Do Cognitive Processes Influence Social Preferences?

Testing the social heuristics hypothesis in a sequential prisoner's dilemma

Kjetil Røiseland Madland

Master's thesis

Thesis submitted in fulfilment of the degree

Master of Economics

Department of Economics, University of Bergen

June 2017

UNIVERSITETET I BERGEN

Acknowledgements

I am indebted my advisor, Eirik Andrè Strømland, for outstanding guidance in the process of writing this thesis, and for organizing the experiment in the Norwegian Citizen Panel. His input and comments have been invaluable to the progress of my thesis, and to the final result. I must also thank Amanda Reigstad, Alexander Daoud and Sidra Choudhary for the cooperation in conducting the lab experiment, and Erla Katrine Løvseth for help with the response time data from the Norwegian Citizen Panel. Finally, I would like to thank Ada, for all her patience and helpful comments.

Bergen, June 2017

Kjetil Røiseland Madland

Abstract

Is cooperation intuitive or deliberative? From an early notion of cooperation as a deliberate suppression of innate selfish preferences, a growing body of literature has turned the general perception towards prosocial behaviour as something intuitive, sometimes actively oppressed for the sake of selfish needs and wishes. If the dual-process framework from psychology gives a better description of decision making than do the classical economic models, this will have important implications for many economic models. Testing the social heuristics hypothesis through a sequential prisoner's dilemma conducted both in the lab and by an online survey, I find no conclusive evidence that increased deliberation systematically changes willingness to cooperate with strangers. This is the first study (to my knowledge) to isolate the effect of a manipulation through preferences. The results hold for both a general regression of cooperation on the deliberation treatment, and for the main analysis, with separate effects through preferences and beliefs.

All calculations and regressions are performed, and all tables and graphs are made, in Stata edition 14.2.

Content

List of tables

List of figures

1 Introduction and literature

The present thesis experimentally investigates the stability of preferences; specifically, preferences that determine our willingness to cooperate with strangers in social dilemmas. The thesis adds to the literature by investigating the assumed conflict between intuitive and reflective thinking in new datasets from both a lab experiment and from the Norwegian Citizen Panel (NCP). I test the predictions of the social heuristics hypothesis (SHH); that increased deliberation will reduce cooperation (Rand et al., 2012, Rand et al., 2014, Bear and Rand, 2016).[1] What makes the present research novel, is the attempt to try and assess whether such effects of cognitive processes, if present, work through preferences or beliefs. Evidence from a study on framing in social dilemmas indicate that social framing effects work through beliefs rather than preferences (Ellingsen et al., 2012).

In a one-shot, sequential prisoner's dilemma, I test whether increased deliberation changes the willingness to cooperate with strangers. Deliberation is promoted by asking half of the subjects to give a written justification prior to making the decision to either cooperate or defect. Holding beliefs constant for the group of second movers, the effect through preferences is isolated. If the observed level of cooperation changes due to increased reflection, this could indicate that preferences are not stable.

In microeconomic theory, we assume that agents behave rationally and that their tastes are complete and transitive. These are axiomatic assumptions, critical to the logical completeness of microeconomic theory and the models therein. Moreover, it is common to assume that agents have a single set of preferences that are stable with respect to time and situation (Stigler and Becker, 1977, Mas-Colell et al., 1995, Chapter 1). The stability of preferences is not a logical necessity, but critical to the predictive power of models attempting to explain social preferences and behaviour, e.g. Rabin (1993) and Fehr and Schmidt (1999).

In a sequential game, beliefs are held constant for the second mover, allowing the separation of the two possible channels responsible for a treatment effect. This is done both in the lab experiment and in the NCP survey. The lab experiment serves as a pilot for the larger NCP dataset, but also adds valuable observation data to the pooled dataset analysed in section 3.3. The NCP is an online survey and experiment tool similar to Amazon Mechanical Turk, but with

[1] The theory, and its predictions, will be discussed further in section 1.2.

an attempted representative selection of the Norwegian population. Participants in the NCP have been chosen at random from the National registry to receive an invite.

In a social dilemma, here understood as a non-zero-sum game, players may pay a personal cost to cooperate to benefit the common good. The collectively rational decision, giving the highest total payoff, is for everyone to cooperate. The personally payoff-maximizing decision, in terms of money, is to defect and hope that everyone else contributes (Dawes, 1980).[2] A much used social dilemma is a public good game (PGG) in which each player's contribution is doubled and then divided between four players (see e.g. Ledyard (1995) and Cone and Rand (2014)). Since each contribution is doubled, contributing is clearly socially optimal. Yet, the personal marginal benefit of a player's contribution is less than the personal marginal cost, so the rational decision for a selfish individual is not to contribute. Thus, the unique Nash-equilibrium is that nobody cooperates.

It is a well-known fact that levels of cooperation in social dilemma games tend to well exceed the levels predicted by economic theory for a selfish agent; the *homo economicus*. Typically, the observed levels of cooperation in prisoner's dilemma games are in the range of 40-60% (Ledyard, 1995, Sally, 1995, Chaudhuri, 2011). Similar results are also found for one-shot PGGs (Dawes and Thaler, 1988). Economists have long struggled to properly explain these high numbers, and to come up with models that accurately predict this sort of behaviour (see e.g. Becker (1976)).

One experimentally motivated explanation for seemingly altruistic behaviour, is that people are concerned not only with their own payoff, but also with the fairness of the outcome. Rabin (1993) models this in a framework of reciprocity, introducing three stylized facts: (i) People are willing to sacrifice some material payoff to be nice to those who are nice to them. (ii) People are willing to sacrifice some material payoff to hurt those who hurt them. (iii) Stylized facts (i) and (ii) have greater impact as material payoff gets smaller. If preferences and/or beliefs are not stable, this poses problems for Rabin's model when used to predict behaviour.

Fehr and Schmidt (1999) propose a different explanation for cooperation in social dilemmas. They show that, if some agents care about equality, equilibria may occur where it is rational to cooperate even for selfish individuals. A weakness with this model, as pointed out by Binmore and Shaked (2010), is that the number of free parameters makes it flexible to the point where it

[2] Such social dilemmas include varieties of public goods games (PGG), the trust game and prisoner's dilemma. The dictator game has some similar attributes, but is conventionally a zero-sum game.

becomes tautological. Thus, the model may well explain behaviour, but is less suited to make out-of-sample predictions about behaviour if preferences are not stable.

How do we model and predict behaviour if preferences are unstable? The notion of dual-process thinking has been popular in the academic literature of psychology and economics for some decades (see e.g. Kahneman et al. (1973), Evans et al. (1983), Chaiken and Trope (1999), Evans (2003), Kahneman (2003)). In recent years, many studies have indicated that there is indeed a difference between intuitive and deliberative decision making, and that cooperation seems to be the intuitive response in a social dilemma, whereas reflective decision making is more selfish (Rand et al., 2012, Zaki and Mitchell, 2013, Rand et al., 2014, Bear and Rand, 2016, Cappelen et al., 2016, Rand, 2016, Strømland et al., 2016). Other studies find no evidence suggesting that intuition is more cooperative than deliberation (Verkoeijen and Bouwmeester, 2014, Hauge et al., 2016, Bouwmeester et al., 2017).

The overall explanation for cooperation in social dilemmas is beyond the scope of the present text. Instead, it considers the role of cognitive modes and processes in forming preferences and beliefs. If beliefs and preferences are sensitive to cognitive processes or changes in cognitive mode, then behavioural models, like the ones presented by Rabin (1993) and Fehr and Schmidt (1999), will give inconsistent predictions. As will the standard rational agent-model. Indeed, all behavioural models that hinge on the assumption that preferences are stable, will give inconsistent predictions.[3] Within a dual-process framework, the present thesis investigates the stability of preferences and the possible conflict between intuitive and deliberative decision making, and its implications for behavioural models.

1.1 Dual-process theory

Dual-process theory is an umbrella term for a set of theories stating that the human mind consists of two distinct cognitive subsystems. For reference, I use the framework presented by psychologist and Nobel prize laureate in Economics, Daniel Kahneman. He describes the dual-process thinking as two opposing systems of our cognition, simply called system one and system two (Kahneman et al., 1973, Kahneman, 2003, Kahneman, 2011). These are competing systems, both with their unique strengths and weaknesses.

In short, system one is the intuitive system; quickly making decisions based on heuristics and simplifications (Kahneman et al., 1973, Evans et al., 1983, Evans, 2003, Kahneman, 2011). It

[3] To my knowledge, all behavioural models within economics hinge on the assumption that preferences are stable.

works fast, and (almost) effortlessly comes up with answers. When faced with a difficult question or problem, system one might come up with an answer to an entirely different question, but an answer nonetheless (Kahneman, 2011).

System two is slow and demands more effort than system one. But it is also much better at evaluating all available information, and importantly, separating important information from the trivial (Evans, 2003). Being costly to use, system two is lazy, and gladly lets system one do all the work, unless specifically commanded to take control (Kahneman, 2011, p. 35). This calls for deliberate reflection, which is the core of the present thesis.

Dual-process theory is different from standard microeconomic theory in some important respects. Notably, in neoclassical microeconomics, few assumptions are made about the actual processes and tastes involved in individual decision making. Instead, it is assumed that people behave *as if* they maximize some underlying objective function. The inputs of such functions are taken as given and are usually not delved any further into (Stigler and Becker, 1977). However, to many economic models (e.g. Rabin (1993) and Fehr and Schmidt (1999)) the stability of preferences and beliefs is important. Formally, most (if not all) behavioural models assume that tastes are not only complete and transitive, but also stable; invariant with regard to time and place.

In the present thesis, I investigate whether preferences are invariant with respect to cognitive mode. Is there, as Kahneman and many with him suggest (Rand et al., 2012, Zaki and Mitchell, 2013, Cone and Rand, 2014, Evans et al., 2015, Bear and Rand, 2016), an internal conflict between intuitive and deliberative thinking (system one and system two)? And if so, through what mechanisms does such a conflict materialize in a social dilemma?

A growing body of literature suggests there is indeed an internal conflict between intuitive and reflective thinking. Many studies find that, in social dilemmas, intuition favours cooperation rather than defection (Rand et al., 2012, Zaki and Mitchell, 2013, Cone and Rand, 2014, Rand et al., 2014, Bear and Rand, 2016, Cappelen et al., 2016, Rand, 2016, Strømland et al., 2016). This conclusion has been challenged both on technical grounds (Tinghög et al., 2013, Myrseth and Wollbrant, 2015, Myrseth and Wollbrant, 2016), and by similar studies finding no effects, or inconclusive evidence (Verkoeijen and Bouwmeester, 2014, Hauge et al., 2016, Bouwmeester et al., 2017). Nevertheless, considering the amount of evidence suggesting prosocial behaviour is intuitive, this serves as my working hypothesis. Specifically, I base the analysis primarily on the social heuristics hypothesis.

1.2 The social heuristics hypothesis

A theory which has gained much ground lately, is the social heuristics hypothesis (SHH), presented by Rand, Greene & Nowak (2012), and refined by Rand et al. (2014) and Bear and Rand (2016). The general idea is that people learn, through socialization and everyday activities, that cooperation with others tends to be beneficial; even to the extent that it will often be worth paying a personal cost to promote the common good. Internalizing the idea that cooperation is beneficial, prosocial behaviour develops as the intuitive, default choice in a social dilemma (Capraro et al., 2014, Rand et al., 2014, Bear and Rand, 2016). When faced with new and unknown surroundings and situations, such as a lab experiment, those who have internalized the idea that cooperation is beneficial, tend to use this as a heuristic for making decisions.

If a share of the population have these social heuristics, one would expect a treatment promoting intuition to increase cooperation (Rand et al., 2014). Conversely, increased deliberation and reflection should decrease cooperation. A one-shot economic game will seem unnatural to most, as most people rarely encounter real one-shot situations in everyday life. The SHH predicts that individuals will draw on former experience and norms, and extrapolate these to the unfamiliar situations of a lab environment or a one-shot game.

This out-of-context use of social heuristics, is predicted only for individuals who are used to an environment where prosocial behaviour has shown to be beneficial, and who are unfamiliar with the lab environment. For individuals who have experience from the lab setting or one-shot games, or are in other ways used to being rewarded for anti-social behaviour, the SHH predicts no effect of manipulating cognitive mode. Importantly, according to the SHH, intuition should never lower cooperation (Rand et al., 2014). The SHH claims that intuition promotes cooperation though preferences (Rand et al., 2012). There are, however, also theories emphasizing other mechanisms than social heuristics. Some of these mechanisms might explain how increased deliberation could also give more cooperation in social dilemmas.

1.3 Reactive egoism and moral deliberation

In the present study, we use a treatment to promote deliberation, rather than intuition. One possible side effect of increased deliberation is perspective taking. When taking the time to deliberate a decision, many subjects will presumably think further ahead, and form empirical expectations. In a game including other players, deliberation could make subjects form a clearer idea of what they expect the others to do. Torsvik et al. (2011) find that announced post-play face-to-face discussion decreases contributions to a common good. These results are interpreted as an effect of perspective taking and empirical expectations. This closely relates to the present study, where subjects are given a deliberation treatment.

There is a large body of literature advocating the positive effects of perspective taking in social interactions and bargaining, e.g. Neale & Bazerman (1983), Batson (1995), Galinsky & Moskowitz (2000) and Savitsky et al. (2005). The general idea is that taking the perspective of others help diminish egocentric biases in judgement (Epley et al., 2006). A common problem in social interactions and bargaining, is that people are egocentrically biased in their assessment of fairness, and feel they deserve more than their fair share of resources (Babcock and Loewenstein, 1997, Paese and Yonker, 2001). Therefore, it has long been a commonly held belief that perspective taking is beneficial in bargaining situations through helping promote prosocial behaviour and reducing egocentric bias (Epley et al., 2006).

