CHECKING AND TEMPORAL EMBEDDEDNESS IN INDEFINITELY AND FINITELY REPEATED ASYMMETRIC TRUST GAMES

ISBN 978 90 393 5279 3

This book was typeset by the author using LaTeX.

CHECKING AND TEMPORAL EMBEDDEDNESS IN INDEFINITELY AND FINITELY REPEATED ASYMMETRIC TRUST GAMES

Checken en temporele inbedding in onbepaald en eindig herhaalde asymmetrische vertrouwensspellen
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus,
prof.dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
donderdag 25 februari 2010 des middags om 4.15 uur

door

Thomas Dirkmaat
geboren op 20 december 1977 te Nijmegen

Promotoren:

Prof.dr. S. Rosenkranz
Prof.dr. P.P. de Gijsel

Voor mijn ouders

Acknowledgement

For me the process of writing a PhD-thesis has been an exciting journey of scientific exploration. Exploration is not possible without inspiration: people who sketch visions of the unknown and who stimulate you to keep asking questions. Like most of us, I have been inspired by others. From the time that I was studying in Utrecht I would like to mention Ton Hol, Gerrit de Geest and Giovanni Russo. Next I would like to thank Ardine de Wit who gave me my first research position at the RIVM. The two months I worked there, strengthened my feeling that I should apply for a PhD-position.

This dissertation would not have been written without the trust and support of my supervisors Peter de Gijsel and Stephanie Rosenkranz.

Peter, I would like to thank you for giving me a PhD position at the Utrecht School of Economics. Although during the first years your responsibilities as dean didn't leave much time for discussions, the discussions were lively and stimulating when we did manage to meet. Your comments always forced me to think carefully about what I was doing, which is something I really learned to appreciate. What I value most about you is your absolute commitment. When you commit yourself to a project, the people involved know that you will be there all the way. Given the difficult times I faced in my personal life, this has been very valuable to me.

Stephanie, you were always there for me when I had questions or needed guidance. Even in the final two years when you became director of the undergraduate school and you were very busy, you invited me to your home in the evenings to talk about my thesis. I owe my knowledge of game theory to you. I look back with fond memories on running experiments with you in the ELSE laboratory and how exited we got to see the data literally appear on our computer screen. Your meticulous readings of the manuscript and the many suggestions you wrote in the margin, significantly improved this thesis. For this I am very grateful. Finally, I want to thank you for listening to me and just being there during the difficult times I faced in 2006 and

2008. It still means a lot to me.

Next I would like to thank the members of the assessment committee: Werner Raub, Jan Potters, Arjen van Witteloostuijn, Patrick Schmitz and Gábor Péli for the careful reading of the manuscript and for their constructive comments.

Regarding the experimental design, programming the treatments and the data analysis I received very useful suggestions and practical help from Vincent Buskens. My trips to his office at the Van Unnik gebouw always proved to be very useful and our conversations a source of inspiration.

With respect to the data analysis I also must mention Rob Alessie, the best econometrics teacher that I know who also has a great sense of humour. He helped me to find the appropriate statistical models on several occasions. I also would like to express my thanks to the other members of the chair of econometrics, Wolter Hassink, Adriaan Kalwij and Yolanda Grift, who helped me with my statistics questions.

On the theoretical side I received support from Kris de Jaegher who was so kind as to take a critical look at my model for finitely repeated asymmetric trust game. I would also like to thank my former colleagues from the chair of theoretical microeconomics, Jeroen van de Ven and Jurjen Kamphorst, for the advice and support they gave me in an earlier stage of this project.

When it comes to practical support I should mention Franziska Schuetze. As the student assistant of the chair of theoretical microeconomics she was a great help with running the experiments and keeping track of the references. For her secretarial support I would like to thank Janet Sartorius.

Working at the Utrecht School of Economics has been a pleasant experience for me. I would like to thank all my colleagues for providing a nice working atmosphere. In particular I would like to mention Utz Weitzel for giving me tips on data analysis and talking about movies, Jaap Bos for being my LaTeX-buddy (together with Vincent) and Piet Keizer for his useful UCU teaching insights. I also would like to mention my fellow PhD-students. I have many fond memories of the PhD-drinks and dinners!

I would like to address two of my fellow PhD-students individually, because they were not only colleagues but are also friends. Metodij, we shared an office for more than 2 years. I got to know you as a disciplined researcher, but also as someone with a broad interest in international developments. I enjoyed the discussions we had about economic and political developments in the Netherlands, Macedonia and the rest of the world.

Bastian, as direct colleagues we were able to share all our hopes and frustrations with writing a PhD-thesis. I got to know you as an passionate researcher, a game enthusiast and a good cook. Regardless of whether the topic was board, card, trust or network games, I always enjoyed our discussions about strategies. Thank you, guys, for being my “paranymphs!

Given the fact that I studied a combination of Economics and Law, I felt truly at home in the faculty of Law, Economics and Governance. The last 2.5 years I had an office at the Boothstraat 6. The Wiardians have my gratitude for accepting me in their group. I must admit that I miss the afternoon tea breaks and the discussions with the legal research master students.

I also enjoyed my work for the PhD-council. I would like to thank Merel, Nora, Liesbeth, Marianne, Hana, Reile and Eva for a co-operative and relaxed atmosphere during our meetings. It was an honour to represent the faculty’s PhD-students on the Board of Graduate Studies, and I would like to thank its members for the lively and interesting discussions we have had.

My parents have always supported and motivated me to pursue my studies and I will always love them for that. To them I dedicate this book.

Finally, I would like to thank my family members, friends and everyone else who gave me their support and never fading interest during the last 6 years. You all asked me about the topic of my thesis. I usually answered this question by saying that my research is about long-term trust relations between people, most of the time followed by some examples and /or more specific information about information asymmetry and experiments. It was a bit surprising to me that the majority of the people I spoke to didn’t believe that this is a topic an economist would normally study (psychology was the clear favourite). I hope that browsing through this book will convince you otherwise.

Thomas Dirkmaat
Utrecht, December 2009

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis presents economic research on trust. Trust is a fuzzy concept and – perhaps because of this – for many people it is a fascinating research object. Given this property, it is important to be clear about how trust is defined and from what perspective it is studied. In Section 1.1 the position of this thesis in the trust literature will be made clear and in Section 1.2 our working definition of trust is presented.

To get a quick impression about the topic of this thesis, the following paragraph cites an example of the type of trust problem analysed in this dissertation. A more formal presentation of the topic of this thesis can be found in Sections 1.3-1.5. In Section 1.6 the main research questions are presented, together with an outline of the thesis. Finally, Section 1.7 gives an overview of the main results.

Consider the following example: John has inherited some money, which he decides he would like to invest. Unfortunately, John does not know how he should invest his money in order to get the best return on his investment. In order to deal with this problem, John decides to hire William, who is a financial expert. Together they decide that William will invest John's money for him. Note that, although William decides how the money is invested, John owns the assets (stock, bonds, etc.). Each year on January 1^{st}, William presents John with an overview of the assets in his investment portfolio at their current value. At this moment in time, John can decide to continue his relationship with William or have his investment portfolio liquidated.

John needs to trust that William will apply all his skills and expertise to ensure that John gets the best return on his investment. William needs to decide whether he is going to exert effort or not. When William chooses to exert effort, it is very likely that John will have a good return on his investment. When, on the other hand, William is lazy, it is most likely that John will face a lower return or perhaps even a loss on his investments. John does not know whether William is a lazy or hard-working person. Given that exerting effort is costly, John might even assume that William has an incentive to be lazy. Only when John expects that William will exert effort is it rational for John to trust William. Given that John does not know whether William will exert effort or not, John's decision to hire William to invest his money for him can be described as a trust problem.

In this specific example, John faces an additional information problem. Given his lack of knowledge about the financial markets, he finds it difficult to interpret the information he receives from William. When on his annual overview he sees a good return on his investment, John does not know if this is due to the efforts of William or due to favourable market conditions. In other words, perhaps even the laziest financial expert should have been able to make a profit. Alternatively, when John makes a loss on his investments, it might be due to unfavourable market conditions or William's lack of effort. Therefore, on the one hand John does not know whether William will exert effort or not. On the other hand, he cannot directly observe whether William has been trustworthy and, hence, he does not know whether he should continue his collaboration with William.

John would prefer a situation where he could check whether William has been trustworthy or not. In order to deal with the last problem, John could, for example, consult an additional financial expert, who would judge if the return on investment achieved by William was reasonable given the market conditions of the previous year (a second opinion). The advantage of checking out whether William has been trustworthy is that John is now in a much better position to judge whether he should continue his relationship with William.

This particular type of trust problem – where the person who places trust cannot directly observe what his partner has done, together with a checking option that can ex post provide him with this information – forms the topic of this thesis.

1.1 Perspectives on trust

In the broader trust literature, many researchers from different backgrounds such as sociology, psychology, political science, and economics are active. Not entirely surprisingly, these researchers use different methods and approaches when studying the concept of trust. On the one hand, this leads to a broad range of interesting insights; on the other, it becomes more difficult to arrive at a common understanding of the phenomenon under discussion.

Some authors give an overview of the broader trust literature and make an attempt to come to a unifying trust concept (see Lane (2000), Korczynski (2000) and Nooteboom (2002)). It is beyond the scope of this thesis to do something similar. A brief outline of the trust literature will be given in order to make clear to which sub-branch this thesis contributes.

The broader trust literature can be roughly divided into two branches. The first studies trust at the *macro-level*. Within the *macro-level* branch, roughly two sub-branches can be distinguished. In the first sub-branch, trust in abstract systems and/or social institutions is studied.[1] The work of sociologists such as Luhmann (1979), Zucker (1986) and Giddens (1990) is well-known in this branch. The second sub-branch deals with the concept of societal trust. Within this concept trust is, next to norms and networks, a component of social capital, the level of which within a society determines how successful that society will be (see Putnam et al. (1993) and Fukuyama (1995a)). Political scientists such as Levi and Stoker (2000) discuss, for example, whether a trustworthy government can promote the economic productivity of a country. Conversely, economists place an emphasise on the effect of trust (as a part of social capital) on the economic performance of countries. Some important studies in this field are authored by La Porta et al. (1997), Knack and Keefer (1997), Zak and Knack (2001) and Bornhorst et al. (2005). Their main findings are that trust can explain differences in the economic performance of countries and that trust levels themselves differ between countries due to cultural differences. In relation to the latter observation, it is often mentioned that in northern European countries trust levels are higher compared to southern European countries.

Some authors are critical about the concept of societal trust. According to Cook et al. (2005), co-ordination and state regulation have gained importance over trust given the long-term change from small communities to mass urban complexes. Hence, Cook et al. (2005) emphasise the roles of institutions and other arrangements as determinants of a successful society. In their opinion, trust is important in interpersonal contexts, where it should be regarded as a complement, not a substitute, for organisational arrangements that make co-operation possible.

This thesis must be positioned in a second extensive branch of literature

[1] The monetary system and product markets are examples of abstract systems, while culturally determined norms are an example of social institutions.

that studies trust at the *micro-level*.[2] In this branch, trust between individual agents is studied. The main difference between the studies in this branch lies in the approaches that are used, including their underlying assumptions. Although approaches can be roughly linked to disciplines, the boundaries are far from clear cut. In general, economists take rational choice theory as their starting point[3], while sociologists and psychologists such as Kramer (1999) and Lane (2000) delve more into the roles of common values, moral considerations, common cognitions and the role of emotions to explain trust. In this thesis, trust is studied from an economic perspective, and game theoretical models including their underlying rationality assumptions will be used to come to testable hypotheses.

