group project, the bigger is the total sum of payoffs of the two group members.
- However, for every possible investment of the other group member your personal payoff is higher the less you invest in the group project.

Please familiarize yourself with the situation now, by entering in the two boxes different investments you and the other group member might make. Afterwards, click on the button "calculate". For practical reasons, you can only invest 0, 3, 6, 9, 12 or 15 Euros in the project.

Your possible investment:

Possible investment of the other group member:

Your investment	Your payoff	Investment of the other	Payoff of the other

Part 2, PREFS1: instructions; screen with only text (as in upper half) omitted.

Now please answer the following questions:

In case you need a calculator, please click on the corresponding icon below.

1. Assume that no group member invests anything into the project.

How many Euros does each participant receive?

Part 2, PREFS1: instructions; comprehension question 1 (upper part as above).

Now please answer the following questions:

In case you need a calculator, please click on the corresponding icon below.

2. Assume that one group member invests the whole 15 Euros in the project. The other group member doesn't invest anything.

How many Euros does the investing group member receive?

How many Euros does the non-investing group member receive?

Part 2, PREFS1: instructions; comprehension question 2 (upper part as above).

Now please answer the following questions:

In case you need a calculator, please click on the corresponding icon below.

3. Assume that both group members invest the whole 15 Euros in the project.

How many Euros does each participant receive?

Part 2, PREFS1: instructions; comprehension question 3 (upper part as above).

Part 2 out of 7

Like in part 1, there are again two types of participants, A and B, and again you don't know which type you are. One group member will be of type A, the other group member will be of type B. First, participant A makes his investment decision. In a second step, participant B makes his investment decision. However, he can adjust his decision depending on participant A's decision.
For practical reasons, you can only invest 0, 3, 6, 9, 12 or 15 Euros in the project.

You are now taking your decisions as participant A.

Amount that is available to you: 15
Your investment in the project:

Part 2, Prefs1: unconditional-contribution choice.

Part 2 out of 7

You are now taking your decisions as participant B.

In the following table, the rows mark the potential decisions of participant A. Please state how you react to each individual investment decision by participant A (marked in yellow in the following).

Please click on how many Euros you invest into the project. Note: by clicking on an investment, you already take your decision.

Investment by A:

0	For you: 15	For you: 14	For you: 13	For you: 12	For you: 11	For you: 10
	For A: 15	For A: 17	For A: 19	For A: 21	For A: 23	For A: 25
3	For you: 17	For you: 16	For you: 15	For you: 14	For you: 13	For you: 12
	For A: 14	For A: 16	For A: 18	For A: 20	For A: 22	For A: 24
6	For you: 19	For you: 18	For you: 17	For you: 16	For you: 15	For you: 14
	For A: 13	For A: 15	For A: 17	For A: 19	For A: 21	For A: 23
9	For you: 21	For you: 20	For you: 19	For you: 18	For you: 17	For you: 16
	For A: 12	For A: 14	For A: 16	For A: 18	For A: 20	For A: 22
12	For you: 23	For you: 22	For you: 21	For you: 20	For you: 19	For you: 18
	For A: 11	For A: 13	For A: 15	For A: 17	For A: 19	For A: 21
15	For you: 25	For you: 24	For you: 23	For you: 22	For you: 21	For you: 20
	For A: 10	For A: 12	For A: 14	For A: 16	For A: 18	For A: 20

Your investment:

Part 2, Prefs1: conditional-contribution choice (preference elicitation).

Part 3 out of 7

Part 3: Like in the previous parts, you will interact with a new randomly assigned participant also in the 3rd part. The situation is the same as in the 2nd part. Therefore, you can again invest 0, 3, 6, 9, 12 or 15 Euros in a project. Your payoff is the result of:

Your payoff =
money transferred to the private account
+ 0.67*sum of the amounts that you and the other group member invested in the project

This results in the following **situational features**:

The more is invested in the group project, the bigger is the total sum of payoffs of the two group members.