In a study of several social interactions, Epley et al. (2006) find evidence suggesting that taking the perspective of the other players, might actually increase taking in a take some-game.[4] Similar results are found in the other games as well, bringing about a theory of *reactive egoism* (Epley et al., 2006). The hypothesis is that agents in a social dilemma will often, when uncertain about the choices of others, think badly of them, and react on this assumed selfishness with selfish behaviour. In other words: when taking another's perspective, people tend to fear that the other will behave selfishly, and counter this assumed selfishness by behaving more selfishly themselves.

Reactive egoism closely relates to the results found by Torsvik et al. (2011): In addition to the prospect of having to defend their (lack of) contribution, the participants might have also considered what the others would have done. The authors interpret the decrease in cooperation as a response to the empirical expectations imposed by the announcement of post-play face-to-face discussion.

Subjects in the present study are not specifically asked to take the perspective of the other player, as is the treatment in the study by Epley et al. (2006). Still, it is plausible that the deliberation treatment could induce some perspective taking.[5] Consider a sequential prisoner's

[4] One of the scenarios studied was a classic tragedy of the commons, where perspective taking led the subjects to take even more of the common resource for themselves, even though the share they deemed was fair for them to take decreased with the perspective taking.
[5] Text data from the NCP responses, although not paired with subject IDs, suggests this applies at least for some of the subjects.

dilemma; the game in question in the present thesis. For player 1 (i.e. the first mover), the action of player 2 (i.e. the second mover) is unknown, and of paramount importance to his own monetary payoff. This is obviously something to reflect upon. For player 2, there is little, if any, room for reflection concerning player 1's choice, since player 2 has full knowledge about this when making her decision.

How will the deliberation treatment affect player 2? Because player 2 knows what player 1 has chosen, the treatment can only work through player 2's preferences. It is possible that increased deliberation will strengthen moral sentiments, and thus enforce norm compliance such as reciprocity (Rabin, 1993, Falk and Fischbacher, 2006), despite reduced personal payoff in terms of money. Through such mechanisms, a deliberation treatment might increase cooperation in the second stage of a sequential prisoner's dilemma, contrary to the prediction of the SHH. An overwhelming share of the subjects in the treatment groups in NCP mention either "fairness" or "equality" in their written response.[6]

If we believe that the human mind works in the manner proposed by dual-process theory, with two distinct sets of preferences, then we need to specifically model the choice of preference set in any given decision-making situation. Otherwise, we will be unable to make predictions about individual decision making.

[6] Although we cannot pair the written justifications with the actual choices made.

2 Experimental design

2.1 Theoretical framework

Consider an individual with two distinct sets of preferences; one intuitive, and the other deliberative. Let the two sets be represented by the utility functions $\mathrm{U}_i = \mathrm{U}_i(\mathbf{X})$ and $\mathrm{U}_d = \mathrm{U}_d(\mathbf{X})$ respectively, where $\mathbf{X}$ is a commodity vector. Except for the conventional assumptions of completeness, transitivity and continuity, no additional assumptions about the functions are needed. As is custom in microeconomic theory, we assume that the agent attempts to maximize the value of her chosen utility function, or at least acts *as if* she does. Within the framework proposed by the SHH, the deliberative preferences, represented by utility function U_d, is assumed to be more selfish than the intuitive preferences (i.e. U_i).

For any given situation, assume that agent j maximizes either utility U_{ij} or U_{dj} with probabilities γ_j and $(1 - \gamma_j)$ respectively. γ_j is the individual specific probability of making an intuitive decision, while $(1 - \gamma_j)$ is the probability of deliberating. A high γ_j means that agent j has a high probability of making an intuitive decision, when faced with a problem, whereas a low γ_j corresponds to a high probability of deliberation.

Whether the same γ_j can be assumed to apply only in social dilemmas or more generally, is beyond the scope of this text.[7] γ_j is likely to be a function of time, place, the type of decision, and many other factors (Bear and Rand, 2016), which can be hard to control for. However, Rubinstein (2016) shows that there is a strong correlation between individual relative response times over a set of different economic games.[8] Assuming relative response time is a valid proxy for relative deliberation, we can assign each agent as a type, corresponding to the probability that he will make an intuitive decision.

The problem posed by the SHH is that experience with the lab environment is predicted to alter an individual's intuitive preferences, rather than the probability of deliberating. The same goes for individuals who are familiar with an environment which promotes selfishness and competitiveness rather than cooperation. These individuals are predicted to be intuitively more prone to making selfish decisions; which is formally distinct from being more prone to deliberating as opposed to making intuitive decisions. SHH predicts that certain factors make

[7] Although, for applied research, it is an important question.
[8] For a more thorough presentation of player typology, see Rubinstein (2016).

the intuitive preferences more similar to the deliberative preferences; that is $U_{ij} \rightarrow U_{dj}$ as experience increases (Rand et al., 2014, Bear and Rand, 2016).

I propose a different theoretical mechanism, which is to let experimental experience work through the probability of deliberating. As an alternative interpretation of the SHH, imagine that experimental experience and socialization in competitive environments decrease γ_j; the probability of making an intuitive decision rather than deliberating. As individuals get more accustomed to the competitive setting of the lab, γ_j tends towards 0, and the probability of deliberation towards 1. This alternative interpretation, with a typology of players where γ_j is the individual type, is valid if, and only if, U_{dj} is inherently more selfish than U_{ij} in a social dilemma. This is what the SHH assumes (Rand et al., 2014, Bear and Rand, 2016). If this is indeed the case, changes in intuitive preferences may well be treated as a change in the probability of deliberating. This approach is also more flexible and easier to apply. Although formally distinct, the two methods yield the same result.[9]

Setting an agent's type equal to the probability that the given agent will make an intuitive decision, we say that agent j is of type γ_j. With no information about the situation for the decision, the best prediction for agent $j's$ behaviour is the behaviour which maximizes the weighted sum

$$\gamma_j max_{\boldsymbol{X}} U_i(\boldsymbol{X}) + (1 - \gamma_j) max_{\boldsymbol{X}} U_d(\boldsymbol{X})$$

Making specific predictions about individual choices would require knowing the specifics of the agent's utility functions, as well as knowing the agent's type, both of which are beyond the scope of this thesis. This presentation is meant merely as an illustration of how dual-process cognition would affect microeconomic models, and a simple example of what a remedy might look like.

2.2 Game setup

In this thesis, I use data from a sequential prisoner's dilemma. The same game has been run both as a pencil-and-paper experiment in a physical lab with N=187 respondents, and in the Norwegian Citizen Panel (NCP) with N=1173 respondents. The main objective of the experiment is to uncover how, if at all, cognitive modes and processes might affect preferences.

[9] Varying only the probability of deliberating is arguably preferable to changing the utility functions. Keeping the utility functions flexible would give so many free parameters that the model would become tautological (Binmore & Shaked, 2010).

Specifically, we want to know whether cooperation with strangers in a prisoner's dilemma is intuitive, deliberative or alternatively that there is no internal conflict when deciding on whether to cooperate or to defect.

Research by Rand et al. (2012), Rand et al. (2014) and Rand (2016) suggest that cooperation might be the default choice, whereas defection requires active reflection. In the present thesis, I ask the following: If deliberation and intuition elicit different degrees of cooperation, through which processes does a deliberation treatment work? Is it by altering the players' preferences, or by changing their beliefs? In addition to having a large dataset, it is this last question that makes this study novel.

To determine whether intuition and cooperation give different degrees of cooperation, we needed a controlled experiment with a control group and a treatment group. In addition to providing valuable observations, the lab experiment served as a pilot for the NCP study. Having first run the game in the lab, we could better decide how many subjects were needed. The lab experiment was conducted in six sessions on the same day. To get well-balanced groups, we had within-session randomization with both control and treatment in each session.

The following game tree illustrates the decisions included:

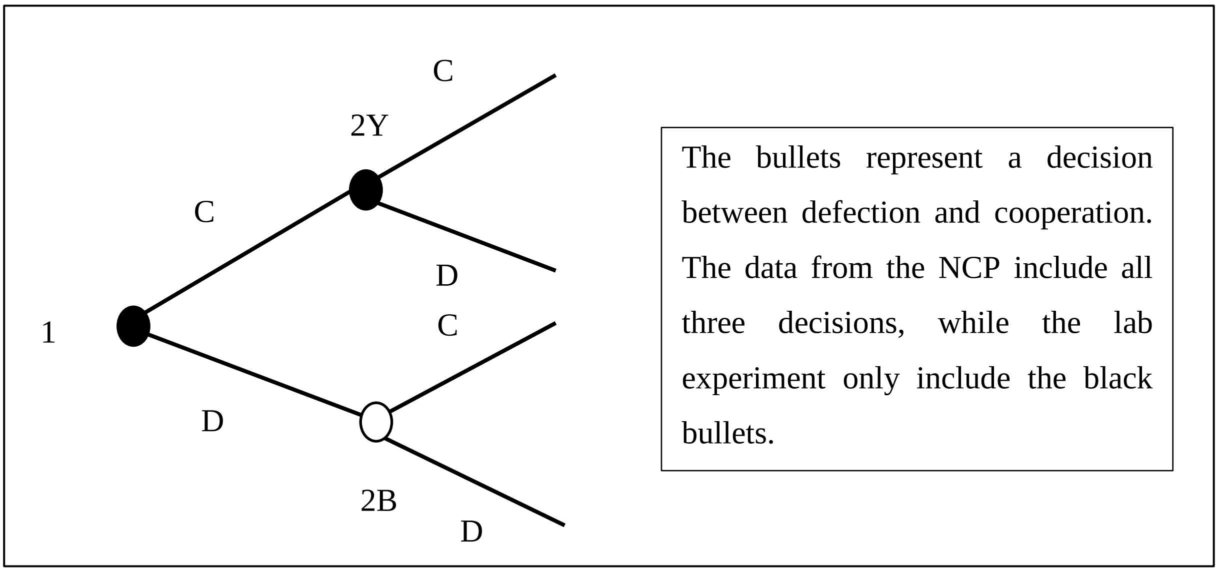

Figure 2.1. Game tree. Numbers indicate which player makes the decision, and letters indicate cooperation or defection. 2Y is the player who is told that player 1 has chosen the yellow box (i.e. cooperated). 2B is the player who is told that player 1 has chosen the blue box (i.e. defected).

To answer the question about what cognitive processes were involved, we ran a sequential 2 (player 1 and player 2) x 2 (control and treatment) prisoner's dilemma game. Half of the subjects were first-movers; player 1, and the other half played second move; player 2. Within each of these groups, there was a control group and a treatment group. Each player chose between a yellow box and a blue box; the yellow box corresponding to cooperation and the blue box to defection.

By playing a sequential game, it is possible to separate the effect through preferences from the effect through beliefs. Player 2 knows, when making her decision, whether player 1 has chosen to cooperate or not. It is then reasonable to assume that the only way a deliberation treatment might change her willingness to cooperate, is through a change in her preferences for cooperation. After all, she already has full knowledge about player 1's decision.[10] For player 1 however, there is also the element of uncertainty. He does not know whether player 2 will choose to cooperate or not, but must make his own assessment of the probability that she will. Thus, the deliberation treatment might affect player 1's beliefs about the game, specifically about player 2's decision.

To distinguish the preference effect from the beliefs effect, one simply subtracts the deliberation effect on cooperation for player 2 (i.e. the preference effect) from the effect for player 1 (the total effect).[11] The identifying assumption then is that, in absence of the treatment, player 1 and 2's preferences are the same in expectation. This makes the randomization vitally important.

In the lab experiment, only the black bullets in the game tree were played out. This was done mainly to get enough statistical power in at least one stage of the game, but also because this is considered the most relevant test of SHH. It was believed that for a player who had been told that the other player had defected (i.e. chosen the blue box), there would be too little cooperation to observe an effect in the lab study (Rabin, 1993, Falk and Fischbacher, 2006).

The player who knows that player 1 has chosen the blue box (i.e. defected) is called player 2B. It seems likely that reciprocity is a strong social norm, and that few would choose to cooperate if they already knew their partner had defected (Rabin, 1993, Falk and Fischbacher, 2006, Gächter et al., 2010, Rand and Nowak, 2013). For this reason, in the lab pilot, all subjects playing second move were told that player 1 had decided to cooperate. In the rest of the text, player 2 who knows that player 1 has cooperated, is called player 2Y.

[10] A different interpretation is that agents reveal their true preferences through deliberation. If the true preferences are systematically different from preferences "clouded" by intuition, which is what the present thesis tests, it is meaningful to speak of distinct sets of preferences.
[11] Assuming preferences are equally affected by the treatment for both players.

Based on results from earlier studies, we were confident that many in the player 1 group would cooperate (Ledyard, 1995, Chaudhuri, 2011). All those assigned as player 2 were matched with a cooperating player 1. Provided that at least one in the player 1 group cooperated, there is no deception, as player 2 was told that she had been matched with a player 1 who had chosen to cooperate. Exactly who this player 1 was, is not important, as long as he existed. Player 1 was matched with a random player 2, and rewarded according to the combination of their decisions according to the following payoff-matrix:

Table 2.2.1 Payoff-matrix for the lab experiment. Payoff in NOK.

Player 2

		Cooperate	Defect
Player 1	Cooperate	210, 210	100, 300
	Defect	300, 100	120, 120

Player 1 plays the game as if it were simultaneous. All players are rewarded in the exact way that they were told in the instructions, and their decisions matter to the other players, so there is no deception.[12]

Game setup and instructions were almost identical in the lab pilot as in the NCP data.[13] The main difference was the payoff-structure. In the NCP, only four out of the 1173 respondents were drawn to receive payment. The payoff-matrix was the following:

Table 2.2.2 Payoff-matrix for the NCP. Payoff in NOK.

Player 2

		Cooperate	Defect
Player 1	Cooperate	1000, 1000	0, 1500
	Defect	1500, 0	500, 500

The expected value of the highest paying selfish outcome is 5,1 NOK, which amounts to weaker incentives than in the lab. Regarding relative desirability of the outcomes, the lab study and the

[12] One unfortunate aspect of this way of matching respondents, is that by telling all player 2s that they will be matched with someone who has chosen to cooperate, one risks getting an expensive level of cooperation from player 2. Then again, playing the entire game tree would require more respondents, so the chosen solution appears to be the most convenient for the purpose of testing the SHH.