The economic literature that studies trust at the *micro-level* focuses on the type of interpersonal trust described in the problem faced by John and William. In our example, John did not know whether he could trust William to invest his money for him. In the trust literature, John is known as a 'trustor' and William as a 'trustee'. In general, a trust problem arises due to the fact that in a trust relation the interests of the trustor and the trustee are not perfectly aligned, while at the same time the trustor has limited knowledge about the trustee's intentions. Whether it is the purchase of goods or services, the hiring of labour, obtaining capital, or putting a project out to tender, the trustor will have to make a decision about the level of discretion he allows the trustee to have.

In economics, it is usually assumed that voluntary exchange is welfare improving. When William realises a good return on the money he invested on behalf of John, and John gives William a decent reward for William's effort, both John and William can benefit from their collaboration. When John trusts William and William honours John's trust, i.e. they overcome the trust problem, they are both better off. Their collaboration is a Pareto improvement.

Trust can facilitate co-operation and, consequently, increase welfare. From this viewpoint, trust is a desirable 'commodity'. Given the fact that trust enables relations and reduces transaction costs, some authors see trust

[2]Some authors such as, for example, Luhmann (1988) make a distinction between confidence and trust. He states that when trust is placed in institutions and social-economic systems (macro-level), it might be better to speak of confidence, while the word 'trust' should be reserved for interpersonal relations (micro-level).

[3]Rational choice theory is also used by other disciplines; take, for example, the work of the rational choice sociologists (for an review study see Buskens and Raub (2008)). But also trust as encapsulated self-interest, introduced by Hardin (2002) (a political scientist), can be considered part of the rational choice approach.

as a lubricant for economic activity, which is reflected in the following quote from Fukuyama (1995b): “If people who have to work together in an enterprise trust one another ... doing business costs less”.[4]

This does not mean that the existence of trust is always beneficial to society. Trust is welfare-enhancing when transactions are not only beneficial to the trustor and the trustee, but also have no negative external effects for society. More precisely, in order to realise a Pareto improvement, the trustor and the trustee should be willing to compensate society for occurring negative external effects out of their profits.

Nevertheless, even if trust is beneficial in the short-run it can have harmful effects in the long-run, as is effectively illustrated in van Witteloosstuijn and van Wegberg (2006). They argue that although trust can protect the alliance between two business partners, it can limit the possibility of entering a new partnership because this might harm the existing relationships. Hence, in dynamic markets, firms face a trade-off between maintaining relationships with partners they know to be trustworthy or entering new and possibly profitable partnerships with firms that still need to prove their trustworthiness.

In this dissertation, we follow most authors in this field by focusing on the positive side of co-operation. Hence, in this thesis it is assumed that trust is good, because it enables relations and reduces transaction costs.[5]

Within the *micro-level* branch of the trust literature we can, roughly speaking, distinguish two sub-branches. The first sub-branch focuses on the foundations, or determinants, of trust. These studies try to answer the question of how trustful and trustworthy behaviour can be explained. This usually happens by relaxing the rationality assumptions or by pointing out that the behaviour of the trustor and the trustee might not only be driven by egoistic preferences, but also by other-regarding, or social, preferences. Examples of studies in this field are: Cox (2004) and Bohnet and Zeckhauser (2004). Cox (2004) introduces an experimental design that enables him to disentangle trustfulness from altruism and trustworthiness from reciprocity. Bohnet and Zeckhauser (2004) use a comparable design to show that the decision to trust someone is not just a risky bet, but that it involves betrayal cost.

The second sub-branch, to which this thesis wishes to contribute, investigates possible solutions to the trust problem. The trustor would prefer

[4] The idea that trust can reduce transaction cost is not new – see, for example, Arrow (1974).

[5] In our analysis of trust, as we focus on the relations enabling aspect of trust, the transactions costs reducing aspect will be ignored.

the situation where the trustee's interests are aligned with his own. Although several solutions have been discussed in the literature, they all have in common that they ask for the trust problem to be socially embedded. Buskens and Raub (2008) make a distinction between three categories of social embeddedness: institutional embeddedness[6], network embeddedness and dyadic, or temporal, embeddedness.[7]

Institutions can help align the interests of the trustor and the trustee and, hence, solve the trust problem. A perfect institution can be seen as a safeguard, because it protects the trustor against the trustee's opportunistic behaviour. Hostages are one example of how institutions can facilitate co-operation (Raub and Keren (1993) and Raub (2004)). A hostage is something of great value to the trustee, which he gives to the trustor. The trustee will lose the hostage when he abuses the trustor's trust. With a hostage the trustee can signal that he is trustworthy, because the use of a hostage enables the trustee to commit himself to his promise of not betraying the trustor's trust. Another example of an institution is contract law. Contracts are, however, most of the time imperfect, which means that not all possible eventualities can be taken into account. Consequently, trust will always play a role.

Relations with more than two actors are mostly studied using network analysis. Being in a network can be advantageous for the actors in the network. When a trustee in a network abuses the trust of a trustor, this trustor can inform other trustors about the untrustworthy behaviour of the trustee. This makes it less likely that other trustors will co-operate with this untrustworthy trustee. Examples of studies on network embeddedness are Coleman (1988), Raub and Weesie (1990), Buskens (1999) and Buskens (2003).

When the trustor and the trustee have several encounters over a longer period of time, this can have an influence on their behaviour. When, for example, the trustee abuses the trustor's trust today, it is questionable whether the trustor will trust the trustee again tomorrow. The longer time horizon changes the dynamics of the trust problem, an effect known as 'temporal embeddedness'.[8]

---

[6]Institutional embeddedness must not be mistaken for trust in the trust problem solving capacities of institutions themselves. This type of trust is studied at the *macro-level* branch.

[7]For a general discussion on the relevance of embeddedness when analysing economic behaviour, see Granovetter (1985).

[8]Trust can, of course, also be studied under both temporal and network embeddedness; see, for example, Barrera (2005).

Recall from our example of John and William that once every year William should present John with an overview of his investment portfolio. In other words, the trust problem that John and William face is temporally embedded. This thesis forms a contribution to the literature on temporal embeddedness. Before we discuss temporal embeddedness more extensively, we will first transform the fuzzy concept of trust into a workable definition.

1.2 Defining trust

1.2.1 The trust problem

Dasgupta (1988) and Coleman (1990) discuss the elements of a trust problem as it can be encountered in interpersonal or inter-firm relations. Using their work as a basis, we come to the following three elements of the trust problem:

(1) The *trustor* is the person who can decide whether to co-operate with the trustee or not.

(2) The *trustee* is the person who receives trust from the trustor and needs to make the decision to honour or abuse it.

(3) When the trustor places trust in the trustee, he enters a state of *uncertainty*. Compared to the situation where he does not place trust, the trustor is worse off when his trust is abused, but he benefits when his trust is honoured.

The first two elements define the actors and their action set. This is a simplification from the reality whereby trust can be placed in more than one object at once. For instance, if someone buys a car, he can place trust in the manufacturer of the car, the local dealer and/or a specific car salesman.

The third element is crucial, because without *uncertainty* it is impossible to speak of trust. If the trustor knows the trustee's response with certainty, there is no need for trust.

Uncertainty must not be mistaken for *risk*. According to Knight (1921), risk is uncertainty that can be measured. We talk of risk when the set of possible outcomes is defined and the probability of an outcome is objectively known, like, for example, in the case of rolling a fair dice. Uncertainty refers to the situation where the set of possible outcomes is defined, but the probability of an outcome is an estimation or a *subjective probability*.

We argued that complete contracts do not exist, because it is impossible to cover every potential hazard. The situation where not only are the probabilities estimates, but also the set of possible outcomes is not closed is referred to as *radical uncertainty*.

In the previous section it was argued that voluntary exchange is welfare-increasing. While defining the trust problem we implicitly assumed that both trustor and trustee should be free individuals in mind and action. This means that the trustor cannot be forced to choose co-operation or no co-operation, and neither can the trustee be forced to honour or abuse the trustor's trust. It is impossible to speak of trust when either the trustor or the trustee is forced to make his decision under pressure. A good example of extreme pressure is reflected in the famous quote from the 1972 movie the Godfather: "I'll make him an offer he can't refuse". This quote reflects the Godfather's use of a highly immoral, and in most countries illegal, business method by forcing someone to sign a contract by holding a gun to his head.

1.2.2 Calculative and non-calculative trust

Although the trustor and the trustee are free to make their own decisions, this does not mean that they do not prefer one action above the other. In the broader trust literature it is argued that the trustor's decision can be based on both calculative and non-calculative motivations, cited, for example, by Zucker (1986) and Lane (2000).

In the literature, different forms of non-calculative trust[9] are discussed. Firstly, trust can arise when the presence of a common understanding is assumed, or, as Garfinkel (1967) describes it: "this ability [to act rationally] depends upon the person being able to take for granted, to take under trust, a vast array of features of the social order". For example, Macneil (1980) argues that we need to trust each other to give the same meaning to the words we use – contracts will lose their usefulness when we need to define every word in advance. Another example comes from Parsons (1951), who argues that trust cannot develop unless individuals have shared common values. The trustor expects the trustee to meet his social obligation and exercise responsibility. Trustworthiness and morality are interwoven in this context. However, it is questionable whether, in advanced and highly differentiated societies, common values and norms that can support trust exist.

Secondly, the trustor might not want to think of the possibility that the trustee will abuse his trust. Nooteboom (2002) links this to cognitive dissonance: "one does not want to contemplate the possibility of opportunism

[9]Non-calculativeness does not necessarily lead to blind trust as long as trust is not unconditional – see Nooteboom (2002).

on the part of friends or family".

Thirdly, reflection ceases to take place in the case of naivety. Here individuals simply fail to label a situation as potentially risky.

Fourthly, as suggested by authors such as Uzzi (1997) and Nooteboom (2002), trusting behaviour can be based on simple heuristics; trust is placed on the basis of available knowledge, routines and perhaps even instinct until evidence to the contrary appears. For instance, when placing trust in the past has worked out well in a specific situation, it might become routine practice to place trust in the future in similar situations. Or, in the words of Nooteboom, trust becomes the default. However, when the trustor learns that his trust is being abused by the trustee, he will adopt a new strategy.

Calculativeness is commonly associated with the rational choice model of human behaviour (e.g. Williamson (1993), Lane (2000) and Nooteboom (2002)), which assumes that people are rational. This means that people have a complete and transitive preference ordering, and given this preference ordering they will maximise their utility. Hence, a trustee will only honour trust when this maximises his utility. It is assumed that the trustor knows this. A rational trustor will only co-operate if he knows the trustee prefers to honour his trust, but this outcome conflicts with our understanding of trust. Element (3) states that the trustor should enter a state of uncertainty. Williamson (1993), in his criticism on trust as risk as proposed by Coleman (1990), raises this point exactly, concluding that: "Calculative trust is a contradiction in terms".