However, for every possible investment of the other group member your personal payoff is higher the less you invest in the group project.

As before, participant A makes his investment decision first. And again, participant B can adjust his decision depending on participant A's decision.

[next](#)

[I need help/have a question](#)

Part 3, PREFS2: instructions.

Part 3 out of 7

Part 3: Like in the previous parts, you will interact with a new randomly assigned participant also in the 3rd part. The situation is the same as in the 2nd part. Therefore, you can again invest 0, 3, 6, 9, 12 or 15 Euros in a project. Your payoff is the result of:

Your payoff =
money transferred to the private account
+ 0.67*sum of the amounts that you and the other group member invested in the project

You are now taking your decisions as participant A.

Amount that is available to you: 15

Your investment in the project:

[OK](#)

[I need help/have a question](#)

Part 3, PREFS2: unconditional-contribution choice.

Part 3 out of 7

You are now taking your decisions as participant B.

Investment by A:

0	For you: 15	For you: 14	For you: 13	For you: 12	For you: 11	For you: 10
	For A: 15	For A: 17	For A: 19	For A: 21	For A: 23	For A: 25
3	For you: 17	For you: 16	For you: 15	For you: 14	For you: 13	For you: 12
	For A: 14	For A: 16	For A: 18	For A: 20	For A: 22	For A: 24
6	For you: 19	For you: 18	For you: 17	For you: 16	For you: 15	For you: 14
	For A: 13	For A: 15	For A: 17	For A: 19	For A: 21	For A: 23
9	For you: 21	For you: 20	For you: 19	For you: 18	For you: 17	For you: 16
	For A: 12	For A: 14	For A: 16	For A: 18	For A: 20	For A: 22
12	For you: 23	For you: 22	For you: 21	For you: 20	For you: 19	For you: 18
	For A: 11	For A: 13	For A: 15	For A: 17	For A: 19	For A: 21
15	For you: 25	For you: 24	For you: 23	For you: 22	For you: 21	For you: 20
	For A: 10	For A: 12	For A: 14	For A: 16	For A: 18	For A: 20

In the following table, the rows again mark the potential decisions of participant A. Please state how you react to each individual investment decision by participant A (marked in yellow in the following).

Please click on how many Euros you invest into the project. Note: by clicking on an investment, you already take your decision.

Your investment:

Part 3, PREFS2: conditional-contribution choice (preference elicitation).

Part 3 out of 7

You just stated how you react to every possible investment decision by participant A. Now we want you to assess how participant A behaved. In the following, you have to state the probability with which participant A chose each investment decision.

In part 3, participant A had to choose between 6 different investment levels. So, for each of these investment levels you have to state the probability with which participant A chose this exact level. Doing so, you can earn an additional 2.00 Euros. Your chance of earning 2.00 Euros increases with the accuracy of your assessment. Your assessment is the better, the closer it matches the actual behavior of participant A. By clicking on the button "details", you can see how we calculate the payoff that rewards the accuracy of your assessment. If you are not interested in the details, you can safely ignore the corresponding explanation.

What is important for you is just that the chance of getting a high payoff is maximized when you correctly assess the choice behavior of participant A. We want you to have an incentive to properly think about the behavior of participant A and to be rewarded if you capture it well and state it accordingly.

You can state your assessment of the probabilities with the help of a bar chart. Below you see an example of such a chart. If you click on the chart, you will notice that you can adjust the height of the bars by clicking on them. At the same time, in the upper right of the diagram a window will open up with which you can define the exact height of the bar. The sum of the bars of the probability distribution always has to equal 100 percent. If you click on "scale", the bars adjust accordingly, while the relative size ratio remains the same. If, by using exact entries, the sum is already 100 percent, the bars are already correctly scaled. Please note that the percent values above the bars are rounded to integers, so that the sum can differ slightly from 100.