[13] See appendix B and C for instruction sheets from the lab experiment and NCP respectively.

NCP study are quite similar, and the SHH makes no predictions about differences in treatment effects following differences in material payoff (Rand et al., 2014, Bear and Rand, 2016). The different payoffs might, however, explain a lower over-all cooperation in the lab than in the NCP, due to a higher expected material payoff (Rabin, 1993).[14] Even if subjects use non-linear probability weighting (as suggested by Kahneman and Tversky (1979)), the difference in expected value is quite substantial.[15]

2.3 Treatment design

Roughly half of the subjects in the present research were asked to give a written justification for their decision before ticking off the yellow or blue box. This treatment is meant to force the subjects to deliberate their decision. In earlier literature, the treatments have largely been either time pressure (Rand et al., 2012, Cone and Rand, 2014), cognitive load (Hauge et al., 2016), ego-depletion (Achtziger et al., 2011), or induction (Liu and Hao, 2011), all meant to stimulate intuition (Rand, 2016). A few have also tried to stimulate deliberation through time delay; e.g. Rand et al. (2012). In the present research, we induced deliberation by asking the respondents to give a written justification of their decision before actually making the decision. This was done in the exact same way in both the lab experiment and the NCP survey.[16] One important motivation for choosing this treatment, was to be able to induce deliberation while sustaining a relatively high level of compliance.

Compliance is often a problem in studies with time pressure, where many of the subjects will fail to respond within the given time frame (Tinghög et al., 2013). Even with time delay, one will see subjects failing to comply with the time delay (e.g. Rand et al. (2012)). Still, compliance is generally much higher with time delay than with time pressure. In our lab experiment, we had full compliance to the treatment. This is not only good for the preservation of scarce observation data, but means we do not have to worry about heterogeneity in compliance. Heterogeneity could give a selection problem in the analysis, and is therefore important to rule out.

To make sure the treatment gave a deliberation effect, and not just an effect on understanding through reading the instructions several times, the instruction sheets in the lab experiment were

[14] This is stated as the third stylized fact in Rabin's famous article.
[15] Prospect theory suggests that agents do not weight probabilities linearly, but put a different emphasis on loss and gain. In the present case, prospect theory predicts that subjects will place an overly large probability weight on the prospect of winning 1500 NOK (Kahneman & Tversky, 1979).
[16] The exact text presented to the subjects are presented in the appendix.

recollected before proceeding to the treatment and decision making. In the NCP survey, subjects were not able to go back to the instructions after proceeding to the decision-making screen. Allowing the subjects to read the instructions several times might have given a better understanding of the payoff-structure, but could also harm the validity of the treatment. If the treatment gave the subjects a better understanding of the game, this could have had its own effects on the outcome. However, if a deliberation effect is conditional on understanding, as suggested by Strømland et al. (2016), the game must be understood by enough subjects.

2.4 Procedure in the lab experiment

The lab experiment was conducted by Sidra Choudhary, Alexander Daoud, Kjetil Madland and Amanda Reigstad as part of a course in experimental economics at the University of Bergen. This experiment also served as a pilot for the seventh round of the Norwegian citizen panel (NCP). I did a power-analysis from the lab data, to get an estimate of how many subjects would be needed in each branch of the game tree in NCP. The empirical results from the lab study, along with the power-analyses and results from NCP, are presented in section 3.

Participants in the lab pilot were recruited from mail lists of all students who had taken part on seminars in the subject ex.phil. at the University of Bergen in 2013 and 2014. This would suggest that the sample is relatively representative for the student mass at UiB. The geography of the campus means we must expect fewer students from the faculty for medicine and odontology to have participated. The lab is located in proximity to all other faculties, meaning the selection of students was probably quite diverse. The subjects were invited to participate voluntarily, and informed that they would be paid at least 100 NOK and possibly more, depending on their choices. They were informed that there would be an economic experiment, but not the purpose of the experiment.

The experiment was conducted in six sessions, all in one day, at the DIGSSCORE *Citizen lab*. At arrival, the subjects were assigned to seats by drawing a random number from a bowl. The seats determined which group the subjects would play, and who would get the treatment. Therefore, it was important to make sure seating was randomized.

After a brief introduction, the subjects were presented with the instruction sheet, which they had three minutes to read through before the sheets were recollected. Then they immediately got their respective decision sheets, and another three minutes to make their decisions (and give a written argument for their decision if they were in the treatment group). All sheets were handed out and collected in the exact same order in all sessions. This way, all subjects had the same amount of time in all stages of the experiment.

Once the decision sheets had been collected, the participants had to wait for the payments to be arranged. All in all, the sessions lasted an average of 25 minutes. The average payment to subjects were 221 Norwegian kroner (NOK). This makes for an hourly wage of 530 NOK; about three times what a student will earn in an average part time job, so the experiment should be well enough incentivized. Since cooperation in this case is about a trade-off between one's own payoff and the payoff of a stranger, the level of material payoff can be expected to be of some importance, if not for the effect size, then at least for the level of cooperation (Rabin, 1993). To ensure subject anonymity, we used a double-blind procedure, so those who matched player 1 and 2 and arranged the payment, were never in the same room as the subjects.

2.5 The Norwegian Citizen Panel

The Norwegian Citizen Panel (NCP) is an online survey panel consisting of around 5000 active individuals, invited and selected on the basis of a probability sample to get a representative selection of the Norwegian population over the age of 18 (DIGSSCORE, 2017).[17] The first panel survey was conducted in November 2013. Each participant will answer a survey approximately twice a year.

Representativeness makes the NCP preferable to Amazon Mechanical Turk (AMT), which is a digital on-demand work force. Here, anyone can sign up to complete *human intelligence tasks*, and the workers can choose which tasks they would like to complete. This potentially makes selection a problem.

Overly experienced workers, is another potential issue with AMT, at least for studies like the present one. One often sees that a relatively small share of AMT workers are responsible for a large portion of responses, and that many of them have extensive experience with classic paradigms within social sciences (Paolacci and Chandler, 2014).

In some situations, where understanding is important, AMT may well prove to be a better option than the NCP, because of the differences in experience. One could argue that AMT workers provide more realistic responses than people who only rarely participate in surveys. Occasional survey participants may perceive the survey as unfamiliar and unnatural, and thus may answer differently than they would have done in "natural" everyday situations. However, for testing

[17] The probability of being selected is a function of geography, gender, age and education among other things.

the social heuristics hypothesis (SHH), AMT would prove futile, as the SHH predicts a deliberation effect only for inexperienced individuals (Rand et al., 2014, Bear and Rand, 2016). For this study, the NCP might therefore be the best format for data collection.

Our experiment was included in a round of survey in the NCP in 2016, and had N=1173 participants. These were participants with little or no experience from economic games, because, according to the SHH, there will be a negative deliberation effect on cooperation only for subjects with limited experience from economic games and other "artificial lab environments" (Rand et al., 2014). Considering recent research on cooperation, the SHH seemed as the best candidate for a theory for explaining cooperation in a dual process framework, so lab experience was a leading concern when designing the experiment.

The present experiment being part of a survey round in the NCP, participants had answered some questions before completing this experiment. These questions were mainly about politics and the organization of society. Examples of questions were "In general, how pleased are you with the Norwegian government?" and "Do you think refugees should have the same rights to social benefits as Norwegian citizens?". Participants were also presented with a set of articles and asked which ones they would choose to read if they had to read a given number of them.

The most troubling section of questions were about voluntary work, which may have primed the subjects to cooperation. Although the chances of framing and priming effects can never be written off entirely, it seems reasonable to assume this is not a significant problem in this study since the treatment is randomized. This illustrates why having control and treatment groups is of paramount importance for acquiring robust results. This way, subjects are evenly distributed across conditions, so that important parameters are the same in expectation.

The survey lasted about 20 minutes in total, with the studied game at the end of the survey. This means that the subjects may have been tired and less concentrated. Although tired and unfocused subjects can give more random replies, and thus give more noisy data, it can also make the subjects "lazier" and have them rely more on their intuition. This could actually be beneficial to our experiment, as a very focused, deliberating control group would make it hard to observe differences between treatment and control groups.

3 Results and interpretation

3.1 Lab experiment

The lab experiment was conducted with 187 subjects, randomized into control and treatment groups as shown in table 3.1.1.

Table 3.1.1 Number of respondents in the lab experiment, by groups.

	Treatment	Control	Total
Player 1	47	45	92
Player 2Y	47	48	95
Total	94	93	187

By running a simple regression of cooperation on the treatment dummy, all available knowledge from the lab study is pooled into a single deliberation effect. This is a relevant test of the SHH because the SHH predicts that, at least for a subsample of individuals, deliberation will always favour defection (Rand et al., 2014, Bear and Rand, 2016). Hence, the SHH predicts that the deliberation treatment will have a negative overall effect on cooperation.

The main regression is the one where the deliberation effect is analysed separately for player 1 and player 2. Here, we are able to separate the deliberation effect through the subjects' preferences from the effect through beliefs. We also control for comprehension of the payoff-matrix and for experience from similar lab experiments. Controlling for all possible interactions gives no significant coefficients, and lacks a solid theoretical foundation.

I use the linear probability model (LPM) estimated with OLS with robust standard errors. In a linear probability model with a binary dependent variable, the standard errors are heteroscedastic by construction (Verbeek, 2012, p. 207, Wooldridge, 2014, p. 205).[18] For all the regressions in the present thesis, probit and logit regressions give equivalent results as those from the LPM. All tables in the main body of the thesis are reported using the LPM estimated with OLS.[19] The same tables with probit and logit are presented in appendix A. The main reason for using OLS is that the assumptions are less restrictive than in the maximum likelihood-

[18] The only exception is if all the coefficients for the independent variables are zero (Wooldridge, 2014).

[19] Although my working hypothesis (i.e. the SHH) is that deliberation will reduce cooperation, only two-sided tests are conducted, because the evidence for the SHH is controversial, and some studies find opposing effects.

models, and are plausible with a randomized dataset (Cameron and Trivedi, 2005, p. 70-81, 860-864, Angrist and Pischke, 2008, p. 147-148).

Table 3.1.2 shows the effects of different binary variables on the likelihood that a given subject in the lab experiment will choose the yellow box, which indicates cooperation. Player 2Y is the baseline. Table 3.1.2. presents four different specifications, all estimated with OLS.

Table 3.1.2 Treatment effects on cooperation in the lab pilot. Linear probability model

	No controls	Separate effects for player 1 and 2Y (main)	Controlling for comprehension	Controlling for experience
Deliberation	0.0603	0.1551	0.0865	0.2563*
	(0.0706)	(0.0989)	(0.2249)	(0.1464)
Player 1		0.0861	0.0795	0.2826**
		(0.0987)	(0.2327)	(0.1364)
Player 1 * Deliberation		-0.1925	0.0743	-0.3277
		(0.1414)	(0.3052)	(0.2036)
Comprehension			-0.1000	
			(0.1892)	
Deliberation * Comprehension			0.0796	
			(0.2517)	
Player 1 * Comprehension			-0.0016	
			(0.2576)	
Player 1 * Deliberation * Comprehension			-0.3581	
			(0.3442)	
Experience				0.1426
				(0.1317)
Deliberation * Experience				-0.1877
				(0.2005)
Player 1 * Experience				-0.4045**
				(0.1930)

Player 1 * Deliberation * Experience				0.2316
				(0.2800)
_cons	0.3333***	0.2917***	0.3750**	0.2174**
	(0.0491)	(0.0663)	(0.1749)	(0.0879)
N	187	187	187	187
R^2	0.004	0.014	0.049	0.050

Robust standard errors in parentheses
Data from the Pencil-and-paper study in the DIGSSCORE lab
* p<0.1, ** p<0.05, *** p<0.01

The only effects that are statistically significant at conventional levels are found in the specification controlling for experience. This is just as predicted by the SHH. The estimated deliberation effect, however, is positive in our data. This is a noisy phenomenon, and doing many regressions will increase the probability of getting what looks like statistically significant results (Morgan and Rubin, 2012).[20] However, our results do not match the predictions from the SHH (Rand et al., 2014, Bear and Rand, 2016).

The positive estimated treatment effect on cooperation increases when explicitly controlling for game experience, relative to no controls. This contrasts with the prediction from the SHH; that deliberation should have an unambiguous negative effect on cooperation, at least for unexperienced subjects (Rand et al., 2012, Rand et al., 2014, Rand, 2016, Bear and Rand, 2016). The lab data suggests that the positive deliberation effect is driven by the very people for whom the SHH predicts a negative effect.

Importantly, the effects are imprecisely estimated, and could very well be just noise. Also, by conducting multiple tests, the probability of some being significant at the 5% level is above 5% (Morgan and Rubin, 2012). Hence, the data should not be used to draw conclusions, but should be analysed primarily as a pilot study prior to the NCP dataset, which is much more comprehensive. The lab findings suggest that this is a noisy phenomenon, so observation data is scarce. Therefore, in section 3.3, the lab data is pooled with the NCP data to gain more statistical power.

Interestingly, there are big differences in predicted willingness to cooperate between player 1 with and without experience. Experience has a positive, though not significant, coefficient. The

[20] With k independent covariates, the chance of at least one covariate showing a “significant difference” between treatment and control groups, at significance level α, is $1 - (1 - \alpha)k$.

interaction between player 1 and experience gives a rather large negative coefficient, significant at the 5% level. Player 1 with no experience, however, cooperates much more than player 2, also significant at the 5% level. This could suggest that there might be some sort of moral deliberation effect for a player 2 who knows that the other has chosen to cooperate.

In the pencil-and-paper lab experiment, there were three control questions at the end, to control whether the participants had understood the payoff-structure of the game. The variable *comprehension* in table 3.1.2 is a dummy, indicating that the subject has answered all three control questions correctly. Strømland et al. (2016) find that a time pressure effect on cooperation is conditional on payoff comprehension. Importantly, there are no significant differences in understanding between groups in our experiment. This is controlled with a regression of understanding on the player 1 dummy, deliberation and the interaction between them, presented in table 3.1.3.