Does this mean that the only true form of trust is non-calculative trust? The simple answer is no. When the trustor is uncertain about the trustee's choice, which itself is in the end determined by the trustee's preferences, trust can still be calculative. On the basis of extensive experimental evidence – see, for example, Roth and Kagel (1995), Camerer and Thaler (1995), Fehr and Gächter (1998), Rabin (1998) and Fehr and Fischbacher (2003) – utility functions have been specified that, next to egoistic preferences, allow for the existence of altruistic and reciprocal preferences (see Rabin (1993), Levine (1998), Fehr and Schmidt (1999) and Bolton and Ockenfels (2000)). The assumption that the trustor knows the trustee's preferences is more difficult to maintain when these social preferences are incorporated. When the trustor does not know the specification of the trustee's utility function, or in terms of the later chapters his 'type', he cannot predict with certainty what the trustee will do. This, however, does not mean that the trustor cannot come to a rational or calculative decision. The trustor can form a belief (subjective probability) about the distribution of trustee types, and on the basis of this he can calculate whether, in expected terms, placing trust should be preferred compared to not trusting the trustee. When the trustor does place trust on the basis of this rational inference, he has reasons to believe that his trust is justified. Although the trustor can place trust in the trustee, due to the highly subjective nature of the trustor's belief, uncertainty will be always present. In other words, the trustee might still choose abuse and, hence, element (3) remains satisfied.

Some authors (e.g. Nooteboom (2002)) argue that trust is never purely calculative, due to radical uncertainty. Whether a utility-maximising trustee prefers to honour or abuse the trustor's trust becomes less important in the light of radical uncertainty, because it can always be argued that for some unknown, unforeseeable reason he might change his behaviour. Given the infinitely small probability with which these unforeseeable events occur, it is impossible to include them in a calculation, therefore non-calculative trust is always present.

1.2.3 The trust game

The simplest model that captures the spirit of the trust problem is the binary trust game introduced by Dasgupta (1988), Camerer and Weigelt (1988) and Kreps (1990). It is a sequential game where the trustor moves first and needs to decide between *Co-operation* and *No co-operation*.[10] When the trustor chooses *No co-operation*, the game ends. When the trustor chooses *Co-operation*, the trustee needs to decide between choosing *Abuse* and *Honour*. In Figure 1.1, the extensive form of the trust game is presented. In the trust game the following pay-off structure is assumed:

$$C_1 > N_1 > S_1 \wedge A_2 > C_2 > N_2. \tag{1.1}$$

C_1, N_1 and S_1 and A_2, C_2 and N_2 represent the utility the trustor and respectively the trustee derive from the different outcomes of the game.

If $C_1 > N_1$, the relationship is beneficial to the trustor in the case where his trust is honoured, while if his trust is abused he is worse off, because $S_1 < N_1$.

Excluding uncertainty about the trustee's preferences for the moment, it is not difficult to find the pure strategy Nash equilibrium. For the trustee it is assumed that the gain A_2 from abusing is greater than the gain C_2 from honouring the trustor's decision to co-operate. The trustee prefers to choose

[10] To highlight strategies in the text, we will write strategies in *italic*. Note that in the isolated encounter binary trust game actions and strategies are the same.

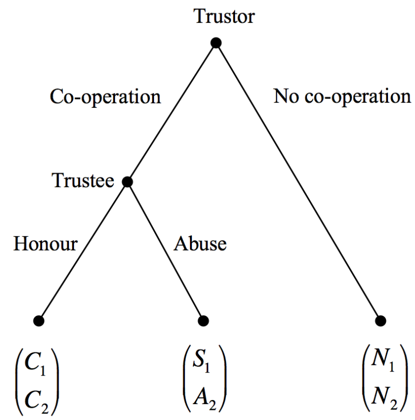


Figure 1.1: The trust game of complete information.

Abuse whenever he has the chance. A rational trustor anticipates that the trustee will respond with *Abuse* and therefore will choose *No co-operation*, because $N_1 > S_1$. The outcome of the trust game is that the trustor and the trustee will not co-operate. For this reason the trust game is often labelled as a social dilemma game.[11] [12]

If, alternatively, the trustor prefers to choose *Co-operation*, this would be a violation of element (3) of the trust problem, because as long as we maintain the assumption that the trustor knows the trustee's preferences, he will only choose *Co-operation* when he knows with certainty that the trustee will choose *Honour*.

In order to fully capture the spirit of the trust problem, the trust game must allow for uncertainty about the trustee's preferences. In the experimental literature on isolated encounter trust games it has been shown that contrary to the equilibrium strategy a significant number of the trustors and trustees do choose *Co-operation* and *Honour*, respectively. Well-known contributions in this field are the experiments of Snijders (1996) and Snijders and Keren (1999), who test the binary trust game, and Berg et al. (1995),

[11] The best known example of a social dilemma game is the prisoner's dilemma.

[12] In the binary trust game it is assumed that the trustor and the trustee cannot communicate with each other, but even if they could, it would be mere cheap talk. Assume the trustee communicates that he will choose *Honour* when the trustor chooses *Co-operation*. However, as soon as he needs to fulfil his promise a rational trustee will still prefer to choose *Abuse*. Hence, without credible commitment the promise is nothing more than cheap talk.

who introduce and experimentally test the investment game.[13] For a detailed review on isolated encounter trust game experiments, we refer to Camerer (2003). Given the extensive experimental evidence on social preferences, the most obvious way to adjust the trust game is to allow for incomplete information about the trustee's pay-off, which can be done by assuming some underlying distribution of social preferences among the trustees. For example, that some trustees might have altruistic preferences, in the sense that they also care about the trustor's pay-off,[14], and due to this prefer to honour the trustor's trust. Other trustees might want to reciprocate the trustor's kindness to choose *Co-operation*, by responding with *Honour*.

Given the pay-off structure of the binary trust game, trustees solely motivated by selfish or egoistic preferences will prefer to choose *Abuse* in response to the trustor's decision to choose *Co-operation*, while trustees motivated by a broader set of preferences might be willing to choose *Honour*. Figure 1.2 reflects this idea. The trustor faces two types of trustees. He does not know what type of trustee he faces, but he has a subjective belief about the distribution of the types in the trustee population.[15] With $(1-P)$ he believes that he faces a trustee who is solely motivated by egoistic preferences. Meanwhile, P denotes his belief about the likelihood with which he faces a trustee who is prepared to choose *Honour* due to his social preferences. This latter group of trustees earns different pay-offs compared to trustees who are only motivated by selfish preferences. When they respond

[13]The investment game differs from the binary trust game with respect to the player's action set, which has the potential to be a continuum of choices. At the beginning of the game, both the trustor and trustee receive an endowment. Next, the trustor can decide to send any share he likes of his endowment to the trustee. The experimenter multiplies the amount sent with a certain factor. Finally, the trustee needs to decide how much money he is going to transfer back to the trustor from the combined wealth of his initial endowment and the amount he received from the trustor.

[14]Condition (1.1), is the standard pay-off structure assumed in the trust game. Next to this we will also assume that $C_1 + C_2 > S_1 + A_2 > N_1 + N_2$, i.e. co-operation between trustor and trustee is socially optimal compared to the situation where only the trustee picks the reward of 'co-operation'. The situation where they do not co-operate at all has the lowest combined pay-off. This additional condition is necessary to allow altruistic preferences to play a role on the side of trustees. This condition also solves another possible problem. Given other pay-off conditions it is possible that trustors, who have extremely altruistic preferences, might prefer choosing *Co-operation* above *No co-operation*, which can happen when the total welfare of choosing *No co-operation* is lower than that of *Co-operation*, followed by *Abuse*. Extreme altruism on the side of trustors violates element (3) of the trust problem presented above, which explicitly assumes that the trustor is worse off when his trust is abused compared to the situation where he chooses *No co-operation*.

[15]This subjective prior belief is assumed to be common knowledge and, hence, it is depicted as a move of nature that is observed by both players.

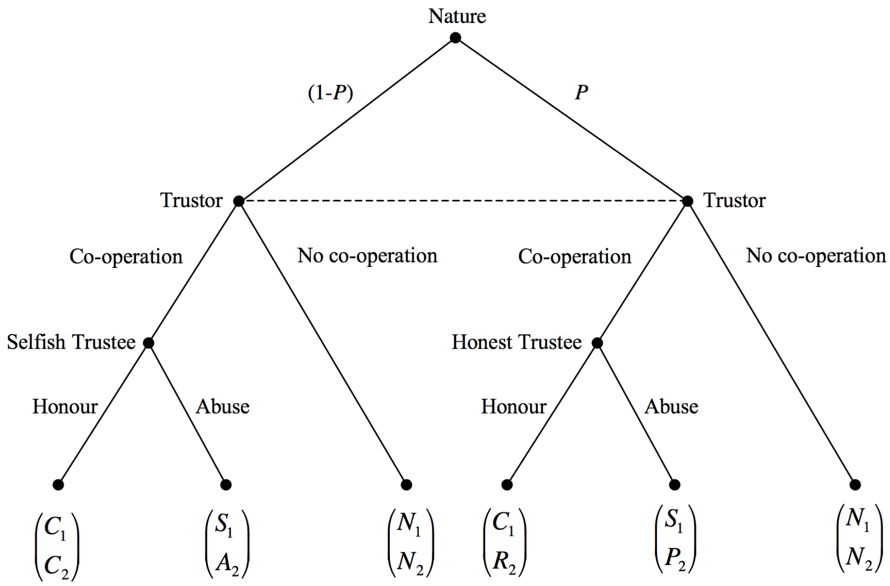


Figure 1.2: The trust game of incomplete information. In the sub-game to the left of the game tree, the selfish trustee prefers choosing *Abuse* above choosing *Honour*, because $A_2 > C_2$. In the sub-game to the right of the game tree, the honest trustee prefers to choose *Honour* above *Abuse*, because $R_2 > P_2$. Selfish and Honest trustees are defined in Section 1.2.4.

with *Honour*, they earn R_2 and when they choose *Abuse*, they earn P_2. It is assumed that $R_2 > P_2$.

When the trustor's belief is sufficiently high, he will choose *Co-operation*. This happens when:

$$P > \frac{N_1 - S_1}{C_1 - S_1}. \tag{1.2}$$

When incorporating social preferences in the binary trust game, the trustor enters a state of uncertainty when he chooses *Co-operation* and all three elements of the trust problem are present. The binary trust game with incomplete information about the type of trustee is regarded as a good representation of the trust problem, and will be the starting point for the models presented later in this thesis.

1.2.4 Trust, trustworthiness and definition of types

To come to a consistent trust lexicon, several terms will be defined. The first is trust:

Trust *is the trustor's decision to choose* Co-operation *in the case of a trust game of incomplete information, given his expectation that the trustee will respond with* Honour*.*

In this definition, trust is a combination of behaviour and disposition. It combines the act of choosing *Co-operation* with the subjective state of expectations that the trustor's trust will be honoured. When the trustor trusts the trustee, he not only accepts the uncertainty, but also he expects that the response of the trustee will be in his interest. In other words, he believes in a *positive outcome* from his interaction with the trustee. A situation can be labelled as 'mistrust' or 'distrust' when either behaviour or disposition is missing. So, when the trustor chooses *Co-operation* but does not believe in a positive outcome, or when he does believe in a positive outcome but does not choose *Co-operation*, it is improper to talk of trust.

The trustor's expectation is based on his beliefs about the trustee's response, which will be determined by the trustee's intentions and capability, respectively.[16] When it comes to the trustee's intentions, Nooteboom (2002) makes the following distinction: "A weak form is lack of care or commitment to act scrupulously, with effort and attention. This includes shirking. A strong form is opportunism or lack of benevolence, which entails theft, expropriation and extortion, or the threat of them". While the strong form is more about *whether* the trustee will honour trust, the weak form is more about *how* he will do it. The research presented in this thesis tries to answer the *whether* he will question. This is also reflected in the theoretical framework. The binary trust game does not allow for intentions of the weak form.