How likely is it that participant A chose an investment of X?

Example chart:

When you click on "next", such a diagram will appear on your screen. Please note that this diagram can be relevant for your payoff. Please work on the diagram on the next page. You can return to these instructions at any time. Afterwards, the 3rd part of the experiment is completed.

Part 3, PREFS2: instructions for the belief-elicitation on A's choice (training for the SIMPGBELIEFS and STABILITYBELIEFS experiments).

Part 3 out of 7

Payoff of the probability assessment:

In case this part will be relevant for your payoff, you will receive an additional payoff of 2 Euros with a certain probability p and no additional payoff with probability $1-p$. Your assessments affect your chance p in the following way:

First, your chance p is set to 100%. Then there will be deductions for wrong assessments. The deviations are first divided by 100, then squared, then halved and then subtracted from 1. The result is multiplied by 100 to get a percentage between 0 and 100.

If you, for example, assigned 70% to a certain investment, but participant A chose a different investment level, your deviation is 0.7. So $0.70 * 0.70 / 2 = 0.245$ (approx. 25%) is subtracted from p .

For the investment level that participant A actually chose, it is bad if your assessment is far away from 100%. That deviation will also be squared, halved and subtracted. If you, for example, assigned only 60% to the correct investment, $(1 - 0.60) * (1 - 0.60) / 2 = 0.40 * 0.40 / 2 = 0.08$ (8%) will be subtracted from p .

In the end, all the subtractions are added up. The smaller these squared deviations are, the better was your assessment. For those who are interested: Here's the mathematical formula that is used to calculate the quality of your assessment and thereby the chance p to earn 2.00 Euros:

$$p = (1 - 0.5 * (\text{sum of the squared probabilities on investment levels that were not chosen by participant A} + \text{squared counter-probability of the actually chosen investment level})) * 100$$

The computer program will calculate the quality of your assessment p and show it to you. Your assessment was the better and the chance to earn additional 2.00 Euros in this part is the higher, the higher your value of p is. If part 3 is relevant for your payoff, a number between 0 and 100 will be determined by rolling dice at the end of the experiment. If this number is smaller or equal to p , you receive 2.00 additional Euros. If the number is larger than p , you won't receive an additional payoff.

Summary:

To have a good chance of receiving the additional payoff, it should be your goal to get as few subtractions from p as possible. The best way to achieve that is to make a good assessment of the choice behavior of participant A and to state it truthfully.

[Hide details](#)

[I need help/have a question](#)

Part 3, PREFS2: instructions for the belief-elicitation on A's choice, details (training for the SIMPGBELIEFS and STABILITYBELIEFS experiments).

You just stated how you react to every possible investment decision by participant A. Now we want you to assess how participant A behaved. In the following, you have to state the probability with which participant A chose each investment decision.

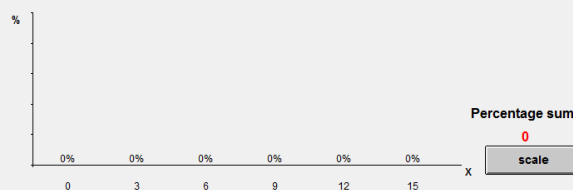
In part 3 participant A had to choose between 6 different investment levels. So, for each of these investment levels you have to state the probability with which participant A chose this exact level. Doing so, you can earn additional 2.00 Euros. Your chance of earning 2.00 Euros increases with the accuracy of your assessment. Your assessment is all the better, the more it matches the actual behavior of participant A. By clicking on the button "details" you can see the calculation of the payoff that rewards the accuracy of your assessment. If you are not interested in the details, you can ignore the corresponding explanation.

What is important for you is just that the chance of getting a high payoff is maximized when you correctly assess the decision behavior of participant A. We want you to have an incentive to properly think about the behavior of participant A and to be rewarded if you capture it well and state it accordingly.