Table 3.1.3 Treatment effects on understanding. Linear probability model.

	OLS (robust SE)
Player 1	-0.0778
	(0.0846)
Deliberation	-0.1099
	(0.0855)
Player 1 * Deliberation	0.0778
	(0.1259)
_cons	0.8333***
	(0.0544)
N	187
R^2	0.011

Robust standard errors in parentheses
Controlling for heterogeneity in comprehension of the payoff-structure.
* p<0.1, ** p<0.05, *** p<0.01

None of the explanatory variables are statistically significant, so I cannot reject the null-hypothesis of no difference in payoff comprehension across conditions.

The most notable change in table 3.1.2 when controlling for understanding of the payoff-structure, is the coefficient for the interaction between player 1, deliberation and comprehension. This coefficient is negative and quite large, as predicted by the SHH (Rand et al., 2014, Bear and Rand, 2016). It is however very imprecisely estimated, and far from significant at conventional levels. The same goes for all other coefficients, except for the interaction between player 1 and experience. Although still insignificant, it looks like controlling for understanding of the payoff-structure alters the estimated effects quite dramatically, as predicted by Strømland et al. (2016).

Assume, for the sake of argument, that the deliberation treatment has a negative effect on cooperation for a comprehending, deliberating player 1. Even if this is the case, the SHH still does not explain the lack of a similar effect for player 2. Here, the same interaction is virtually zero, only slightly positive. If there is indeed an effect for a comprehending, deliberating player 1, the present data suggests that this works through beliefs rather than preferences, just as Ellingsen et al. (2012) find for social framing effects. Player 2 knows, when making her decision, what player 1 has chosen. Therefore, there is no room for a treatment effect through beliefs in the second stage of the game.

For the lab pilot, the main specification is the one estimating separate effects for player 1 and player 2, without controlling for comprehension and experience. This is the same specification as in the NCP; decided before conducting the experiment. This specification allows for a separation of a deliberation effect through preferences from an effect through beliefs. This specification gives no statistically significant coefficients.

Interpreting a null-effect can be problematic, as the interpretation is ambiguous. One obvious interpretation of a null-effect is that there is simply no internal conflict between intuitive and deliberative thinking when deciding whether or not to cooperate with a stranger in a social dilemma. This is the scenario assumed by most microeconomic models trying to predict behaviour, e.g. Rabin (1993), Fehr & Schmidt (1999), Becker (1976) and Stigler & Becker (1977). Section 4 with discussion and concluding remarks contains a deeper discussion of interpretations.

Results from the lab experiment are not conclusive, but they give a sense of what to expect from the NCP data. Power-analyses are conducted to estimate the sample sizes needed to confidently observe a deliberation effect. The power-graphs show estimated statistical power both for the specification where all knowledge is pooled into a single deliberation effect, and estimated separately for the two different players. Statistical power is measured as (1-β), where β is the probability of making a type 2 error (i.e. not rejecting the null hypothesis when it is false) (Verbeek, 2012, p. 31). In addition to estimated effect sizes from the lab data, some hypothetical effect sizes are included for comparison. The observed effect sizes from the lab experiment are represented by green curves with triangular markers in all three figures.

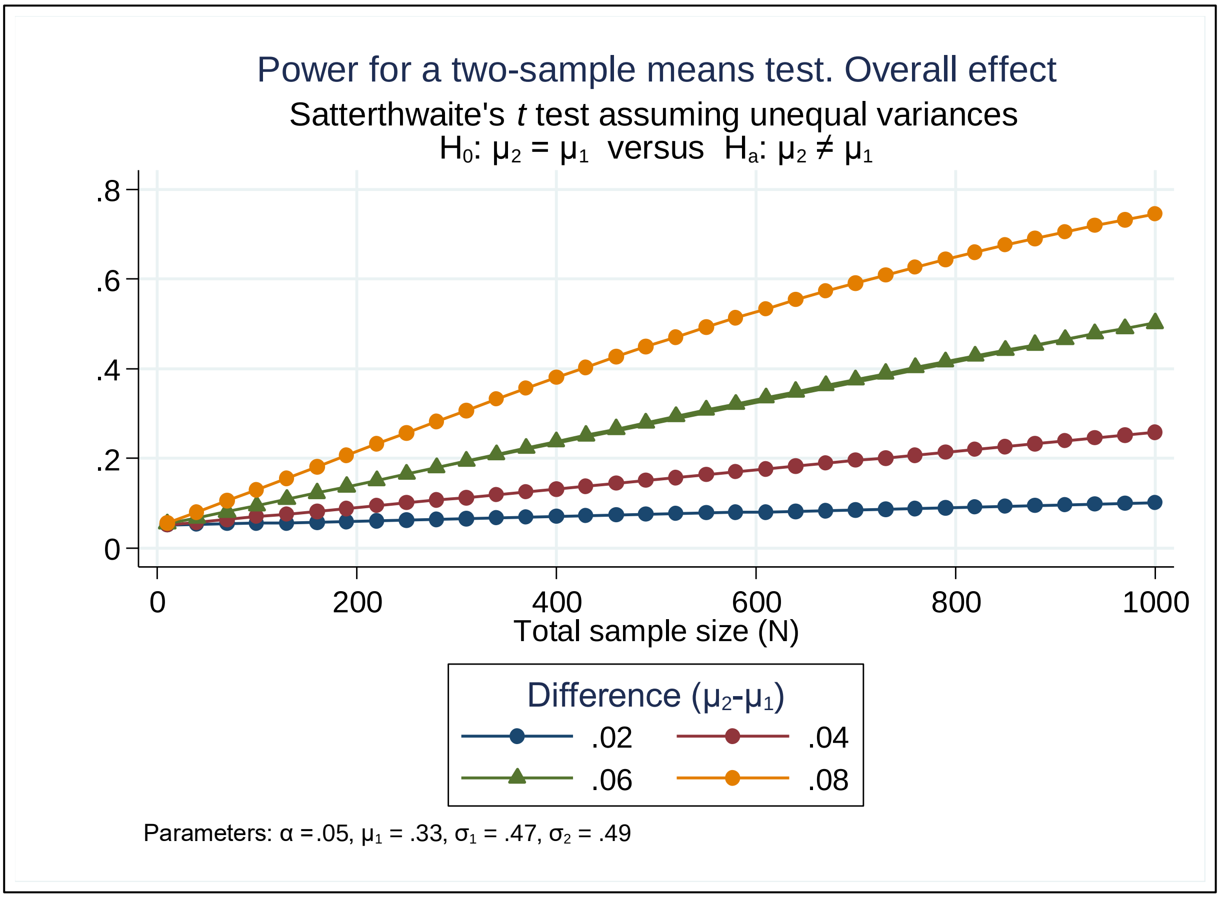

Figure 3.1 Power for two-sample means test of general deliberation effect, assuming equal sample size in each treatment group. No controls.

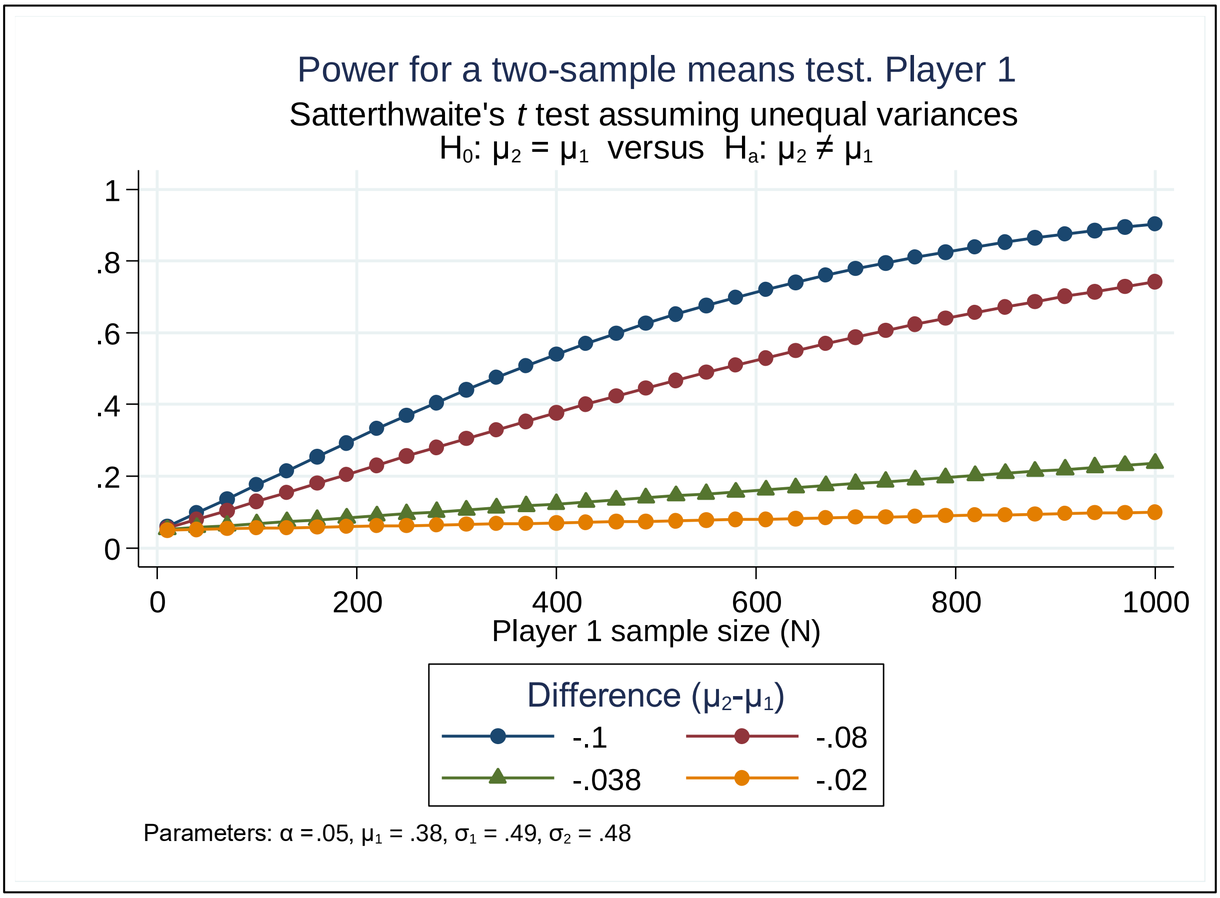

Figure 3.2 Two-sample means test of deliberation effect on cooperation for player 1, assuming equal sample size in each treatment group. No controls.

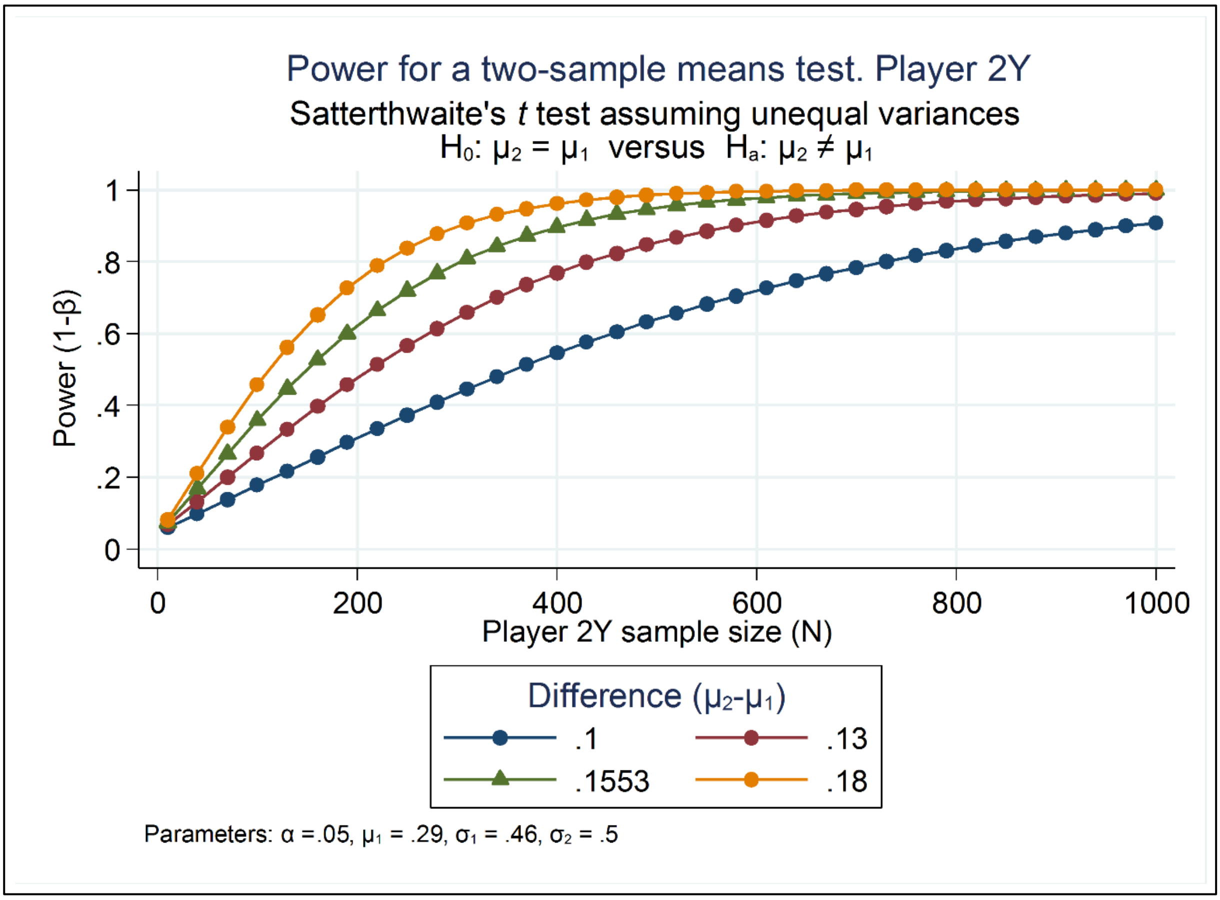

Figure 3.3 Two-sample means test of deliberation effect on cooperation for player 2Y, assuming equal sample size in each treatment group. No controls.