Next, we define trustworthiness:

Trustworthiness *is the extent to which the trustee chooses* Honour *after the trustor has chosen* Co-operation*.*

Trustworthiness, in this definition, reflects the trustee's behaviour, and, un-

[16] In this thesis, the trust game of incomplete information is the starting point of our analysis. In this game, capability is ignored, i.e. the trustee's decision is purely determined by his intentions.

like trust, no specific state of mind is required on the side of the trustee. Hence, in the binary trust game a trustee is considered trustworthy when he responds with *Honour* after the trustor has decided to choose *Co-operation*, regardless of his motivations for so doing. The word 'extent' points to repeated trust games, where the trustee might only choose *Honour* for a fraction of the entire game length, and to real-life trust situations where the trustee's response is not necessarily a binary decision but a continuum of possible actions. Given the discussion so far, it makes sense to distinguish two types of trustees.

Selfish trustee *is a trustee who prefers to choose* Abuse *in the isolated encounter trust game of incomplete information ($A_2 > C_2$).*

In the isolated encounter binary trust game of incomplete information, selfish trustees will choose *Abuse* whenever they have the chance. In finitely and in(de)finitely repeated encounters of the binary trust game, it can be in the self-interest of the trustee to choose *Honour* for several or all periods of the game, due to temporal embeddedness, as will be explained in Section 1.3. When this is the case, a selfish trustee can behave in a trustworthy manner.

Honest trustee *is a trustee who prefers to choose* Honour *in the isolated encounter binary trust game of incomplete information ($R_2 \geq P_2$).*[17]

The reason why the honest trustee prefers this decision can be explained by the fact that his preference set contains both egoistic and socially-oriented preferences. Of course, an honest trustee should also prefer to be honest in repeated encounters.

The main differences between the selfish and the honest trustee can be summarised as follows: Under the assumption that preferences are stable over time, an honest trustee will always act in a trustworthy way. A selfish trustee, by definition, will act in an untrustworthy manner in isolated encounter binary trust games. Whether he will act in an untrustworthy or trustworthy way in repeated trust games is determined by the parameters of the game and the equilibrium strategies these parameters support.

[17] Due to social preferences, the situation might arise that a trustee is indifferent between choosing *Honour* or *Abuse*. When this situation arises, it is assumed that the trustee will choose *Honour*.

A similar distinction can be made for trustors, although this time the distinction is not based on the trustor's preferences but on his beliefs.

Non-trustful trustor *is a trustor that prefers to choose* No co-operation *in the isolated encounter binary trust game of incomplete information, given his beliefs about the distribution of trustee types in the population* ($P < \frac{N_1 - S_1}{C_1 - S_1}$).

This group of trustors might choose *Co-operation* when they know that it is in the self-interest of the selfish trustee to choose *Honour*. This situation can arise in finitely and in(de)finitely repeated versions of the binary trust game, where, as mentioned above, it can be in the self-interest of the trustee to choose *Honour*.

Trustful trustor *is a trustor that prefers to choose* Co-operation *in the isolated encounter binary trust game of incomplete information, given his beliefs about the distribution of trustee types in the population* ($P \geq \frac{N_1 - S_1}{C_1 - S_1}$).[18]

Trustful trustors are even prepared to choose *Co-operation* when it is not in the self-interest of the selfish trustee to choose *Honour*, because they believe that a sufficient part of the trustee population can be labelled as honest. The behaviour of this group of trustors can also be explained using a non-calculative argument. For example, it might be that honest trustors are simply too naive to label the trust problem as a potentially risky situation.

Now that we have defined trust and concluded that the binary trust game of incomplete information is a good representation of the trust problem, it is time to return to temporal embeddedness as a possible solution to the trust problem.

1.3 Temporal embeddedness

1.3.1 Control and Learning

The repeated nature of games can create incentives that cannot be studied in isolated encounters (Mailath and Samuelson (2006)). When the same trustor and trustee meet each other several times in succession, this enables them to play different and potentially co-operative strategies, depending on

[18]Given the trustor's belief, the situation might arise that the trustor is indifferent between choosing *Co-operation* and *No co-operation*. When this situation arises, it is assumed that the trustor will choose *Co-operation*.

the parameters of the game. In this thesis, we refer to this effect as 'temporal embeddedness'.[19]

Games can be repeated in two different ways: A repeated game can have a certain end, like, for example, a temporary contract, and a game can have an uncertain end, like, for example, a permanent contract. In the latter case the player knows with certainty that it will end – he will retire one day – but he does not know when the relation will end exactly – he might retire early. This difference in time horizon is important and is also reflected in our daily behaviour. The best example might be life itself. We do know that we will die one day, we just don't know when. Now assume that from your birth onward you know with absolute certainty you are going to die in a car accident at the age of 40. Will you live your life in the same way?

Games with a certain end are analysed using finitely repeated games. While games with an uncertain end can be studied using infinitely repeated games, or the strategically equivalent indefinitely repeated games. The difference between finitely and indefinitely repeated games will be referred to in this thesis as the difference in horizon perspective.[20]

Buskens (1999), Buskens and Raub (2002) and Buskens and Raub (2008) distinguish two mechanisms that are responsible for the change in strategic behaviour when trust is temporally embedded: *learning* and *control*.[21]

Control reflects the idea that the trustor can impose a sanction on the trustee when he refrains from choosing *Co-operation*, and that the trustee understands that his current behaviour might have an influence on the trustor's future decision to impose this sanction. In other words, the trustee is aware of the fact that when he chooses *Abuse* today, it is very unlikely that the trustor will choose *Co-operation* tomorrow or, possibly, ever again.

Control is mostly studied in the context of in(de)finitely repeated social dilemma games of complete information.[22] When the sanction is triggered and how long the sanction will be applied depends on the strategy played. However, the basic idea of all strategies boils down to this: Because choosing *No co-operation* is an equilibrium strategy in the isolated encounter game, the trustor can ex ante make a credible threat that he will punish

[19]Some authors, such as Barrera (2005) and Buskens and Raub (2008), talk about dyadic embeddedness.

[20]Note that the number of repetitions has no effect on this theoretical distinction.

[21]These mechanisms are at work in situations where trust is temporal and/or network embedded. Given the focus of this thesis, we will only study the influence of control and learning on trust relations that are temporally embedded.

[22]Finitely repeated games of complete information have no future, because when backward induction is applied it can be easily shown that the first round is strategically equivalent to the final round.

the trustee by ceasing to choose *Co-operation* if he observes that his trust has been abused. Whether this threat will make the trustee prefer to choose *Honour* depends on the harshness of the trustor's sanction and the trustee's time preference. The latter means that the trustee is more likely to choose *Honour* the more he cares about future transactions. According to the Folk Theorem of non co-operative game theory, there are many equilibrium strategies that allow for a certain degree of co-operative behaviour, as long as these strategies satisfy the minmax criterion and the end of the game lies far enough away in the future.

Learning refers to the fact that the trustor can update his beliefs about the type of trustee he is interacting with, based on his observation of the trustee's past behaviour and the extent to which the trustee is aware of this. The combined effect of learning and control is mostly studied in finitely repeated social dilemma games of incomplete information.

The main theoretical contribution on reputation effects in finitely repeated games of incomplete information comes from Kreps and Wilson (1982a) and Kreps et al. (1982), with the development of the sequential equilibrium concept (Kreps and Wilson (1982b)). For later applications of the sequential equilibrium on the binary trust game, see Camerer and Weigelt (1988) and Bower et al. (1996). In Chapter 4, the sequential equilibrium will be discussed in more detail, but for now we will limit ourselves to the intuition which goes roughly as follows: The trustor does not know whether he faces an honest trustee, but he might have a belief about how likely it is that he does. When his initial belief is sufficiently high, the trustor will choose *Co-operation*. When the trustor observes that the trustee has chosen *Honour*, he will update his initial belief, which in this case will be strengthened. Consequently, the trustee has a good reason to choose *Co-operation* again. When, on the other hand, he observes that his trust has been abused, he will update his belief about the type of trustee he faces, and given this bad experience he will very likely choose *No co-operation* in the future. The selfish trustee is controlled in the sense that given the trustor's strategy, he might favour building up a reputation as being honest by choosing *Honour* for some rounds of the game. In the finitely repeated trust game of incomplete information, the selfish trustee has to take the future into account. Once the selfish trustee reveals himself as selfish, the trustor will sanction him by choosing *No co-operation* for the rest of the game.

In the indefinitely repeated game of complete information, the trustor's strategy determines when the sanction must be applied. In the finitely repeated trust game of incomplete information, however, this moment is determined by the updating rule which the trustor uses to update his belief, i.e. learn about the type of trustee he faces.[23]

It should, nevertheless, be noted that learning in repeated trust games differs from the idea of 'trust building' often found in the trust literature. With 'trust building' it is usually meant that at first the trustor is not willing to trust a trustee he doesn't know. When they have known each other for some time, the trustor will start to trust the trustee. Given good experiences, their bond of trust will become stronger over time. This idea has been brought forward by many authors, for example Granovetter (1985), Sabel (1993) and Hardin (2002). A good example is the way friendships evolve. Most people will not share their most intimate secrets with perfect strangers, while they might do so with a friend with whom they have developed a mutual bond of trust over a longer period of time. The same principle is at work in commercial relations. Gulati (1995) found that firms who formed alliances in the past are more likely to form alliances again, while at the same time the probability that the alliances are equity based decreases.[24]

'Trust building' also involves some form of learning in the sense that one strengthens one's trust relation through past experiences. Learning in the sequential equilibrium also involves some sort of 'trust building' in the sense that the trustor's belief about how likely it is that he faces a honest trustee will increase the longer the relation lasts. However, there is an important difference between the two. 'Trust building' suggests that in trust relations, where the trustor and trustee meet each other over a longer period of time, on average more co-operative behaviour can be observed. Learning, as described in finitely repeated trust games, suggests that on average co-operation decreases over time, due to the observation of untrustworthy behaviour. 'Trust building' therefore cannot be supported as an equilibrium strategy in the repeated binary trust game of incomplete information.[25] In this game the trustor does not receive any additional information about what type of trustee he faces when he does not choose *Co-operation*, and, hence, his initial belief remains unaltered. So, when his prior belief is not high enough to choose *Co-operation* in the first period, he will never prefer

[23]It is possible that in indefinitely repeated games of incomplete information equilibrium, strategies involving learning might exist.

[24]Gulati (1995) defines this safeguard as follows: "an equity joint venture had been created or that a firm had taken a substantive minority position in another with the intent of pursuing joint projects".

[25]In a repeated investment game of incomplete information, 'trust building' might be part of an equilibrium strategy. For example, the trustor sends a small amount to the trustee in the first period. Upon observing that the trustee has rewarded his trustfulness, by sending a large amount in return, the trustor sends a larger amount in the second period.

to choose *Co-operation* in later periods.

Control and learning can induce the trustee to choose *Honour* in repeated games. At first sight this looks like a violation of element (3) of the trust problem. This is not the case, though, because in repeated encounters many equilibrium strategies are possible, as will be made clear in Chapters 2 and 4. In other words, the trustee might prefer to choose *Honour* within the context of one strategy, but this still leaves many other strategies where *Abuse* can be the trustee's preferred action. This leaves the trustor in a state of strategic uncertainty with regard to the strategy the trustee will play.