You can state your assessment of the probabilities with the help of a bar chart. Below you see an example of such a chart. If you click on the chart, you will notice that you can adjust the height of the bars by clicking on them. At the same time in the upper right of the diagram a window will open up with which you can define the exact height of the bar. The sum of the bars of the probability distribution always has to equal 100 percent. If you click on "scale", the bars adjust accordingly, while the relative size ratio stays the same. If, by using exact entries, the sum is already 100 percent, the bars are already correctly scaled. Please note that the percent values above the bars are rounded to integers, so the sum can slightly differ from 100.

[back to instructions](#)

How likely is it that participant A chose an investment of X?



[next](#)

Part 3, PREFS2: belief-elicitation on A's choice (training for the SIMPGBELIEFS and STABILITYBELIEFS experiments).

Part 4 out of 7

Reaction of your newly assigned group member as B in part 2:

Investment of A:	0	3	6	9	12	15
0	For B: 15 For A: 15	For B: 14 For A: 17	For B: 13 For A: 19	For B: 12 For A: 21	For B: 11 For A: 23	For B: 10 For A: 25
3	For B: 17 For A: 14	For B: 16 For A: 16	For B: 15 For A: 18	For B: 14 For A: 20	For B: 13 For A: 22	For B: 12 For A: 24
6	For B: 19 For A: 13	For B: 18 For A: 15	For B: 17 For A: 17	For B: 16 For A: 19	For B: 15 For A: 21	For B: 14 For A: 23
9	For B: 21 For A: 12	For B: 20 For A: 14	For B: 19 For A: 16	For B: 18 For A: 18	For B: 17 For A: 20	For B: 16 For A: 22
12	For B: 23 For A: 11	For B: 22 For A: 13	For B: 21 For A: 15	For B: 20 For A: 17	For B: 19 For A: 19	For B: 18 For A: 21
15	For B: 25 For A: 10	For B: 24 For A: 12	For B: 23 For A: 14	For B: 22 For A: 16	For B: 21 For A: 18	For B: 20 For A: 20

You now make your decision.

Amount that is available to you: 15

Your investment in the project:

Part 4, SIMPG: contribution choice; screen with only instructions (as in upper half) omitted.

Part 5 out of 7

Reaction of the other participant as B in part 2:

Investment of A:	0	3	6	9	12	15
0	For B: 15 For A: 15	For B: 14 For A: 17	For B: 13 For A: 19	For B: 12 For A: 21	For B: 11 For A: 23	For B: 10 For A: 25
3	For B: 17 For A: 14	For B: 16 For A: 16	For B: 15 For A: 18	For B: 14 For A: 20	For B: 13 For A: 22	For B: 12 For A: 24
6	For B: 19 For A: 13	For B: 18 For A: 15	For B: 17 For A: 17	For B: 16 For A: 19	For B: 15 For A: 21	For B: 14 For A: 23
9	For B: 21 For A: 12	For B: 20 For A: 14	For B: 19 For A: 16	For B: 18 For A: 18	For B: 17 For A: 20	For B: 16 For A: 22
12	For B: 23 For A: 11	For B: 22 For A: 13	For B: 21 For A: 15	For B: 20 For A: 17	For B: 19 For A: 19	For B: 18 For A: 21
15	For B: 25 For A: 10	For B: 24 For A: 12	For B: 23 For A: 14	For B: 22 For A: 16	For B: 21 For A: 18	For B: 20 For A: 20

How likely is it that the other participant chose an investment of X in part 4?

Percentage sum: 0

Part 5, SIMPGBELIEFS: belief elicitation; screen with only instructions (as in upper half) omitted. Clicking on “details” led to an analogous screen as in Part 3.