What do the power graphs tell us? The power graph for the entire population (figure 3.1.) suggests that if we want to test for an overall effect of deliberation on cooperation, like predicted by the SHH, even the 1173 observations from the NCP may not be sufficient to get a satisfying level of statistical power. Importantly, the power-analyses hinge on the assumption that the effect size is correctly estimated in the lab pilot.

Figure 3.2. shows statistical power for the estimated effect size for player 1. The estimated effect size is very small, and will demand thousands of observations to be precisely estimated, given that the effect size has been correctly estimated in the lab study. If the true effect sizes are larger or smaller than what has been estimated from the lab data, respectively less and more observation data is needed from the NCP to achieve an adequate level of statistical power.

For player 2Y, the estimated effect size is quite large. If this is correctly estimated in the lab pilot, the NCP dataset presented in section 3.2 should provide enough power to detect a treatment effect for player 2Y. In the present thesis, this is the most interesting effect, as it measures the pure preference effect of the treatment.

3.2 Results from the Norwegian citizen panel

In this section, I analyse the results from the NCP data. This contains the same experiment as in the lab pilot, but without control questions for understanding and experience, due to very limited space in the survey. The NCP data does, however, include response times, which are useful for testing whether the treatment worked as intended. Also, the NCP survey included all branches of the game tree, including the part where player 2 was told that player 1 had chosen to defect (i.e. the blue box). Players in this group are denoted player 2B.

I start with a general deliberation effect on cooperation without controls, as was done with the data from the lab experiment. As before, this is a relevant test of the social heuristics hypothesis, which predicts that deliberation should give an unambiguous negative effect on cooperation for both player groups, at least for a subsample of subjects.

The main specification is the one estimating separate effects for each player group, thus separating deliberation effects through preferences from effects through beliefs. Probit and logit regressions are presented in the appendix. Table 3.2.1 displays the treatment effects estimated with OLS. Player 2Y is the baseline.

Table 3.2.1 Treatment effects on cooperation for subjects in the NCP. Linear probability model

	No controls	Separate effects for each player group
Deliberation	0.0479	0.0473
	(0.0297)	(0.0509)
Player 1		-0.0197
		(0.0505)
Player 2B		-0.2140***
		(0.0514)
Player 1 * Deliberation		-0.0016
		(0.0704)
Player 2B * Deliberation		-0.0177
		(0.0729)
_cons	0.5253***	0.6034***
	(0.0209)	(0.0372)
N	1126	1126

R^2	0.002	0.042

Robust standard errors in parentheses
Effects estimated with OLS. Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y (i.e. player 2 who knows that player 1 has chosen the yellow box) is the baseline. Only data from the NCP
* p<0.1, ** p<0.05, *** p<0.01

The positive coefficient for deliberation lends no support to the SHH, and is more in line with the notion of a moral deliberation effect (Epley et al., 2006). With the observed coefficients and levels of significance however, it would be merely speculative to suggest that the NCP data support either of the theories. As was the case in the lab pilot, the estimated treatment effects are not significant at conventional levels.[21] This is the same both in the main specification and for the regression where all knowledge is pooled into a general deliberation effect on cooperation.

In the present NCP round, only unexperienced respondents were drawn for participation in the game studied, according to explicit instructions to external programmers. Experimental experience from other studies than the NCP is not controlled for.

As was clear from the lab pilot, a deliberation effect, if present, can depend on many factors. It also appeared to vary a great deal between player groups. The specification to the right in table 3.2.1 shows estimated deliberation effects and interactions on the different player groups, with player 2Y as the benchmark.

There does not seem to be any significant difference in the willingness to cooperate between subjects assigned as Player 1 and those assigned as Player 2Y. The only statistically significant coefficient is that for player 2B; the ones who are told that player 1 defected. The negative and significant coefficient is not surprising, and merely confirms that there is probably an element of reciprocity in cooperation between strangers in a social dilemma (Rabin, 1993). After all, player 2B knows, at the time of the decision, that the other player has chosen the blue box (i.e. defect).

The coefficient for the interaction between knowing that player 1 has defected, and getting the treatment (*Player 2B * deliberation*), is virtually zero. This indicates that, once controlling for knowing that the other player has defected (i.e. player 2B), there is no room for a treatment effect. This confirms the assumptions made when deciding to exclude this branch of the game tree from the lab pilot. Few people would choose to give away all the (potential to earn) money

[21] The point estimates in NCP have smaller magnitude than those in the pilot. This might just be because of the larger sample size.

to the other player, when the other player has already shown him- or herself as a defector. None of the other variables have significant estimated effects either. Baseline cooperation in the NCP is 60 %, which is relatively high, but within normal, given that the baseline is the case where player 2 knows that player 1 has cooperated (Ledyard, 1995, Chaudhuri, 2011).

Both specifications in table 3.2.1 show no statistically significant effect of the deliberation treatment, and thus no support for the SHH. However, the exact interpretation of a null-effect is ambiguous, as several mechanisms could explain a null-effect. For example, did the treatment fail to induce deliberation? Is there a selection problem in missing responses? Did too many subjects misunderstand the instructions or the payoff-structure? Although there were no questions to control for comprehension, the text data may shed some light on whether the subjects have understood the game or not. It may also tell us something about subjects' motivation for their choices.

3.2.1 Treatment compliance, comprehension and missing responses

Going through the text data from the Norwegian citizen panel, I have divided the written justifications into five categories, in accordance with the motivation expressed in writing. The "Other"-category includes responses which just state for example "Blue box", or that lack a clear interpretation. The categorization is done in a very ad hoc way, especially since the answers are anonymized, and thus cannot be connected to actual choices. Hence, the descriptive statistics in table 3.2.2 should be interpreted with extreme caution. Note also that what the subjects have written, does not necessarily reflect their true motivation.

Table 3.2.2 Categorization of answers from subjects in the treatment groups.

	Equality	Reciprocity	Selfish optimization	Misunderstood	Other
Subjects:	228	12	83	73	69
Percent:	48,93 %	2,58 %	17,81 %	15,67 %	14,81 %

Looking at the answers given, some tendencies materialize: Almost half of the subjects in the NCP state equality or fairness as the primary motivation for their choice. In many cases this is hard to separate from reciprocity. Also, some answers may be artefacts of demand effects. One might speculate that the way the treatment was conducted led subjects to make the decision they felt was easiest to justify. It is, however, not clear which direction this would "nudge" the subjects. Subjects might also act differently if they do not believe that they are anonymous, or if they believe that what they write will be read.

The most disturbing number in table 3.2.2, is the share of subjects writing something that indicates they have clearly misinterpreted the payoff-structure, or the game altogether. Still, as long as there is a control group and a treatment group, and there is no selection bias in understanding, differences in understanding will not cause inconsistent predictions. However, misunderstanding could increase the number of random responses and make the data noisier. This could be part of an explanation for a null-effect, and is an important reason why more observation data is needed.

The text data from the NCP shows that 117 out of 582; that is 20% of all participants in the treatment groups, skipped the writing altogether. This is a lower treatment compliance than we hoped for, and clearly shows that there are some significant differences between a lab study and a web based survey. Still, 80% compliance is much higher than what can be expected from studies with time pressure on this subject (Tinghög et al., 2013, Bouwmeester et al., 2017) (see e.g. Rand et al. (2012)).

Importantly, there does not seem to be any significant differences in compliance to treatment between player groups in the NCP, so heterogeneity between groups should not be a problem.[22] On the actual cooperation decision, the response rate was 96%, with no statistically significant differences between the treated and the control group. Regressions controlling for selection in missing responses are presented in table 3.2.4.

The 47 observations with missing response on at least one question, are dropped from the regressions. I do not have access to the subject ID numbers for the written justifications for each player; they are anonymized. As already pointed out, treatment compliance is about 80%, and virtually the same for all player groups. This is important to the validity of the treatment. While low compliance makes the estimates noisier, selection bias in compliance would have had more dramatic effects on estimated treatment effects.

[22] Treated groups had following treatment compliance rates: Player 1: 82%. Player 2Y: 81%. Player 2B: 76%. Numbers are calculated for all players assigned the respective groups.

Table 3.2.3 Descriptive statistics from the NCP data

Number of respondents by groups				
	Player 1	Player 2B	Player 2Y	Total
Control	209	190	174	573
Treatment	197	167	189	553
Total	406	357	363	1126

Only subjects who answered the questions completely.

Since the text data are anonymized, I cannot know whether the same subjects skipped both the decision and the written justification. With data connecting the written justification and the cooperation decision, one could use treatment as an instrument variable to estimate treatment effect on the treated. In the present case, since treatment is randomized, it seems plausible that the instrument (deliberation treatment) affects the dependent variable (cooperation) only through the instrumented variable (giving a written justification) (Imbens and Angrist, 1994). Thus, being in the treatment group works as an instrument for giving a written response. Hopefully, this connection between text data and cooperation decisions will be feasible in later datasets.

The concern is that the same subjects skip both the decision and the written justification. If this is the case, there is a selection problem. What we measure is not really the effect of the treatment, but the effect of "intention-to-treat". If there is a selection bias in missing responses, the difference can be substantial. If missing responses are (approximately) uniformly distributed between treatment groups, the difference is negligible. Doing a regression on missing response; propensity to drop out, it is possible to estimate whether or not there is a selection problem. Table 3.2.4 shows the results from regressions of missing responses on the decisions.

Table 3.2.4 Linear regression on propensity to drop out.

		OLS (robust SE)	OLS (robust SE)
Player control	1	-0.0146	
		(0.0163)	
Player treatment	1	0.0104	
		(0.0196)	
Player control	2B	0.0071	
		(0.0194)	
Player treatment	2B	0.0285	
		(0.0225)	
Player treatment	2Y	0.0121	
		(0.0200)	
Treatment			0.0194*
			(0.0115)
_cons		0.0333**	0.0305***
		(0.0134)	(0.0071)
N		1173	1173
R^2		0.004	0.002

Robust standard errors in parentheses
Measuring the impact of treatment, and on specific player group, on propensity to drop out. In the regression on all player groups, player 2Y in the control group is the baseline. Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y is player 2 who knows that player 1 has chosen the yellow box).
* p<0.1, ** p<0.05, *** p<0.01

Regressing missing response only on treatment, gives a coefficient which is statistically significant at the 10% level, although the magnitude of the coefficient sows doubt about real significance. In the main regression, estimating separate effects for the different player groups, there are no statistically significant coefficients. As in the other regressions, player 2Y in the control group is used as baseline.[23] Considering both the magnitude of the coefficient, and the

[23] Using player 1 in the control group as baseline, player 2B in the control group gets a coefficient of 0.043 (P=0.035), but this is hardly a theoretically relevant comparison.

P-value compared to the number of regressions conducted, it seems safe to assume there is no (serious) selection problem in missing responses (Morgan and Rubin, 2012).

3.2.2 Response times

Was the deliberation treatment successful in inducing a deliberative mindset? A significant advantage with the NCP dataset compared to pencil-and-paper experiments, except from being quite comprehensive, is that we have access to response times for the decision for all subjects. Adding the time taken by the treatment group to give a written justification for their decision, to the time taken to choose between the boxes, and comparing this to the decision time in the control group, one gets an indication of whether or not the deliberation treatment worked as intended. If total response time increased significantly with the treatment, one can be relatively confident that the treatment had the desired effect (Rand et al., 2012, Rand, 2016). Therefore, it is important to test the effect of the deliberation treatment on response time. Table 3.2.5 shows the effect, estimated with OLS, of the treatment on total response time, measured in seconds.

Table 3.2.5 Effect of deliberation treatment on response time in seconds.

	Player 1	Player 2Y	Player 2B
Deliberation	61.1141***	62.3544***	65.1933***
	(7.1217)	(8.4491)	(9.2993)
_cons	88.8038***	95.4382***	108.0102***
	(4.9478)	(6.0555)	(6.3575)
N	404	366	368
R^2	0.155	0.130	0.118

Standard errors in parentheses
Shows the effect of the deliberation treatment on response time measured in seconds. Subjects taking more than 10 minutes are excluded as outliers. For the control group, I consider the time taken to make the decision between the two boxes. For the treatment group, I consider the total time spend on the written justification and the decision between the two boxes combined. Data from the NCP.
* p<0.1, ** p<0.05, *** p<0.01

To prevent the coefficients from being inflated by participants leaving their computers in the middle of the survey to have dinner,[24] the regressions are run only for participants with a total response time less than 10 minutes. Including all participants, gives estimated treatment effects that are clearly artefacts of a few extreme outliers. The cut-off time of 10 minutes was chosen visually as a natural cut-off in response times. Other cut-offs were also considered; the results

[24] Or go on vacation. A few participants exceeded response times of 100 000 seconds, and one managed a total response time of over one million seconds; just short of twelve days.

are quite sensitive to the choice of cut-off time. However, the tendencies are the same for all reasonable exclusions.[25]

Choosing a specific cut-off time, is a common way to deal with extreme right outliers (Ratcliff, 1993, Whelan, 2008). A problem with the method is the lack of a clear, general rule for the cut-off (Whelan, 2008). Log-transforming response times is another way to minimize the effects of outliers, and make the distribution closer to normal, without losing observations (Ratcliff, 1993, Whelan, 2008, Wooldridge, 2014, p. 157, 264-269).[26] The histograms in figure 3.4. show the distributions of the natural log of response time in seconds. It is evident from the histograms, that the deliberation groups have more density towards longer response time than the control groups. Whether the differences are large enough to compensate for the extra task is an open question.