1.3.2 Evidence on temporal embeddedness

Repeated trust games have also been experimentally tested, although the number of experiments is limited compared to the number of isolated encounter binary trust game and investment game experiments.

Several contributions can be found on finitely repeated trust games. Camerer and Weigelt (1988), followed by Neral and Ochs (1992), Brandts and Figueras (2003) and Anderhub et al. (2002) form a series of studies that aim at testing the behaviour of trustors and trustees in finitely repeated games of incomplete information. A detailed overview of these studies can be found in Camerer (2003). These experiments confirm that temporal embeddedness in the form of control and learning leads to higher levels of co-operation. Trustees are more likely to choose *Co-operation* as long as they observe *Honour* compared to the isolated encounter setting. Near the end of the game, the frequency with which trustors and trustees choose *Co-operation* and *Honour* decreases sharply, which is also known as the 'endgame effect'. The overall conclusion of these articles is that on the individual level the Sequential Equilibrium (SE) fails to predict the behaviour of participants, but at the aggregate level it explains the data reasonably well.

It should be noted that in these experimental studies the subjects play many repeated games, which allows for what Chong et al. (2006) call 'cross-sequence learning'. When subjects become more experienced with playing a finitely repeated trust game, they behave closer to SE compared to when they were inexperienced during earlier sequences of this game.

Bornhorst et al. (2004) tested a finitely repeated version of the investment game. They let subjects play six period games. Their results support the findings of finitely repeated binary trust game experiments. They found that trustors and trustees show a higher degree of trusting and trustworthy behaviour compared to the one-shot experiments. Towards the end of the game, the degree of trust and trustworthiness declines.

Bohnet and Huck (2004) look at three different matching protocols and how they influence future decision making. The first ten rounds of their experiment participants either were matched with the same partner for all ten rounds or they were randomly matched with a new partner every round. In the case of the stranger matching, two situations were possible.Trustors either received information about the past behaviour of the new trustee they were matched with, or did not. In the next ten rounds, all subjects were matched under the stranger protocol without feedback. They found that in the first phase trustors and trustees playing with the same partner were more likely to co-operate compared to those under stranger matching with feedback. Stranger matching with feedback in turn allowed for more co-operation compared to stranger matching without feedback.

The data from Bohnet and Huck (2004) again shows that finitely repeated trust games yield higher levels of co-operative behaviour compared to isolated encounter trust games. The graphs reported in their article also show a clear end-round effect as well. Next to this they found that in the second phase of their experiment participants who had been playing under partner matching in the first phase were more likely to co-operate compared to the participants who had been playing under the other two matching protocols. This result suggests that some form of cross-sequence learning can exist, even between finitely repeated and isolated encounter games.

Engle-Warnick and Slonim (2006) test an indefinitely repeated version of a binary trust game with a continuation probability of 0.8. The subjects play 20 repeated games, also called ‘supergames’, in total. Given the continuation probability in this experiment, the expected duration of a supergame is five rounds. Subjects were rematched for every new supergame. They tested whether the length of a supergame has an influence on future decision-making. The experimental design consists of two treatments. In the long-short treatment, the first ten supergames have an expected average length of 6.5 rounds, while the second ten supergames have an expected average length of 3.6 rounds. The short-long treatment is a mirror version of the long-short treatment.

Engle-Warnick and Slonim (2006) found that when trustors and trustees played longer supergames they were more likely to choose *Co-operation* and *Honour* in the next supergame compared to if they played shorter supergames. They also found a decrease per round in the overall rate of co-operation within each supergame.

Engle-Warnick and Slonim (2004) and Slonim et al. (2006) compare indefinitely with finitely repeated binary trust games. The finitely repeated games had a fixed length of five periods. The indefinitely repeated games used a continuation probability of 0.8 and had, therefore, an expected duration of five periods.

Engle-Warnick and Slonim (2004) found no significant difference in the overall level of co-operation between the indefinitely and finitely repeated trust game when subjects are inexperienced in playing repeated games. However, when trustors and trustees become more experienced, a higher level of co-operation is visible in indefinitely repeated games. Engle-Warnick and Slonim (2004) also find evidence that trustors use distinct strategies to play the different repeated games. In the indefinitely repeated trust game, trustors played the grim-trigger strategy, while in the finitely repeated trust game, different strategies were played, e.g. they always chose *Co-operation*, *No co-operation* or they first chose *Co-operation* for a number of rounds followed by *No co-operation* forever. In the graphs presented in Engle-Warnick and Slonim (2004), a clear decrease over time is visible in the overall level of co-operation within repeated games.

Slonim et al. (2006) look at the effect history has on current decision-making. They make a distinction between a *good history* when both players have always co-operated with each other, a *bad history* when at least one player did not co-operate at least one time and *no history* when players have no experience with each other. This study shows that in indefinitely repeated games more co-operation is observed compared to finitely repeated games, regardless of whether there is a good or bad history. However, in the case of no history, there is no significant difference between finitely and indefinitely repeated games.

If control can sustain co-operation in indefinitely repeated games, it should, according to theory, do so in all periods of the game. Nevertheless, the experimental results of the studies discussed above show a decrease in co-operation over time within supergames; control does not seem to be as strong as theory suggests. However, if cross-sequence learning is taken into account, the data show that control does allow for higher levels of co-operation and is also able to sustain co-operative behaviour in later periods in indefinitely repeated games.

1.4 Information asymmetry and the *asymmetric* trust game

When, in a bi-lateral relation, one party has private information, this is known as 'information asymmetry'. The best known example comes from Akerlof (1970), where a car salesman knows whether a car is a lemon or not and the buyer does not have this knowledge. In our example of John and William, William also has private information. He knows whether he will honour John's trust or not.

In the trust problem, information asymmetry is always present. *Ex ante* the trustor never knows the trustee's preferences and thus whether he will honour his trust, whereas the trustee does. The trustor can form a belief about the proportion of honest trustees in the population. Depending on this belief, the trustor might place trust or not. This difference in information is also referred to as 'incomplete information'.

In the example of John and William, it was argued that John faced another information problem. While in the binary trust game the trustor knows *ex post* if his trust has been abused or not, John, on the other hand, could not directly observe William's action. In other words, *ex post* the trustor does not know for certain whether his trust has been honoured or abused. In this thesis, asymmetric information refers to information asymmetry about the trustee's choice.

In general, in trust relations with fluctuations in external conditions such as market prices or the weather, information asymmetry about the trustee's choice can arise. When the task that the trustee is required to fulfil depends heavily on external conditions which the trustor cannot observe, it can be difficult for the trustor to establish whether the result he observes displays the trustee's trustworthiness or not. Furthermore, trust problems where the trustee can be labelled as an expert can suffer from information asymmetry. For example, for most people it is impossible to determine by themselves whether the trust they place in their doctor is justified.

In order to allow for information asymmetry in the binary trust game we will introduce an adjusted version of this game, which we call the *asymmetric* trust game.[26] [27] In the *asymmetric* trust game, the trustor's pay-offs when choosing *Co-operation* are expected values based on two possible earnings with two different probability distributions. Which distribution is used

[26] I am grateful to Jeroen van de Ven for some helpful suggestions when constructing this game.

[27] To remind the reader that the specific situation as described by the *asymmetric* trust game is meant, we write the word 'asymmetric' in *italic*.

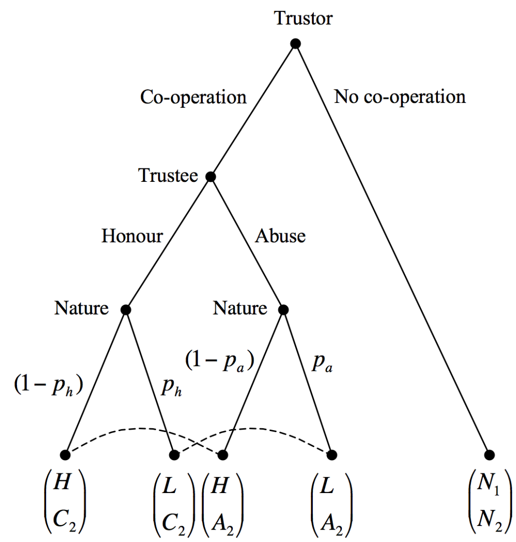


Figure 1.3: The *asymmetric* trust game. The dotted lines reflect that the trustor does not know in which node he is.

depends on the trustee's decision. Nature determines which earning is realised.

Of the two possible earnings, one is a high pay-off (H_y) and the other a low pay-off (L_y). $y \in [a, h]$ is the trustee's action set. Let p_a be the probability linked with L_a, the low pay-off in case of *Abuse*. Thus, $(1 - p_a)$ is the probability of H_a, the high pay-off in case of *Abuse*. Let p_h be the probability of L_h, the low pay-off in case of *Honour*, so the probability of H_h, the high pay-off in case of *Honour*, is $(1 - p_h)$. When the trustor should not be able to distinguish between his trust being abused or honoured, then it needs to be the case that $L_h = L_a = L$ and $H_h = H_a = H$. In other words, the trustor only observes H and L.

The trustee has the same information that he has in the binary trust game, with the exception that he does not know the realisation of the trustor's pay-off. Therefore, the trustee can focus on his decision between honouring and abusing trust. He does not need to take the realisation of nature's draw into account. For a graphical representation of the *asymmetric* trust game, see Figure 1.3.

It should be noted that in order to speak of a trust game, p_a and p_h cannot take on every possible value in the interval $]0, 1[$. To maintain

strategic equivalence to the original trust game we need in expected terms $E(C_1) > E(S_1)$, where $E(C_1)$ stands for $p_h L + (1 - p_h)H$ and $E(S_1)$ represents $p_a L + (1 - p_a)H$.
This implies:

$$p_h L + (1 - p_h)H > p_a L + (1 - p_a)H, \tag{1.3}$$

which simplifies to:

$$p_h < p_a. \tag{1.4}$$

In other words, p_h must be chosen from the interval $]0, p_a[$ and p_a must be chosen from the interval $]p_h, 1[$.

Additionally, it should be that $E(S_1) < N_1$, which implies that:

$$p_a > \frac{H - N_1}{H - L}. \tag{1.5}$$

Because p_a is a probability between 0 and 1, this means that:

$$L < N_1. \tag{1.6}$$

The main difference between the binary trust game and the *asymmetric* trust game is that the trustor's pay-offs do not reveal the trustee's decision in the latter, while they do in the former. Nonetheless, when players play the *asymmetric* trust game as an isolated encounter, their equilibrium behaviour should not differ from the binary trust game. In other words, the pure strategy Nash equilibrium is the same in both games, namely (*No co-operation*, *Abuse*).

In this thesis, trust is studied under asymmetric information as defined above and under incomplete information. This is reflected in Figure 1.4, where the first move of nature reflects the incomplete information about the trustee's type, while the last move of nature represents the information asymmetry about the trustee's decision.

In the *asymmetric* trust game of incomplete information, the trustor will choose *Co-operation* when his belief P is:

$$P > \frac{N_1 - E(S_1)}{E(C_1) - E(S_1)}. \tag{1.7}$$

A comparison of (1.2) and (1.7) makes clear that, also in case of incomplete information, there is no real difference between the equilibrium conditions for the isolated encounter binary trust game of Figure 1.2 and the *asymmetric* trust game of Figure 1.4.