Part 6 out of 7

Part 6: In part 2 and part 3 you and the other participants had to make an investment decision, where you, as participant B, could adjust your decision depending on the decisions of your randomly assigned participant A. In part 6 we want you to assess the behavior of four other type-B-participants in part 3. In the following, you have to state with which probability these participants made their investment decisions, depending on the behavior of the corresponding type-A-participants. For assistance, you can see the behavior of each participant in the role of participant B in part 2.

Each of the four B-participants had to state how they would react to each of the 6 possible investment decisions of the corresponding A-participant. So, for each of the 6 possible investment decisions you have to make an assessment of the probability with which the respective B-participant chose a certain reaction. Therefore, this time you see 6 diagrams for each of the four B-participants.

Like in the previous parts, you can earn money with your assessment of the behavior of the other participants. For each assessment you receive either 6.00 Euros or nothing. The chance to earn 6.00 Euros increases with the accuracy of your assessment. First, for each of the four B-participants one of the 6 diagrams is randomly drawn. The chance of earning 6.00 Euros for that randomly drawn diagram of the respective B-participant is then calculated in the same way the chance of earning the extra amount of 20.00 Euros in part 5 was calculated. Again, you can see the details of the calculation of your chance to earn 6.00 Euros each, by clicking on "details". If you are not interested in the details, you can ignore the corresponding explanation.

What is important for you to know is that the chance of getting a high payoff is maximized when you correctly assess the decision behavior of the other participants. We want you to have an incentive to properly think about the behavior of the B-participants and to be rewarded if you capture it well and state it accordingly.

details

next

I need help/have a question

Part 6, STABILITYBELIEFS: instructions; "details" led to an analogous screen as in Part 3.

Part 6 out of 7

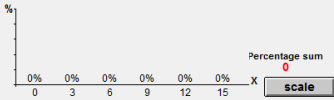
On the right, you see the behavior as participant B in part 2 of the first participant for whom you have to make a guess on the behavior in part 3.

Instructions

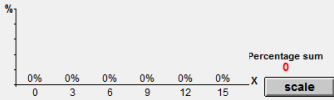
Reaction of the respective participant as B in part 2:

Investment of A:	0	3	6	9	12	15
0	For B: 15 For A: 15	For B: 14 For A: 17	For B: 13 For A: 19	For B: 12 For A: 21	For B: 11 For A: 23	For B: 10 For A: 25
3	For B: 17 For A: 14	For B: 16 For A: 16	For B: 15 For A: 18	For B: 14 For A: 20	For B: 13 For A: 22	For B: 12 For A: 24
6	For B: 19 For A: 13	For B: 18 For A: 15	For B: 17 For A: 17	For B: 16 For A: 19	For B: 15 For A: 21	For B: 14 For A: 23
9	For B: 21 For A: 12	For B: 20 For A: 14	For B: 19 For A: 16	For B: 18 For A: 18	For B: 17 For A: 20	For B: 16 For A: 22
12	For B: 23 For A: 11	For B: 22 For A: 13	For B: 21 For A: 15	For B: 20 For A: 17	For B: 19 For A: 19	For B: 18 For A: 21
15	For B: 25 For A: 10	For B: 24 For A: 12	For B: 23 For A: 14	For B: 22 For A: 16	For B: 21 For A: 18	For B: 20 For A: 20

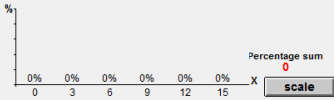
How likely is it that the participant's reaction on an investment of 0 was X in part 3?



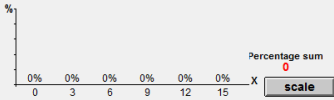
How likely is it that the participant's reaction on an investment of 3 was X in part 3?



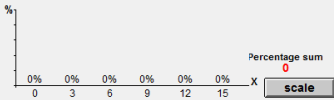
How likely is it that the participant's reaction on an investment of 6 was X in part 3?



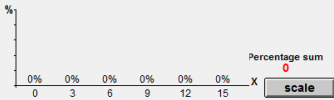
How likely is it that the participant's reaction on an investment of 9 was X in part 3?