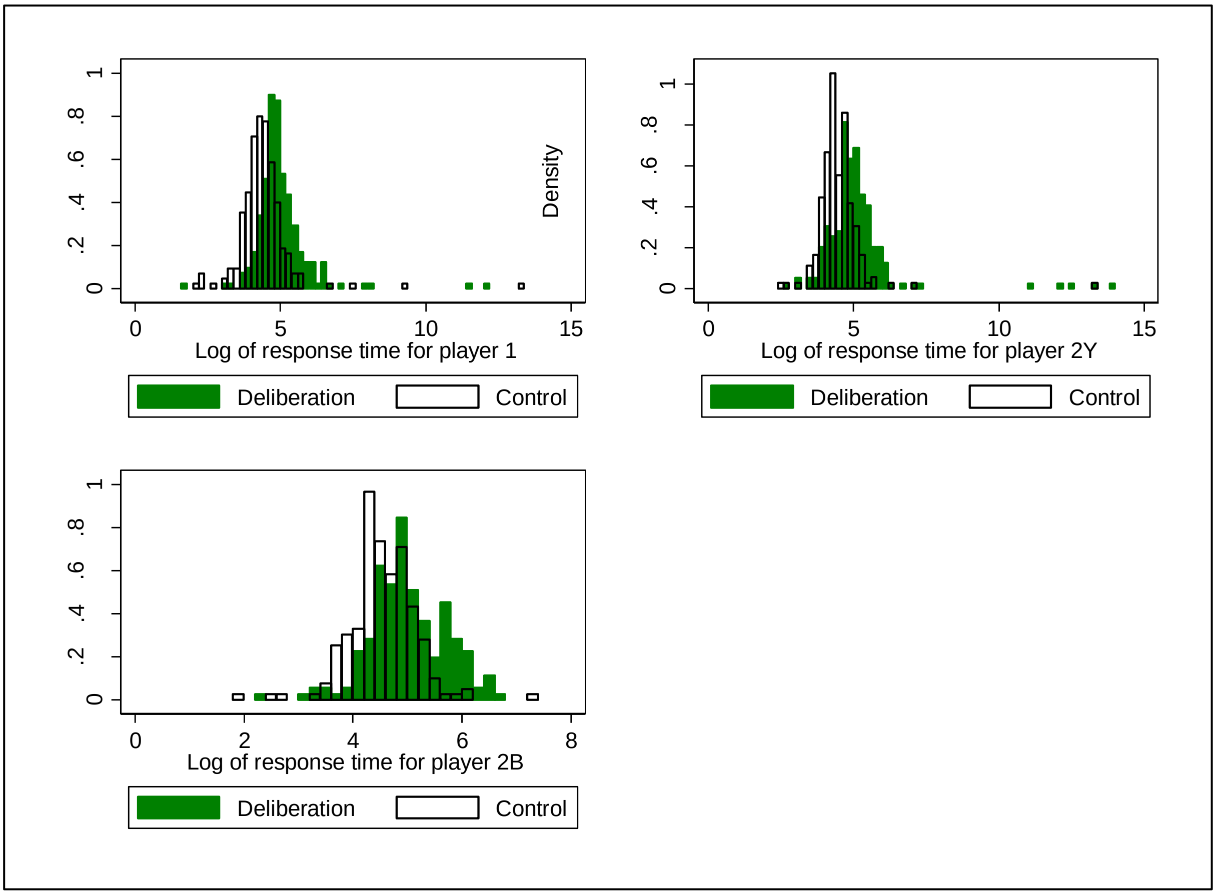

Figure 3.4 Log of response time in seconds. One histogram for each group of players.

[25] Table 3.2.5A in appendix A shows the estimated effects of the treatment on the natural log of response times, without excluding outliers. Results are similar to those in table 3.2.5.

[26] Taking the natural log is a monotonic transformation; hence keeping the ordering of values unchanged. Log-transforming response times reduces the effects of extreme outliers.

The effects of the treatment on response time are sensitive to cut-off time, but the estimated coefficients are large and highly significant for all player groups. This may not be so surprising, as the treatment group is asked to do an extra task. Using response time as a proxy for deliberation is in accordance with earlier research (Rand et al., 2012, Cone and Rand, 2014, Rand, 2016), albeit controversial (Myrseth and Wollbrant, 2015). It is believed that writing down a justification for a decision of whether to cooperate or not, is likely to stimulate to increased deliberation.

If there are (probably) no selection problems in treatment compliance or missing responses, and the treatment appears to have worked as intended, how should we interpret the null-effect? In addition to the possibility that cognitive modes do not affect preferences for cooperation in social dilemmas, there is a very real possibility that we simply do not have enough statistical power to detect the effect.

In section 3.3, the NCP data is combined with the data from the lab experiment to get a pooled analysis, with more statistical power.

3.3 Pooled analysis

Pooling the data from the lab experiment and the NCP survey gives a comprehensive dataset, covering the same social dilemma. All questions were formulated in a very similar way; the main difference being the payoff-structure. In the lab experiment, the participants were told that if both cooperated, they would earn 210 NOK each. If one cooperated and one defected, the defector would earn 300 NOK, while the co-operator would earn the minimum guaranteed sum of 100 NOK. If both defected, they would both earn 120 NOK. In the NCP survey, both cooperating, would earn both players 1000 NOK, both defecting would earn both 500 NOK, and one defecting and one cooperating would give 1500 NOK to the defector and 0 to the co-operator. However, in the NCP, only four people were randomly drawn to receive payment. Thus, the payoff in the two experiments was not identical, but the relative desirability of scenarios similar.

Running the linear probability regression on the entire pooled dataset, both with a general deliberation effect, and with the specification estimating separate effects for the first- and second mover, table 3.3.1 is obtained.

Table 3.3.1 Treatment effects on cooperation. Pooled analysis with OLS

	No controls	Separate effects for each player group
Deliberation	0.0497*	0.0698
	(0.0273)	(0.0452)
Lab	-0.1858***	-0.2529***
	(0.0381)	(0.0393)
Player 1		0.0013
		(0.0449)
Player 2B		-0.2013***
		(0.0490)
Player 1 * Deliberation		-0.0394
		(0.0629)
Player 2B * Deliberation		-0.0401
		(0.0690)
_cons	0.5244***	0.5907***
	(0.0201)	(0.0338)
N	1313	1313
R^2	0.019	0.053

Robust standard errors in parentheses
Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y (i.e. player 2 who knows that player 1 has chosen the yellow box) is the baseline.
Data from both the NCP and the lab experiment
* p<0.1, ** p<0.05, *** p<0.01

Lab is a dummy equal to one for the participants in the lab experiment. Overall cooperation is substantially higher in the NCP survey than in the lab study. One possible explanation is the weaker incentives in the NCP. Although the relative desirability of the alternative outcomes are properly incentivized, the expected payoff in all scenarios are a fraction of that in the lab experiment. Rabin (1993) argues that preferences for fairness and reciprocity compete against preferences for material payoff, and thus will be more pronounced when material incentives are weaker. This could to some extent explain the huge difference in levels of cooperation, in an otherwise identically framed game.

Selection might also be an issue. The subjects in the lab experiment were mostly students, some of whom might see economic experiments only as an easy way to earn some cash. This explanation is more speculative, although easily follows from Rabin's model if adding the common assumption of diminishing marginal utility of money. Also, sample differences do not devalue the results, as it is the treatment effects we are interested in, not levels of cooperation.

The overall estimated (without controls) deliberation effect is positive, and significant at the 10% level, contradicting the predictions of the SHH. Again: When running so many regressions, some are bound to be statistically significant by conventional levels (Morgan and Rubin, 2012). The results should therefore be interpreted with great caution. When allowing player group differences, the estimated deliberation effect is no longer statistically significant at conventional levels, but the magnitude of the coefficient for deliberation is somewhat larger.

Table 3.3.2 shows both individual datasets, as well as the pooled dataset, all analysed using OLS. The regression presented, is the one estimating the model separately for player 1 and player 2 (i.e. separating the effect through preferences from the effect through beliefs). As before, player 2Y is the benchmark.

Table 3.3.2 Treatment effects on cooperation. Comparison estimated with OLS

	Pooled	NCP	Lab
Deliberation	0.0698	0.0473	0.1551
	(0.0452)	(0.0509)	(0.0989)
Lab	-0.2529***		
	(0.0393)		
Player 1	0.0013	-0.0197	0.0861
	(0.0449)	(0.0505)	(0.0987)
Player 2B	-0.2013***	-0.2140***	
	(0.0490)	(0.0514)	
Player 1* Deliberation	-0.0394	-0.0016	-0.1925
	(0.0629)	(0.0704)	(0.1414)
Player 2B* Deliberation	-0.0401	-0.0177	
	(0.0690)	(0.0729)	
_cons	0.5907***	0.6034***	0.2917***
	(0.0338)	(0.0372)	(0.0663)
N	1313	1126	187
R^2	0.053	0.042	0.014

Robust standard errors in parentheses

Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y (i.e. player 2

who knows that player 1 has chosen the yellow box) is the baseline.
Data from both the NCP and the lab experiment
* p<0.1, ** p<0.05, *** p<0.01

Although not statistically significant, there are tendencies towards a treatment effect that is positive for player 2Y, and negative both for player 2B and for player 1. While the SHH predicts a negative deliberation effect for all groups, the negative coefficient for the interaction between player 1 and the deliberation treatment could also be explained as reactive egoism (Epley et al., 2006). The positive (though not significant) estimated treatment effect for player 2Y, might be interpreted as a moral response to a reciprocity norm (Rabin, 1993, Falk and Fischbacher, 2006). Knowing that the other has defected has a large, and not surprisingly, negative effect on cooperation.

The most interesting numbers are the coefficients for *Deliberation* and the interactions between *Deliberation* and the dummy for player 1. Neither of these coefficients are significantly different from zero. The deliberation effect for player 2Y is positive in the individual datasets, as well as in the pooled dataset, albeit well within two standard deviations from zero. Thus, the results presented here lend no support to the social heuristics hypothesis, and merely indicative support to the hypothesis of reactive egoism paired with moral deliberation (Epley et al., 2006).

The magnitude of the estimated effect of the deliberation treatment on cooperation for player 2Y, suggests there is need for follow-up studies with more data. From the present data, the effect size is estimated to be between 4,7 and 15,5%.[27] If this is a real population effect, the magnitude suggests that cognitive mode could be significant to decision making.

[27] Depending on which regression specification is used, and if looking at the datasets pooled or individually.

4 Discussion and conclusion

4.1 General discussion

The data presented in this thesis show no statistically significant deliberation effect on cooperation. This does not necessarily mean that the deliberation treatment has no effect on the subjects' decision making. If a sizable portion of the subject pool have experience with similar experiments, or are otherwise used to a competitive environment, one might not be able to observe a deliberation effect (Rand et al., 2012, Rand and Nowak, 2013, Rand et al., 2014, Bear and Rand, 2016). In the pilot, this is specifically tested for, and the results give no support to the SHH's explanation.

In a general regression of deliberation on cooperation, opposing effects through preferences and beliefs might cancel each other out. Thus, observing a null-effect with this specification, could mean that there is a negative effect though beliefs and a positive effect through preferences. This scenario is partly described by Epley et. al. (2006), where subjects, knowing that the other player has chosen to cooperate with them, follow a reciprocity norm, strengthened by their reflection and deliberation. This promotes cooperation. When uncertain of the other's decision however, like in the case of player 1, deliberation prompts perspective taking. Perspective taking then brings about mistrust towards the other player, resulting in reactive egoism (Epley et al., 2006, Torsvik et al., 2011).

Epley et al. (2006) find that subjects cooperate more when they know that the other players have cooperated, but do not explicitly suggest the moral deliberation in their theory. Instead, they focus solely on the effects on perspective taking when there is no information about others' decisions (Epley et al., 2006). They find that, when uncertain about the decisions of others, perspective taking increases taking. Torsvik et al. (2011) find similar effects of announcing post-play face-to-face discussions between participants in a PGG. This is interpreted as effects of perspective taking and empirical expectations. Our results suggest that the same goes for a deliberation treatment. The hypothesis developed by Epley et al., seems fully compatible with different deliberation effects conditional on what sort of knowledge one possesses when making the choice between cooperation and defection.

A moral deliberation effect is also on par with earlier research on perspective taking, and assumed to be the main benefit of taking others' perspective in social interactions (Neale and

Bazerman, 1983, Batson, 1995, Galinsky and Moskowitz, 2000, Savitsky et al., 2005, Epley et al., 2006).

Reactive egoism can give the same results as those predicted by the SHH; namely less cooperation. However, in SHH, the effect of cognitive processes on cooperation is claimed to go through preferences for cooperation, rather than beliefs (Rand et al., 2012). A moral or reciprocal deliberation effect has (to my knowledge) no firm theoretical footing, but would contradict the SHH.

The SHH states that the reason we should observe a deliberation effect is that people tend to over-use heuristics that work well in familiar situations. People who have experience with lab experiments might have formed a distinct set of heuristics about what kind of behaviour is likely to be beneficial in the lab. If so, they are likely to use these heuristics the next time they find themselves in a lab experiment, instead of extrapolating heuristics from situations that do not well resemble a one-shot economic game (Rand et al., 2014, Bear and Rand, 2016). One will likely observe similar results for people who are used to competitive environments where cooperation is generally not a good strategy (Rand et al., 2014). This may be the case for economics students, who have learned to recognise an economic game when they see it, and to find and play the Nash-equilibria.

Observing a treatment effect in the second part of the game (i.e. for player 2) would suggest that the treatment, through increased deliberation, alters the subjects' preferences for cooperation. Since player 2 already knows what player 1 has chosen, the is no room for changes in her beliefs. SHH predicts a decreased rate of observed cooperation, following from increased deliberation (Rand et al., 2012, Rand et al., 2014). If we had observed such an effect for player 2, the SHH would explain this as self-interested cognition; or system two thinking (Chaiken and Trope, 1999, Kahneman, 2003, Kahneman, 2011), overruling a more cooperative and benevolent intuition. There is, at least to my knowledge, no reason to believe that a deliberation effect through preferences would work differently on player 1 and player 2 when treatment groups are randomized.