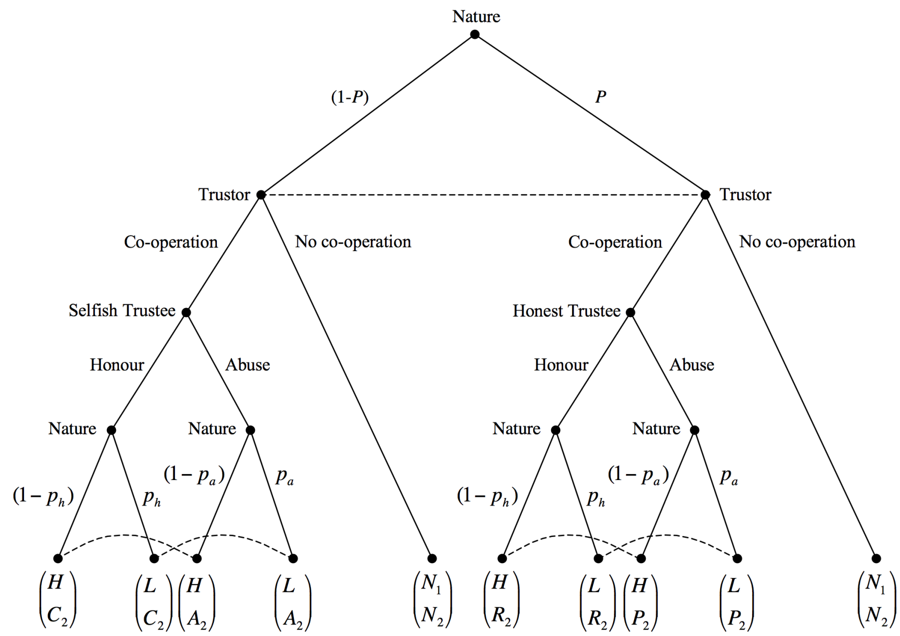


Figure 1.4: The *asymmetric* trust game of incomplete information. Observe that nature moves twice in this game.

The economic literature on the type of information asymmetry discussed in this section is very limited. At first sight this might not be very surprising, because we showed that the presence of asymmetric information in isolated encounter trust games has no effect on the behaviour of players in equilibrium. However, information asymmetry does have an effect on strategic behaviour when social dilemma games are repeated.

The role of asymmetric information in infinitely repeated games has been theoretically analysed by Green and Porter (1984). Green and Porter take the Cournot model of duopoly as a starting point to analyse collusion under asymmetric information. They show that collusion is less easily achieved under asymmetric information. Tirole (1988) developed an alternative model based on Green and Porter (1984), but he takes the Bertrand model of duopoly as his starting point. This model can also be found in Ivaldi et al. (2003).

The study of Aoyagi and Fréchette (2008) is an experimental paper and is closely related to the research presented in this thesis. Aoyagi and Fréchette (2008) test an indefinitely repeated prisoner's dilemma game with a continuation probability of 0.9. In their experiment, the pay-offs can vary depending on the degree of information asymmetry. The control treatment is a standard indefinitely repeated prisoner's dilemma game, whereas in the other treatments the players earn the same pay-off as they would in the control treatment, only this time a random additional pay-off is added. The random pay-off is drawn from the same distribution per treatment, but different distributions are used between treatments. Due to the additional random pay-off, which could be either positive or negative, the reported pay-off could have come from all possible strategy combinations. The likelihood differed per treatment, so depending on the treatment the pay-offs were more or less revealing about the opponent's choice. Aoyagi and Fréchette (2008) found that for all degrees of information asymmetry the level of co-operation is positive, but the level of co-operation decreases with the degree of information asymmetry.

To our knowledge, there are no studies on information asymmetry in finitely repeated social dilemma games. There are, however, some experimental studies on isolated encounter games where a roughly comparable form of information asymmetry is studied. In these studies, though, information asymmetry does not arise on the level of players' actions.

Coricelli et al. (2006) study the investment game of Berg et al. (1995) after changing it by introducing a multiplier that can take on two values. Although both trustor and trustee know the likelihood of facing the two different values of the multiplier, only the trustee knows the realisation and, hence, the size of the surplus. Coricelli et al. (2006) found that the introduction of asymmetric information did not reduce the amount sent and returned when compared to the standard investment game.

Mitzkewitz and Nagel (1993) and Güth et al. (1996) study the effect of asymmetric information for ultimatum games by introducing different endowment sizes. The pie can be larger or smaller, but the true size is known only to the proposer. In these two studies, the players know the action of their opponent, unlike the *asymmetric* trust game. However, because the receivers do not know the actual size of the pie, they cannot judge whether the offer is 'fair'. Although from a theoretical perspective this should not matter, evidence from earlier experimental studies on ultimatum games shows that fairness does play a role (see Roth and Kagel (1995) for a survey). Mitzkewitz and Nagel (1993) and Güth et al. (1996) found that proposers do hide behind the asymmetric information in order to increase their share of the pie.

1.5 Opportunism, monitoring and checking

1.5.1 Monitoring

The trustor in the repeated *asymmetric* trust game of incomplete information faces two serious problems: Honest trustees choose *Honour*, while selfish trustees behave opportunistically and choose *Abuse*. Unfortunately, the trustor does not know what type of trustee he faces. Secondly, when the trustor chooses *Co-operation*, he cannot directly observe what the trustee, regardless of his type, has chosen in response.

The first problem is well-documented. In the economic literature one can find many articles about contractual relations and the effect that incentive devices such as reward and punishment can have on opportunistic behaviour (e.g. Alchian and Demsetz (1972) and Prendergast (1999)). The traditional economic view is perhaps best described by Williamson (1993): "... because opportunistic agents will not self-enforce open-ended promises to behave responsible, efficient exchange will be realized only if dependencies are supported by credible commitments". According to this point of view, it is in the principal's (trustor's) interest to use incentive devices to make sure that opportunistic agents (trustees) will behave 'responsibly'.

Opportunism on the side of selfish trustees can be countered by embedding the trust problem, as was suggested in Section 1.1. In Section 1.3, it was made clear that the two mechanisms of temporal embeddedness – learning and control – can make the selfish trustee prefer to choose *Honour*. Network embeddedness, although beyond the scope of thesis, can also solve the trust problem through the same two mechanisms.

Institutional embeddedness is an alternative way to align the trustor and trustee's interests. In the case of institutional embeddedness, the trustor takes active measures to influence the trustee's preferences in such a way that in the isolated encounter trust game the trustee's interests are aligned with those of the trustor. In the literature, these measures are known under the term 'monitoring', which can be both positive or negative in nature, in the sense that it can stimulate the trustee to choose *Honour* or deter the trustee from choosing *Abuse*.

In the literature, one can find all kinds of monitoring measures, but they all have the same goal: ensuring that the trustee will not abuse the trustor's trust.[28] Monitoring measures alter the trustee's preferences in

[28]Some authors argue that monitoring can have the opposite effect and, hence, crowd out co-operative behaviour instead of increasing it (see, for example, Frey (1993) and Dickinson and Villeval (2005)).

such a way that the trustor knows that the trustee will choose *Honour* *before* the trustor makes the decision to choose *Co-operation*. This makes intuitive sense, because most trustors would like to know whether a trustee will honour their trust before they enter into a relationship with the trustee.

One form of monitoring is contracting. The trustor can, for example, set an incentive wage that would make the trustee prefer to choose *Honour* in the trust game. Another example of contracting can be the threat of litigation in the case of breach of contract. An illustration of the latter is the experimental study of Bohnet et al. (2001), where the earnings of the trustee depend on the outcome of a litigation process.

The most strict form of monitoring is the situation where the trustee's choice set itself is limited by the trustor; for instance, Falk and Kosfeld (2006) assume that the trustor can determine the trustee's choice set directly. Either the trustee has complete freedom in choosing his effort level or a minimum level is required. In Ploner (2006), the trustor can detect the trustee's intention before he needs to make a decision. When this detection technology is used, the trustee must respond according to his revealed intentions.

Perfect monitoring, as tested in the experimental studies mentioned above, is difficult to match with the definition of trust problem presented in this thesis. When perfect monitoring, as a consequence, determines that the trustee prefers to respond with *Honour*, it violates element (3) of the trust problem. However, monitoring activities are hardly ever perfect and subsequently leave room for trust.

1.5.2 Checking and the *symmetric* trust game

This leaves the second problem, namely that the trustor cannot directly observe the trustee's response. Hence, in the repeated *asymmetric* trust game, an additional form of opportunistic behaviour can arise, because the trustee can hide his true behaviour behind this information asymmetry. The work of Green and Porter (1984) and Aoyagi and Fréchette (2008) shows that information asymmetry makes it more difficult to reach co-operative outcomes in repeated social dilemma games. It will be made clear by the models presented in Chapters 2 and 4 that information asymmetry about the trustee's decision has a negative influence on the mechanisms of temporal embeddedness. Information asymmetry makes control and learning less effective and can, in turn, prevent trustors and trustees from reaching the socially optimal outcome.

In the example of John and William, it was suggested that this second problem can be countered by using a checking option.[29] Although monitoring and temporal embeddedness can align the trustor and trustee's interests, they are ineffective against the problem of information asymmetry. The only way to deal with information asymmetry is to close the information gap. Checking does exactly this.

Checking can be implemented in the *asymmetric* trust game by adding a decision node for the trustor; the trustor can use the checking option *after* both he and the trustee made their decisions, but before they receive their pay-offs. When the trustor uses the checking option, he retrieves the trustee's response in the *asymmetric* trust games.

In the absence of the checking option, the trustor can only form an expectation about the trustee's choice based on the pay-off he earned. When the trustor uses his checking option, the trustor will, next to his pay-off, also learn which of the trustee's decisions triggered this pay-off. In other words, checking resolves the information asymmetry.

In isolated encounter *asymmetric* trust games, the checking option should have no effect. What is the point of finding out afterwards what the trustee did if the trustor has no future with this trustee? The consequences of checking, however, become apparent in repeated *asymmetric* trust games. The trustor can use the information he receives from the checking process to determine whether he will choose *co-operation* the next time he meets the trustee.

Checking changes the character of repeated *asymmetric* trust games. When the trustor uses his checking option in every period, he can change the game into one of symmetric information in the sense that at the end of the period he knows, just like the trustee, what the trustee's response was. Checking therefore enhances the mechanisms of temporal embeddedness. Control becomes more effective because the trustor can punish more precisely, given the information he receives. The increase in information enables the trustor to update his belief more adequately, improving the effectiveness of learning.

Figure 1.5 presents the *asymmetric* trust game of incomplete information with the checking option. When the trustor makes use of the checking option in this game, his pay-off is still determined by nature, but he can in addition observe the trustee's response that determined with which likelihood nature would draw his pay-off. When he does not make use of the checking option,

[29] The term 'checking' is used to make a distinction with respect to 'monitoring', which is often found in the literature, and to separate it from 'control' (one of the mechanisms of temporal embeddedness.)