How likely is it that the participant's reaction on an investment of 12 was X in part 3?



How likely is it that the participant's reaction on an investment of 15 was X in part 3?



next

I need help/have a question

Part 6, STABILITYBELIEFS: belief elicitation.

Part 7 out of 7

Part 7: In this part of the experiment you once more form a group of 2 with another randomly assigned participant. The situation is a mixture of parts 2, 3 and 4. Participant A and participant B make their decisions successively, whereby participant B can make his decision dependent on participant A's decision. Before you make your decision as participant A, you learn about how the participant that was randomly assigned to you behaved as participant B in part 2. You will find this information in a table on the right. Participant B, however, doesn't learn about your behavior in part 2.

Keep in mind: The participants make their decisions successively, so participant B can make his decision depending on the decision of participant A.

[Complete situation description](#)

Reaction of the newly assigned participant as B in part 3:

	0	3	6	9	12	15
0	For B: 15 For A: 15	For B: 14 For A: 17	For B: 13 For A: 19	For B: 12 For A: 21	For B: 11 For A: 23	For B: 10 For A: 25
3	For B: 17 For A: 14	For B: 16 For A: 16	For B: 15 For A: 18	For B: 14 For A: 20	For B: 13 For A: 22	For B: 12 For A: 24
6	For B: 19 For A: 13	For B: 18 For A: 15	For B: 17 For A: 17	For B: 16 For A: 19	For B: 15 For A: 21	For B: 14 For A: 23
9	For B: 21 For A: 12	For B: 20 For A: 14	For B: 19 For A: 16	For B: 18 For A: 18	For B: 17 For A: 20	For B: 16 For A: 22
12	For B: 23 For A: 11	For B: 22 For A: 13	For B: 21 For A: 15	For B: 20 For A: 17	For B: 19 For A: 19	For B: 18 For A: 21
15	For B: 25 For A: 10	For B: 24 For A: 12	For B: 23 For A: 14	For B: 22 For A: 16	For B: 21 For A: 18	For B: 20 For A: 20

You are now making your decision as participant A.

Amount that is available to you: 15

Your investment in the project:

[OK](#)

[I need help/have a question](#)

Part 7, PREFS3: unconditional contribution; "Complete situation description" led to a screen similar to the instructions screen in Part 3.

Part 7 out of 7

You are now making your decisions as participant B.

Have in mind that these decisions could be relevant for your payoff!

Investment of A:

0	For you: 15 For A: 15	For you: 14 For A: 17	For you: 13 For A: 19	For you: 12 For A: 21	For you: 11 For A: 23	For you: 10 For A: 25
3	For you: 17 For A: 14	For you: 16 For A: 16	For you: 15 For A: 18	For you: 14 For A: 20	For you: 13 For A: 22	For you: 12 For A: 24
6	For you: 19 For A: 13	For you: 18 For A: 15	For you: 17 For A: 17	For you: 16 For A: 19	For you: 15 For A: 21	For you: 14 For A: 23
9	For you: 21 For A: 12	For you: 20 For A: 14	For you: 19 For A: 16	For you: 18 For A: 18	For you: 17 For A: 20	For you: 16 For A: 22
12	For you: 23 For A: 11	For you: 22 For A: 13	For you: 21 For A: 15	For you: 20 For A: 17	For you: 19 For A: 19	For you: 18 For A: 21
15	For you: 25 For A: 10	For you: 24 For A: 12	For you: 23 For A: 14	For you: 22 For A: 16	For you: 21 For A: 18	For you: 20 For A: 20

In the following table, the rows mark the potential decisions of participant A. Please state how you react to each individual investment decision by participant A (marked in yellow in the following).

Now please click on how many Euros you invest into the project.

Your investment: 0 3 6 9 12 15

[I need help/have a question](#)

Part 7, PREFS3: conditional contribution (preference elicitation).