The novel interpretation of a null-effect in the regression specification not allowing player group differences, is that the treatment affects player 1 and player 2 differently. If deliberation works through the players' beliefs, as is suggested to be the case for framing effects (Ellingsen et al., 2012), we would expect an effect only for player 1. If, on the other hand, the treatment works through preferences for cooperation, this should affect both players.

There are several theories suggesting that moral preferences, inequity aversion and reciprocity might be of some importance to the observed level of cooperation, e.g. Rabin (1993), Fehr and Schmidt (1999), Epley et al. (2006), Falk and Fischbacher (2006). Importantly, if deliberation has opposing effects through beliefs and preferences, the effects might cancel out in a general regression of cooperation on deliberation.

Results from the main specification in the present study lend no support to the possibility that a deliberation treatment has opposing effects though preferences and beliefs. When estimating different effects in different player groups, I find no statistically significant deliberation effects. Nor do the magnitudes of the estimated coefficients suggest any real significance. Thus, the most obvious interpretation of the present data is that there simply is no internal conflict between intuitive and deliberative decision making in a social dilemma with strangers. At least, the effect is not big enough to be precisely estimated with the statistical power available in this study.

An alternative interpretation of a null-effects is that the treatment did not work well enough, given the statistical power available. This could be because the general degree of deliberation was too high in the control groups. All participants had time to reflect and deliberate their decision, so it might be the case that too few relied enough on their intuition when deciding which box to tick off, i.e. to cooperate or defect. In the NCP survey, all subjects had unlimited time, as opposed to earlier studies, where intuition groups often have maximum 10 seconds (Rand et al., 2012, Tinghög et al., 2013, Rand, 2016).

It seems likely that the treatment in the present study induced reflective, deliberative thinking in the treatment groups, at least considering the relatively high rate of compliance to the treatment. One might suspect however, that most participants in the control group were also in a reflective state of mind. If this is the case, then the treatment will not lead to a significant difference in cognitive mode, at least considered within a dual-process-framework a la Kahneman (2003, 2011) and Chaiken & Trope (1999) where "choice" of cognitive subsystem is discrete rather than continuous.

Ideally, I would also control for understanding and for experience from experimental settings in the NCP data. This was not possible in the present data due to lack of space, but will hopefully be included in later datasets from the Norwegian and Swedish citizen panels. As we did not get enough space to control for understanding, we have no way of knowing if the subjects in the NCP understood the game properly, other than indications from the written replies. This is a problem if the deliberation effect is conditional on understanding, as claimed by Strømland et al. (2016) and Bear and Rand (2016). The findings from the lab experiment also indicate that a deliberation effect, if present, could be conditional on understanding. If many subjects in the NCP misunderstood the questions or the payoff structure, this could help explain a null-effect and imprecise estimates. The theoretical foundation for a deliberation effect on cooperation which is contingent on misunderstanding, is weak at best.

Whether the treatment worked as intended or not, is an issue which is much harder to test for. Clearly, deliberation and intuition are individual specific processes, and the time required for one individual to make a well thought out, deliberative response to a question, is not necessarily the same as for another. Using response time as a proxy for deliberation, the treatment appears to have worked well.

Not using time pressure as a treatment, meant all the subjects were given (potentially) the same amount of time to choose between the blue and the yellow boxes. Thus, they all had the same time to deliberate their decision. If average level of deliberation was very high in the control group, this will diminish the chances of being able to precisely estimate a treatment effect.

An important question is whether the treatment could affect the subjects in other ways than just by promoting more deliberation. With time pressure, there is the problem that the treatment group might experience more stress, which is not the same as acting intuitively. With our deliberation treatment, some subjects may experience a demand effect, and may think that they are supposed to cooperate. One indication that this might be the case, is that so many respondents in the NCP state equality and fairness as the motivation for their decision. Even some of those who know that the other player has chosen the blue box, write that they will choose the blue box themselves "because it is fair that both get the same amount". As is almost always the case, it is impossible to distinct demand effects from an actual desire to cooperate. The phrase "Whatever the other person chooses, you will earn most by choosing the blue box" at the end of the instruction sheet should serve as a remedy for potential demand effects for cooperation.

4.2 Concluding remarks

I find no statistically significant evidence that increased deliberation changes the willingness to cooperate in a prisoner's dilemma, neither through preferences, nor through beliefs. Thus, I find no support for the social heuristics hypothesis (SHH). If anything, the results presented in this thesis seem to suggest some support for reactive egoism, paired with a moral deliberation effect promoting reciprocity, as suggested by Epley et al. (2006). However, the evidence are not conclusive, and should be interpreted with caution. Given that we may have underestimated the statistical power using the data from the pilot study, more data is needed to conclude.

With so much evidence for the SHH coming from time pressure studies (Rand, 2016), the results presented in the present thesis could suggest these evidences are artefacts of the treatment. Perhaps time pressure causes stress rather than intuition? To properly control for this explanation, a follow-up study could include both intuition priming and deliberation priming, or time pressure and time delay, in addition to a control group as baseline. This sort of two-way treatment could also help deciding whether most subjects in cooperation experiments are making relatively intuitive or deliberative decisions.

As already pointed out, in the present study, one possible explanation for a null-effect is simply that the level of deliberation is so high in the control group that there is little room left for a deliberation effect on cooperation.

5 References

ACHTZIGER, A., ALÓS-FERRER, C. & WAGNER, A. K. 2011. Social preferences and self-control. Working Paper, University of Constance.

ANGRIST, J. D. & PISCHKE, J.-S. 2008. *Mostly harmless econometrics: An empiricist's companion,* Princeton University Press.

BABCOCK, L. & LOEWENSTEIN, G. 1997. Explaining Bargaining Impasse: The Role of Self-Serving Biases. *Journal of Economic Perspectives,* 11**,** 109-126.

BATSON, C. D. 1995. Prosocial motivation: Why do we help others. *Advanced social psychology***,** 333-381.

BEAR, A. & RAND, D. G. 2016. Intuition, deliberation, and the evolution of cooperation.(PSYCHOLOGICAL AND COGNITIVE SCIENCES: EVOLUTION). *Proceedings of the National Academy of Sciences of the United States,* 113**,** 936.

BECKER, G. 1976. Altruism, egoism, and genetic fitness: economics and sociobiology. *Journal of economic literature,* 14**,** 817-826.

BINMORE, K. & SHAKED, A. 2010. Experimental economics: Where next? *Journal of Economic Behavior and Organization,* 73**,** 87-100.

BOUWMEESTER, S., VERKOEIJEN, P. P., ACZEL, B., BARBOSA, F., BÈGUE, L., BRAÑAS-GARZA, P., CHMURA, T. G., CORNELISSEN, G., DØSSING, F. S. & ESPÍN, A. M. 2017. Registered Replication Report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science***,** 1745691617693624.

CAMERON, A. C. & TRIVEDI, P. K. 2005. *Microeconometrics,* Cambridge, Cambridge University Press.

CAPPELEN, A. W., NIELSEN, U. H., TUNGODDEN, B., TYRAN, J. R. & WENGSTROM, E. 2016. Fairness is intuitive. *Experimental Economics,* 19**,** 727-740.

CAPRARO, V., JORDAN, J. J. & RAND, D. G. 2014. Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments. *Forthcoming in Scientific Reports*.

CHAIKEN, S. & TROPE, Y. 1999. *Dual-process theories in social psychology,* Guilford Press.

CHAUDHURI, A. 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *A Journal of the Economic Science Association,* 14**,** 47-83.

CONE, J. & RAND, D. G. 2014. Time pressure increases cooperation in competitively framed social dilemmas. *PLoS one,* 9**,** e115756.

DAWES, R. M. 1980. Social dilemmas. *Annual review of psychology,* 31**,** 169-193.

DAWES, R. M. & THALER, R. H. 1988. Anomalies: cooperation. *The Journal of Economic Perspectives,* 2**,** 187-197.

DIGSSCORE. 2017. *Methodology* [Online]. Available: http://digsscore.uib.no/methodology [Accessed 02.02.2017].

ELLINGSEN, T., JOHANNESSON, M., MOLLERSTROM, J. & MUNKHAMMAR, S. 2012. Social framing effects: Preferences or beliefs? *Games and Economic Behavior,* 76**,** 117-130.

EPLEY, N., CARUSO, E. M., BAZERMAN, M. H. & DOVIDIO, J. F. 2006. When Perspective Taking Increases Taking: Reactive Egoism in Social Interaction. *Journal of Personality and Social Psychology,* 91**,** 872-889.

EVANS, A. M., DILLON, K. D. & RAND, D. G. 2015. Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General,* 144**,** 951-966.

EVANS, J. S. B. T. 2003. In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences,* 7**,** 454-459.
EVANS, J. S. B. T., BARSTON, J. L. & POLLARD, P. 1983. On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition,* 11**,** 295-306.
FALK, A. & FISCHBACHER, U. 2006. A theory of reciprocity. *Games and Economic Behavior,* 54**,** 293-315.
FEHR, E. & SCHMIDT, K. M. 1999. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114**,** 817-868.
GALINSKY, A. D. & MOSKOWITZ, G. B. 2000. Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology,* 78**,** 708.
GÄCHTER, S., NOSENZO, D., RENNER, E. & SEFTON, M. 2010. Sequential vs. simultaneous contributions to public goods: Experimental evidence. *Journal of Public Economics,* 94**,** 515-522.
HAUGE, K., BREKKE, K., JOHANSSON, L.-O., JOHANSSON-STENMAN, O. & SVEDSÄTER, H. 2016. Keeping others in our mind or in our heart? Distribution games under cognitive load. *A Journal of the Economic Science Association,* 19**,** 562-576.
IMBENS, G. W. & ANGRIST, J. D. 1994. IDENTIFICATION AND ESTIMATION OF LOCAL AVERAGE TREATMENT EFFECTS. *Econometrica,* 62**,** 467-475.
KAHNEMAN, D. 2003. A perspective on judgment and choice - Mapping bounded rationality. *Am. Psychol.*
KAHNEMAN, D. 2011. *Thinking, fast and slow,* New York, Farrar, Straus and Giroux.
KAHNEMAN, D. & TVERSKY, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society***,** 263-291.
KAHNEMAN, D., TVERSKY, A. & MANDLER, G. 1973. On the psychology of prediction. *Psychological Review,* 80**,** 237-251.
LEDYARD, J. 1995. Public goods: A survey of experimental research. *Handbook of Experimental Economics, Princeton University Press, Princeton***,** 111-194.
LIU, C.-J. & HAO, F. 2011. An application of a dual-process approach to decision making in social dilemmas. *American Journal of Psychology,* 124**,** 203-212.
MAS-COLELL, A., WHINSTON, M. D. & GREEN, J. R. 1995. *Microeconomic theory,* Oxford university press New York.
MORGAN, K. L. & RUBIN, D. B. 2012. Rerandomization to improve covariate balance in experiments. *The Annals of Statistics***,** 1263-1282.
MYRSETH, K. & WOLLBRANT, C. 2015. Intuitive cooperation refuted: Commentary on Rand et al. (2012) and Rand et al. (2014). *IDEAS Working Paper Series from RePEc*.
MYRSETH, K. O. R. & WOLLBRANT, C. E. 2016. Commentary: Fairness is intuitive. *Frontiers in Psychology,* 7**,** 2.
NEALE, M. A. & BAZERMAN, M. H. 1983. The role of perspective-taking ability in negotiating under different forms of arbitration. *Industrial & Labor Relations Review,* 36**,** 378-388.
PAESE, P. W. & YONKER, R. D. 2001. TOWARD A BETTER UNDERSTANDING OF EGOCENTRIC FAIRNESS JUDGMENTS IN NEGOTIATION. *International Journal of Conflict Management,* 12**,** 114-131.
PAOLACCI, G. & CHANDLER, J. 2014. Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science,* 23**,** 184-188.
RABIN, M. 1993. Incorporating fairness into game theory and economics. *The American economic review***,** 1281-1302.
RAND, D. G. 2016. Cooperation, Fast and Slow. *Psychological Science,* 27**,** 1192-1206.

RAND, D. G., GREENE, J. D. & NOWAK, M. A. 2012. Spontaneous giving and calculated greed. *Nature,* 489**,** 427-430.
RAND, D. G. & NOWAK, M. A. 2013. Human cooperation. *Trends in Cognitive Sciences,* 17**,** 413-425.
RAND, D. G., PEYSAKHOVICH, A., KRAFT-TODD, G. T., NEWMAN, G. E., WURZBACHER, O., NOWAK, M. A. & GREENE, J. D. 2014. Social heuristics shape intuitive cooperation. *Nature Communications,* 5.
RATCLIFF, R. 1993. Methods for dealing with reaction time outliers. *Psychological bulletin,* 114**,** 510.
RUBINSTEIN, A. 2016. A typology of players: between instinctive and contemplative.(Report). *The Quarterly Journal of Economics,* 131**,** 859-890.
SALLY, D. 1995. Conversation and cooperation in social dilemmas a meta-analysis of experiments from 1958 to 1992. *Rationality and society,* 7**,** 58-92.
SAVITSKY, K., VAN BOVEN, L., EPLEY, N. & WIGHT, W. M. 2005. The unpacking effect in allocations of responsibility for group tasks. *Journal of Experimental Social Psychology,* 41**,** 447-457.
STIGLER, G. J. & BECKER, G. S. 1977. DE GUSTIBUS NON EST DISPUTANDUM. *American Economic Review,* 67**,** 76-90.
STRØMLAND, E., TJOTTA, S. & TORSVIK, G. 2016. Cooperating, fast and slow: Testing the social heuristics hypothesis.
TINGHÖG, G., ANDERSSON, D., BONN, C., BÖTTIGER, H., JOSEPHSON, C., LUNDGREN, G., VÄSTFJÄLL, D., KIRCHLER, M. & JOHANNESSON, M. 2013. Intuition and cooperation reconsidered. *Nature,* 498**,** E1.
TORSVIK, G., MOLANDER, A., TJØTTA, S. & KOBBELTVEDT, T. 2011. Anticipated discussion and cooperation in a social dilemma. *Rationality and society,* 23**,** 199-216.
VERBEEK, M. 2012. *A guide to modern econometrics,* Rotterdam, Wiley.
VERKOEIJEN, P. P. & BOUWMEESTER, S. 2014. Does intuition cause cooperation? *PloS one,* 9**,** e96654.
WHELAN, R. 2008. Effective analysis of reaction time data. *The Psychological Record,* 58**,** 475.
WOOLDRIDGE, J. M. 2014. *Introduction to econometrics,* Andover, Cengage Learning.
ZAKI, J. & MITCHELL, J. P. 2013. Intuitive Prosociality. *Current Directions in Psychological Science,* 22**,** 466-470.