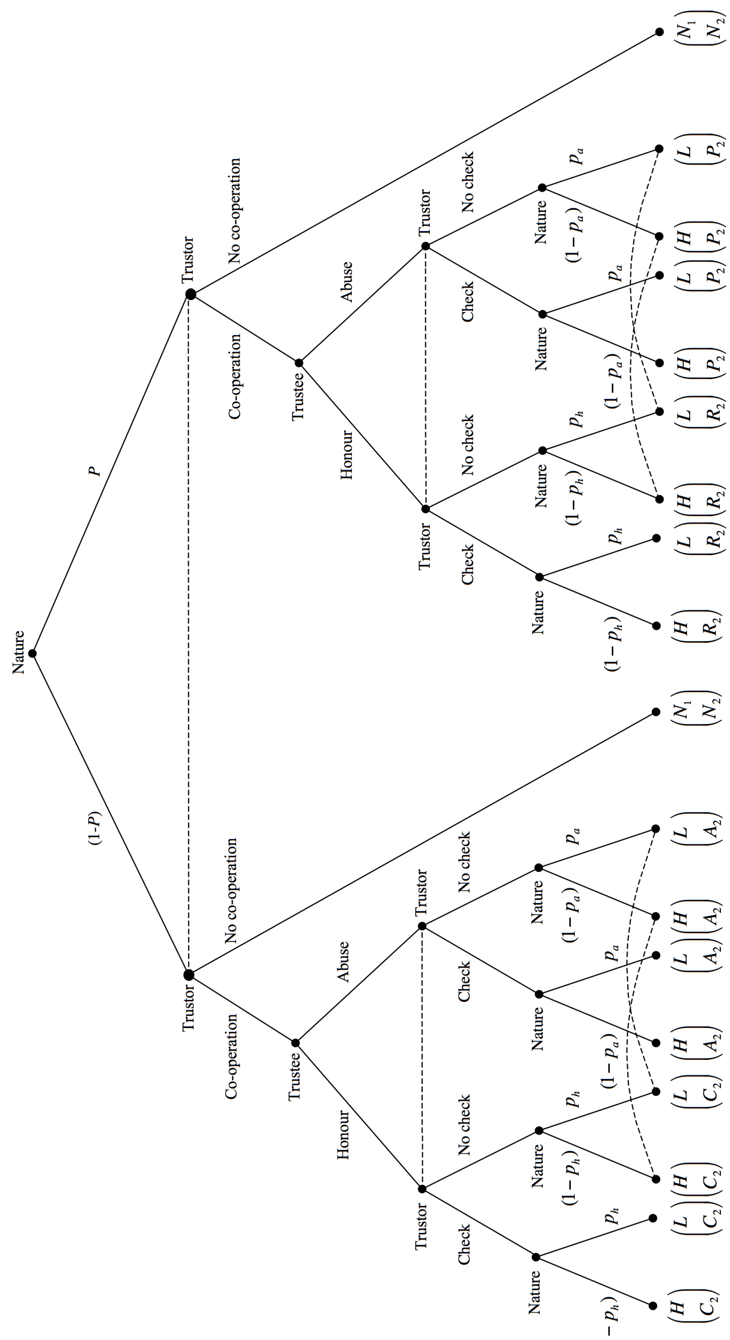


Figure 1.5: The *asymmetric* trust game of incomplete information with the checking option.

he will have the same information as in the *asymmetric* trust game of incomplete information, without the checking option as presented in Figure 1.4.

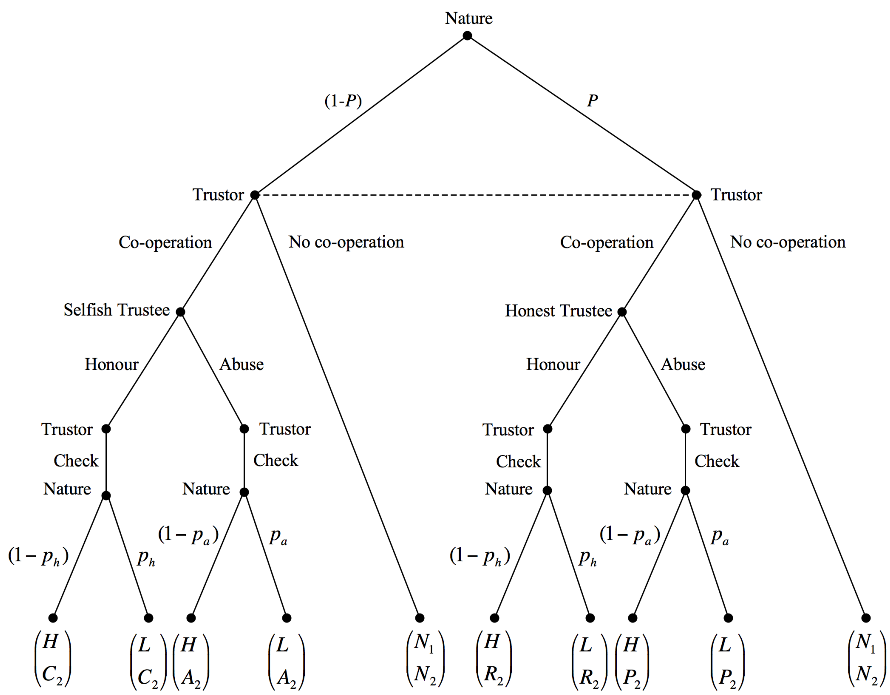


Figure 1.6: The *symmetric* trust game of incomplete information.

In this thesis, as will be explained in Section 1.6, checking will be studied under the assumption that it is costless. Given this assumption, checking is a dominant action for the trustor every time he has the opportunity to use it. When the trustor uses the checking option in every period, he no longer faces information asymmetry about the trustee's decision, although the trustor's pay-offs are, of course, still determined by nature. The game in Figure 1.6 reflects this situation, and is referred to in this thesis as the *symmetric* trust game of incomplete information.[30]

Although checking itself does not restrict the trustee's behaviour in a direct sense, the trustor will use the additional information he acquires when

[30] To remind the reader that the specific situation as described by the *symmetric* trust game is meant, we will also write the word 'symmetric' in *italic*.

making future decisions. The selfish trustee should take this into account when he knows that the trustor has a checking option. Honest trustees might even prefer a situation where the trustor has a checking option, because the trustor no longer needs to be unsure about their trustworthiness.

Checking, in contrast to perfect monitoring, does not aim at rewarding or punishing the trustee in a direct material sense, nor does it limit the trustee's action set. Checking can stimulate co-operative behaviour by strengthening the mechanisms of temporal embeddedness.

1.6 Aim and outline of the thesis

1.6.1 Research questions

In Section 1.1, we argued that trust is a desirable 'commodity' under the assumption that voluntary co-operation leads to a Pareto improvement. Experimental research tells us that trust can enable relations in isolated encounter settings. Most trustors have long-run trust relations with their trustee, i.e. they have several encounters with the same trustee over a longer period of time. The literature on finitely and indefinitely repeated trust games suggests that due to the mechanisms of temporal embeddedness, learning and control, higher levels of co-operation can be obtained in long-run trust relations. The few existing studies on asymmetric information suggest that information asymmetry has a negative influence on the level of co-operation in repeated games. Hence, being faced with asymmetric information is a situation a rational trustor will not prefer.

Recent developments regarding the global financial crisis make clear that trust is much less easily sustained when trustors cannot correctly validate the actions of trustees. Deposit holders are less likely to trust banks with their money when they do not know how banks are investing their savings. The crisis also emphasises the importance of independent regulatory agencies and central banks in their roles of supervisor. These government-established organisations have the task to reduce information asymmetry in financial markets. The checking option, which we introduced in the previous section, can be seen as an abstract representation of regulatory agencies or supervisors that are capable of closing the information gap for the trustor.

Besides the example of the financial crisis, there are many other economic situations were trustors are confronted with information asymmetry about the trustees' actions. This is basically true for each situation where the trustor's outcome is not only determined by the trustee's choice, but also by some external process.

Take, for example, the relation between voters and their democratically-appointed representative. The voters want the representative to defend their point of view in political debates and negotiations. The outcomes of the negotiations depend not only on the determination of the representative and the quality of his arguments, but also on the possibility of forming a coalition with other representatives. Voters have usually limited knowledge about the ins and outs of a negotiation process; for them it is unclear whether the compromise reached by the representative is a relatively good deal or a relatively bad deal, which makes it difficult for voters to determine if they should vote for the same representative at the next election.

There are many real-world economic situations that can be described by repeated *asymmetric* trust games. It is, therefore, important to gain a better insight into the sustainability of trust when people are faced with asymmetric information and the effectiveness of the checking option to stimulate co-operation in this type of situation.

Next to this main research objective, this thesis will also investigate whether the theoretically fundamental difference in horizon perspective leads to significant differences in trustor and trustee behaviour, when comparing indefinitely and finitely repeated *asymmetric* trust games. Engle-Warnick and Slonim (2004) and Slonim et al. (2006) found that the difference in horizon perspective has an impact on the strategies played and on the overall level of co-operation in repeated trust games. When comparing finitely and indefinitely repeated *asymmetric* trust games, it is questionable whether a similar difference in the level of co-operation can be found. The models from Chapters 2 and 4 suggest that this difference will be small. However, when it comes to strategies the two mechanisms of temporal embeddedness – control and learning – suggest that trustors should react differently under the two different horizon perspectives.

Finally, we will explore whether the two types of trustors and trustees, as they were defined in Section 1.2.4, also behave differently in indefinitely and finitely repeated *asymmetric* and *symmetric* trust games as theory suggests. In Section 1.2.3, we concluded that the binary trust game of incomplete information was a good representation of the trust problem, as defined in Section 1.2.1. In the experimental literature on repeated trust games, incomplete information is primarily studied in the context of finitely repeated games. Studies aimed at testing the sequential equilibrium explicitly incorporate different types of trustees in their experiments. They do so by either forcing a proportion of the subjects to choose *Honour* or by assigning the role of trustee to the experimenter, who will make the same decision. In this thesis, subjects are not forced to play a strategy. Trustor and trustee types are determined by an additional isolated encounter treatment, which is fully detailed in Chapter 3. It is important to know whether the difference in trustor and trustee types, a distinction based on behaviour in the isolated encounter trust game, remains significant once applied in repeated encounter settings.

This thesis aims at answering the following three research questions:

1. What is the theoretical and empirical effect of the checking option in repeated *asymmetric* trust games of incomplete information?
2. Can we empirically identify the effect of the horizon perspective in repeated *asymmetric* trust games of incomplete information?
3. Can we confirm theoretical predictions on the behaviour of trustor and trustee types in indefinitely and finitely repeated *asymmetric* and *symmetric* trust games of incomplete information?

1.6.2 Approach and outline

The literature on trust games with asymmetric information at the level of players' actions is very limited. This thesis tries to fill this gap. To do so we will extend the theoretical work on asymmetric information in repeated games.

In Chapters 2, the model of Green and Porter (1984) will be applied on the indefinitely repeated *asymmetric* trust game and we will extend this by allowing for social preferences. A perfect Bayesian equilibrium will be described for the situation where the grim-trigger strategy does not align the behaviour of all players.

To investigate how learning and control in finitely repeated trust games is affected by information asymmetry, a version of the Kreps and Wilson model will be developed that incorporates information asymmetry in Chapter 4.

Game theory assumes that players have complete and transitive preferences. In the language of Section 1.2.2, the theoretical models describe calculative trust, i.e. non-calculative motivations of the trustor and radical uncertainty will be ignored in the models presented in this thesis.

The theoretical models provide insights into how learning and control are affected by information asymmetry. To test the models and the effectiveness of the checking option, we decided to run laboratory experiments. A laboratory experiment has the clear advantage of a controlled economic environment, which allows us to generate an environment that closely resembles the theoretical framework and its underlying assumptions. The checking option is a good example of an instrument that is easily created in the laboratory, but much harder to test in the field. Secondly, it is easier and considerably faster to simulate long-run trust relations in a laboratory experiment compared to a field experiment.