6 Appendix

Appendix A – Probit and logit tables

Table 3.1.2A Marginal effects on cooperation in the lab pilot. Probit estimation

	No controls	Separate effects for player 1 and 2 (main)	Controlling for comprehension	Controlling for experience
Deliberation	0.0602 (0.0698)	0.1539 (0.0964)	0.0800 (0.2051)	0.2570* (0.1440)
Player 1		0.0880 (0.0994)	0.0737 (0.2121)	0.2808** (0.1352)
Player 1 * Treatment		-0.1910 (0.1377)	0.0668 (0.2773)	-0.3217* (0.1895)
Comprehension			-0.1006 (0.1792)	
Deliberation * Comprehension			0.0820 (0.2323)	
Player 1 * Comprehension			0.0057 (0.2392)	
Player 1 * Deliberation * Comprehension			-0.3529 (0.3170)	
Experience				0.1519 (0.1387)
Deliberation * Experience				-0.1929 (0.1924)
Player 1 * Experience				-0.4080** (0.1927)
Player 1 * Deliberation *				0.2245

Experience				
				(0.2789)
N	187	187	187	187
R^2				

Standard errors in parentheses
Data from the Pencil-and-paper study in the DIGSSCORE lab
* p<0.1, ** p<0.05, *** p<0.01

Table 3.1.2B Marginal effect of deliberation on cooperation. Logit estimation

	No controls	Separate effects for player 1 and 2 (main)	Controlling for comprehension	Controlling for experience
Deliberation	0.0602	0.1537	0.0785	0.2583*
	(0.0698)	(0.0963)	(0.2017)	(0.1459)
Player 1		0.0886	0.0723	0.2815**
		(0.1001)	(0.2085)	(0.1374)
Player 1 * Treatment		-0.1908	0.0651	-0.3215*
		(0.1374)	(0.2724)	(0.1887)
Comprehension			-0.1009	
			(0.1781)	
Deliberation * Comprehension			0.0828	
			(0.2292)	
Player 1 * Comprehension			0.0076	
			(0.2364)	
Player 1 * Deliberation * Comprehension			-0.3523	
			(0.3129)	
Experience				0.1550
				(0.1425)
Deliberation * Experience				-0.1951
				(0.1930)
Player 1 * Experience				-0.4106**
				(0.1957)

Player 1 * Deliberation * Experience				0.2235
				(0.2829)
N	187	187	187	187
R^2				

Standard errors in parentheses
Data from the Pencil-and-paper study in the DIGSSCORE lab
* p<0.1, ** p<0.05, *** p<0.01

Table 3.1.3A Marginal effects on understanding in the lab pilot. Probit estimation

	Probit ME (SE)
Player 1	-0.0848
	(0.0909)
Deliberation	-0.1153
	(0.0885)
Player 1 * Deliberation	0.0848
	(0.1244)
N	187
R^2	

Standard errors in parentheses
Controlling for heterogeneity in comprehension of the payoff-structure.
* p<0.1, ** p<0.05, *** p<0.01

Table 3.1.3B Marginal effects on understanding in the lab pilot. Logit estimation

	Logit ME (SE)
Player 1	-0.0869
	(0.0935)
Deliberation	-0.1171
	(0.0905)
Player 1 * Deliberation	0.0869
	(0.1252)
N	187
R^2	

Standard errors in parentheses
Controlling for heterogeneity in comprehension of the payoff-structure.
* p<0.1, ** p<0.05, *** p<0.01

Table 3.2.1A Marginal effects on cooperation for subjects in the NCP. Probit estimation.

	No controls	Separate effects for each player group
Deliberation	0.0479	0.0479
	(0.0295)	(0.0514)
Player 1		-0.0195
		(0.0498)
Player 2B		-0.2079***
		(0.0498)
Player 1 * Deliberation		-0.0024
		(0.0706)
Player 2B * Deliberation		-0.0186
		(0.0727)
N	1126	1126
R^2		

Standard errors in parentheses
Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y (i.e. player 2 who knows that player 1 has chosen the yellow box) is the baseline.
Effects measured with Probit. Only data from NCP
* p<0.1, ** p<0.05, *** p<0.01

Table 3.2.1B Marginal effects on cooperation for subjects in the NCP. Logit estimation.

	No controls	Separate effects for each player group
Deliberation	0.0479	0.0481
	(0.0295)	(0.0515)
Player 1		-0.0194
		(0.0496)
Player 2B		-0.2063***
		(0.0495)
Player 1 * Deliberation		-0.0026
		(0.0707)
Player 2B * Deliberation		-0.0188

		(0.0727)
N	1126	1126
R^2		

Standard errors in parentheses
Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y (i.e. player 2 who knows that player 1 has chosen the yellow box) is the baseline.
Effects measured with Logit. Only data from NCP
* p<0.1, ** p<0.05, *** p<0.01

Table 3.2.4A Marginal effects on propensity to drop out. Probit estimation

		Probit ME (SE)	Probit ME (SE)
Player control	1	-0.0210	
		(0.0232)	
Player treatment	1	0.0106	
		(0.0202)	
Player control	2B	0.0075	
		(0.0206)	
Player treatment	2B	0.0251	
		(0.0200)	
Player treatment	2Y	0.0122	
		(0.0203)	
Treatment			0.0195*

Standard errors in parentheses
Measuring the impact of treatment, and on specific player group, on propensity to drop out. In the regression on all player groups, player 2Y in the control group is the baseline. Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y is player 2 who knows that player 1 has chosen the yellow box).
* p<0.1, ** p<0.05, *** p<0.01

Table 3.2.4B Marginal effects on propensity to drop out. Logit estimation

		Logit ME (SE)	Logit ME (SE
Player control	1	-0.0225	
		(0.0252)	
Player treatment	1	0.0108	
		(0.0206)	
Player control	2B	0.0076	
		(0.0211)	
Player treatment	2B	0.0248	
		(0.0201)	
Player treatment	2Y	0.0124	
		(0.0206)	
Treatment			0.0197
			(0.0120)
N		1173	1173
R^2			

Standard errors in parentheses
Measuring the impact of treatment, and on specific player group, on propensity to drop out. In the regression on all player groups, player 2Y in the control group is the baseline. Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y is player 2 who knows that player 1 has chosen the yellow box).
* p<0.1, ** p<0.05, *** p<0.01

Table 3.2.5A Effect of deliberation treatment on log of response time in seconds. OLS estimation

	Player 1	Player 2Y	Player 2B
Deliberation	0.5915***	0.6150***	0.4531***
	(0.0943)	(0.1204)	(0.0673)
_cons	4.4216***	4.4917***	4.5467***
	(0.0661)	(0.0869)	(0.0463)
N	419	376	374
R^2	0.086	0.065	0.109

Standard errors in parentheses
Shows the effect of the deliberation treatment on the natural log of response time. For the control group, I consider the time taken to make the decision between the two boxes. For the treatment group, I consider the total time spent on both the written justification and the decision between the two boxes. Data from the NCP
* p<0.1, ** p<0.05, *** p<0.01

Table 3.3.1A Marginal effects on cooperation. Pooled analysis with probit

	No controls	Separate effects for each player group
Deliberation	0.0497*	0.0705
	(0.0272)	(0.0455)
Lab	-0.1859***	-0.2470***
	(0.0386)	(0.0381)
Player 1		0.0016
		(0.0446)
Player 2B		-0.1944***
		(0.0478)
Player 1* Deliberation		-0.0402
		(0.0631)
Player 2B* Deliberation		-0.0412
		(0.0686)
N	1313	1313
R^2		

Standard errors in parentheses
Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y (i.e. player 2 who knows that player 1 has chosen the yellow box) is the baseline.
Data from both the NCP and the lab experiment
* p<0.1, ** p<0.05, *** p<0.01

Table 3.3.1B Marginal effects on cooperation. Pooled analysis with logit

	No controls	Separate effects for each player group
Deliberation	0.0496*	0.0706
	(0.0272)	(0.0456)
Lab	-0.1859***	-0.2453***
	(0.0387)	(0.0378)
Player 1		0.0013
		(0.0445)
Player 2B		-0.1928***

		(0.0474)
Player 1* Deliberation		-0.0402
		(0.0631)
Player 2B* Deliberation		-0.0415
		(0.0685)
N	1313	1313
R^2		

Standard errors in parentheses
Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y (i.e. player 2 who knows that player 1 has chosen the yellow box) is the baseline.
Data from both the NCP and the lab experiment
* p<0.1, ** p<0.05, *** p<0.01

Table 3.3.2A Marginal effects on cooperation. Comparison estimated with probit

	Pooled	NCP	Lab
Deliberation	0.0705	0.0479	0.1539
	(0.0455)	(0.0514)	(0.0964)
Lab	-0.2470***		
	(0.0381)		
Player 1	0.0016	-0.0195	0.0880
	(0.0446)	(0.0498)	(0.0994)
Player 2B	-0.1944***	-0.2079***	
	(0.0478)	(0.0498)	
Player 1* Deliberation	-0.0402	-0.0024	-0.1910
	(0.0631)	(0.0706)	(0.1377)
Player 2B* Deliberation	-0.0412	-0.0186	
	(0.0686)	(0.0727)	
N	1313	1126	187
R^2			

Standard errors in parentheses
Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y (i.e. player 2 who knows that player 1 has chosen the yellow box) is the baseline.
Data from both the NCP and the lab experiment
* p<0.1, ** p<0.05, *** p<0.01

Table 3.3.2B Marginal effects on cooperation. Comparison estimated with logit

	Pooled	NCP	Lab
Deliberation	0.0706	0.0481	0.1537
	(0.0456)	(0.0515)	(0.0963)
Lab	-0.2453***		
	(0.0378)		
Player 1	0.0013	-0.0194	0.0886
	(0.0445)	(0.0496)	(0.1001)
Player 2B	-0.1928***	-0.2063***	
	(0.0474)	(0.0495)	
Player 1* Deliberation	-0.0402	-0.0026	-0.1908
	(0.0631)	(0.0707)	(0.1374)
Player 2B* Deliberation	-0.0415	-0.0188	
	(0.0685)	(0.0727)	
N	1313	1126	187
R^2			

Standard errors in parentheses
Player 2B plays second move and knows that player 1 has chosen the blue box (defected) Player 2Y (i.e. player 2 who knows that player 1 has chosen the yellow box) is the baseline.
Data from both the NCP and the lab experiment
* p<0.1, ** p<0.05, *** p<0.01

Appendix B – Instruction sheet for the lab pilot

Instrukser

Vi ønsker å forstå hvordan folk oppfører seg i økonomiske valgsituasjoner og vil be deg om å ta én anonym beslutning. Ingen kan finne ut hvilken beslutning du tar.

Du og en annen tilfeldig person kan velge mellom gul eller blå boks. Betalingen din avhenger av valgene som tas:

Hvis begge velger den gule boksen får dere 210 kr hver. Hvis du velger blå og den andre velger gul får du 300 kr og den andre får 100. Hvis begge velger den blå får dere 120 kr hver.

Uansett hva den andre personen velger, tjener du mest på å velge den blå boksen.

Appendix C - Instruction sheet for the Norwegian Citizen panel

Vi er opptatt av å forstå hvordan folk oppfører seg i økonomiske valgsituasjoner. Vi vil derfor be deg om å ta én anonym beslutning hvor du kan trekkes til å motta penger.
Du og en annen tilfeldig person kan velge mellom en gul eller blå boks.

Hvis begge velger den gule boksen får dere 1000 kroner hver. Hvis du velger blå og den andre gul får du 1500 kroner og den andre får 0 kroner. Hvis begge velger blå får dere 500 kroner hver. Uansett hva den andre gjør tjener du mest på å velge blå.

Vi trekker ut ti personer som mottar penger.

Appendix D – Decision sheet for player 1 in the control group

Vennligst ta en beslutning:

Sett ring rundt den beslutningen du ønsker å ta

Appendix E – Decision sheet for player 1 in the treatment group

Før du tar beslutningen ønsker vi at du begrunner valget ditt.

Vennligst ta en beslutning:

Sett ring rundt den beslutningen du ønsker å ta

Appendix F – Decision sheet for player 2 in the control group

Den andre personen har valgt den gule (blå) boksen.

Vennligst ta en beslutning:

Sett ring rundt den beslutningen du ønsker å ta

Appendix G – Decision sheet for player 2 in the treatment group

Den andre personen har valgt den gule (blå) boksen.

Før du tar beslutningen ønsker vi at du begrunner valget ditt.

Vennligst ta en beslutning:

Sett ring rundt den beslutningen du ønsker å ta

Appendix H – Control questions in the lab pilot

1. **Hvilket kjønn er du?**

☐ Mann ☐ Kvinne

2. **Har du tidligere deltatt på slike eksperimenter?**

☐ Ja ☐ Nei

Vi ber deg til slutt svare på noen spørsmål som sjekker din forståelse av beslutningssituasjonen du har vært i:

3. **Den andre har valgt den gule boksen. Hvilket valg må du ta hvis du skal tjene mest mulig penger?**

☐ Gul ☐ Blå

4. **Den andre har valgt den blå boksen. Hvilket valg må du ta hvis du skal tjene mest mulig penger?**

☐ Gul ☐ Blå

5. **Hvilket alternativ gir høyest samlet betaling?**

☐ Begge velger gul

☐ Begge velger blå

☐ En velger gul og en velger blå