We are aware of the fact that laboratory experiments do not come without drawbacks.[31] Laboratory settings differ from real-world conditions, so it is important to question the external validity of experimental results (Loewenstein (1999) and Levitt and List (2007)). It is also important to understand the limitations of the experimental testing of theoretical models (Binmore (1999) and Rubinstein (2001)). The models from Chapters 2 and 4 are fairly complex. It is reasonable to question whether our subjects (even when they are hyper-rational) will calculate the equilibrium strategy. According to Rubinstein (2001): "...economic theory is an abstract investigation of the concepts and considerations involved in real life economic decision making rather than a tool for predicting or describing real behaviour". Following this line of argument, we aim to verify whether the concepts and considerations contained in economic models are sound and useful, instead of testing whether theory predicts behaviour.

The models from Chapters 2 and 4 are used to determine the parameter values of the games used in the experiments and to come to testable hypotheses in order to answer the main research questions. In Chapters 3 and 5, the indefinitely repeated *asymmetric* trust game and the finitely repeated *asymmetric* trust game are tested.

Each experiment consists of three treatments. An experiment starts with a Type Treatment (TT) that provides the option to label both trustors and trustees according to their type, as they have been defined in Section 1.2.4. The remaining two treatments in the experiment are the repeated *asymmetric* trust game with the checking option, the so-called Checking Treatment (CT), and the same game only now without the checking option, the No-Checking Treatment (NCT). The experiments presented in Chapters 3 and 5 use the 'within subjects' design.

In most real-life situations, checking will be costly for the trustor. In the experiments presented in this thesis, checking is, however, made costless. Given the fact that this is the first time that the effect of checking is researched, we wanted to make sure that we would have enough observa-

[31] For a broader discussion on experiments in economics, see also Roth (1994), Samuelson (2005).

tions where the checking option was used. Making checking costly would probably have led to a reduction in the number of checking observations. Although it is interesting to investigate the effect of the price of checking on the trustor's checking behaviour, it is beyond the scope of this study.

In Chapter 6, the models from Chapters 2 and 4 are compared. The experimental data from Chapters 3 and 5 cannot directly be compared. On the basis of the theoretical models pay-off values where calculated, that would yield clear differences between the treatments; it was impossible to find one set of pay-off values that would provide clear differences in both experiments, so different values were used for the pay-off parameters. To make a comparison, a new experiment was run, testing the indefinitely repeated *asymmetric* trust game using the pay-off values from Chapter 5. This is also reflected by the dotted line in Figure 1.7 that provides a schematic presentation of the structure of this thesis. The third experiment uses the 'between subject' design and consists of two treatments: the Indefinitely repeated *asymmetric* trust game Treatment (IT) and the finitely repeated *asymmetric* trust game Treatment (FT). The latter is the NCT from the finitely repeated *asymmetric* trust game experiment presented in Chapter 5.

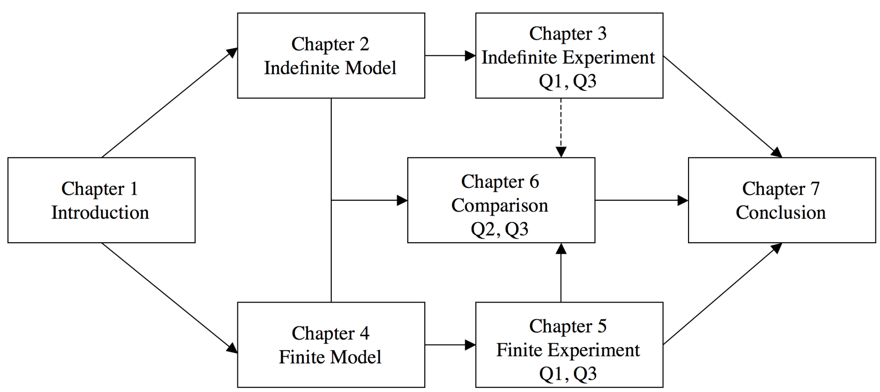


Figure 1.7: Structure of the thesis. Q1, Q2 and Q3 refer to the main research questions and indicate the experimental data used to answer the research questions.

The third research question is answered using all the available experimental data. The theoretical models from Chapters 2 and 4 suggest the following: When the grim-trigger strategy is the equilibrium strategy for selfish trustees in indefinitely repeated *symmetric* trust games, the trustee's type becomes irrelevant, because all types of trustees should choose *Honour*. The control mechanism aligns the preference of both types of trustees. As a consequence, both types of trustors should choose *Co-operation*. When the grim-trigger strategy is not an equilibrium strategy for selfish trustees, the trustor might still play this strategy if he believes that the proportion of honest trustees in the population is significantly large (see Section 2.2.4). In other words, trustor and trustee types should matter.

In situations where both control and learning allow for co-operation, as seen, for instance, in finitely repeated *asymmetric* and *symmetric* trust games, a difference in the behaviour of trustee type is important because the trustor can only update his belief when the two types of trustees play different strategies. Trustor type is also important because trustful trustors have a higher prior belief about the proportion of honest trustees in the trustee population compared to non-trustful trustors. This should have a positive influence on the frequency with which they choose *Co-operation*.

Table 1.1: Mechanisms of temporal embeddedness and type indicators.

Experiment	Treatment	Control trigger by	Type
Indefinite	CT	Grim-trigger strategy	Type irrelevant
	NCT	Grim-trigger strategy	Type relevant
	IT	Grim-trigger strategy	Type irrelevant
Finite	CT	Learning	Type relevant
	NCT=FT	Learning	Type relevant

The pay-off values for the indefinitely repeated *asymmetric* trust game are chosen in such a way that the grim-trigger strategy is not an equilibrium strategy for selfish trustees in the case of the NCT, but it is in the case of the CT and the IT. This means that type indicators should be relevant in the case of the NCT, while they should not matter at all for the CT and IT. In the finitely repeated *asymmetric* trust game, learning determines when control is triggered. Trustee and trustor types should be relevant in both the NCT=FT and the CT. See Table 1.1 for the relation between the mechanisms of temporal embeddeness, the relevance of type and the experimental design.

In the final chapter of this thesis, the overall conclusions with respect to the three research questions are presented. In this chapter, we will also point out some interesting lines of investigation for future research.

1.7 An overview of the main results

Chapter 2 describes the grim-trigger strategies for both the indefinitely repeated *symmetric* and *asymmetric* trust games. A comparison of the models makes clear that in the case of asymmetric information about the trustee's decision, co-operation is more difficult to establish. The main reason lies in the fact that the trustor needs to play a stricter grim-trigger strategy in the presence of asymmetric information. This strict grim-trigger strategy does not punish the trustee when he chooses *Abuse*; rather, when the trustor observes a low pay-off. The strict grim-trigger strategy leads to less easily satisfied equilibrium conditions. Given the fact that the trustor's pay-offs do not perfectly reflect the trustee's decision, good behaviour might be punished, while bad behaviour can be left unpunished. Control becomes less effective and, thus, co-operation is less easily established.

Chapter 3 presents the results of the indefinitely repeated *asymmetric* trust game experiment. The trustors use the checking option 86% of the time in the CT. The results show that in the CT trustors are more likely to choose *Co-operation* and trustees are more likely to respond with *Honour*. Almost all the results on the type indicators support the theoretical predictions. Trustful trustors are more likely to choose *Co-operation* in the NCT, while trustor and trustee types do not matter in the CT, as predicted by theory. However, we found that honest trustees are not more likely to choose *Honour* compared to selfish trustees in the NCT, while according to theory they should.

In the CT, the trustor's pay-off realisation appears to matter. First, when their trust is honoured, trustors are more likely to choose *Co-operation* in the next period, but even more so when they have earned a high pay-off. Secondly, there is a small chance that trustful trustors still choose *Co-operation* after they have observed that their trust has been abused, as long as they have earned a high pay-off.

Chapter 4 presents a model of the finitely repeated *asymmetric* trust game. The sequential equilibrium is based on the principle that the trustor can learn about the type of trustee he is facing, on the basis of the information he receives during the course of the game. In comparison to the finitely repeated *symmetric* trust game, the quality of the information is lower in the presence of information asymmetry. In equilibrium, selfish trustees should start to randomise over their actions earlier, because the trustor's threshold increases sooner. The reason behind this lies in the fact that in the case of information asymmetry the trustor needs to have a higher assurance that he faces an honest trustee. When selfish trustees start randomising between *Honour* and *Abuse*, they should assign a higher probability to *Abuse* compared to the *symmetric* trust game in order to make the trustor indifferent. Given the stricter equilibrium conditions of the finitely repeated *asymmetric* trust game, it becomes more difficult to sustain co-operation compared to the finitely repeated *symmetric* trust game.

Chapter 5 presents the results of the finitely repeated *asymmetric* trust game experiment. Trustors use the checking option 77.54% of the time in the CT. Both trustors and trustees are more likely to co-operate in the CT compared to the NCT. In line with theory, trustful trustors are more likely to choose *Co-operation* compared to non-trustful trustors, and honest trustees are more likely to choose *Honour* compared to selfish trustees. The theoretical assumption that honest trustees remain honest throughout the entire game is supported by the experimental data. The data also provide evidence for learning. According to theory, trustors should use Bayes's Rule to update their belief. Trustful trustors have a relatively high prior belief about the proportion of honest trustees compared to non-trustful trustors, which means that for trustful trustors a larger drop in the likelihood with which they choose *Co-operation* should be visible upon the appearance of evidence that they are facing a selfish trustee compared to a non-trustful trustor. We found that, after they have observed *Abuse* in the CT, trustful trustors lower the frequency with which they choose *Co-operation* even more compared to non-trustful trustors. This result is also supported by the finding that both types of trustors are less likely to choose *Co-operation* in later periods of the game. However, this effect is again stronger for trustful trustors.

Due to the nature of the *asymmetric* trust game, trustors have different pay-off realisations. The pay-off realisation does have an influence on trustor behaviour, even when they use their checking option and know the trustee's response to their trust. Trustors, regardless of their type, are more likely to choose *Co-operation* when they have earned a high pay-off.

In Chapter 6, a comparison is made between finitely and indefinitely repeated *asymmetric* trust games. In line with earlier findings, the data show no significant difference in the likelihood with which trustors and trustees choose *Co-operation* and *Honour* in finitely and indefinitely repeated trust games, when they are inexperienced in playing these games.

Trustors who observe a low pay-off are less likely to choose *Co-operation* in both treatments. This effect is, however, significantly stronger in the indefinitely repeated *asymmetric* trust games, the result of which appears to support the theoretical difference between control triggered by the grim-trigger strategy and control trigger by learning. Control in indefinitely repeated trust games is based on the principle that punishment will start as soon as *Abuse* is observed, whereas learning and control in finitely repeated trust games suggest that co-operative behaviour will stop as soon as the trustor's belief that he faces an honest trustee is not sufficiently high any longer. This difference also becomes visible in the following result: In the FT, trustful trustors initially choose *Co-operation* with a larger frequency compared to non-trustful trustors. Over time, the likelihood with which trustful trustors choose *Co-operation* decreases stronger compared to that of non-trustful trustors. Conversely, in the IT, there is no difference between trustful and non-trustful trustors when we compare the likelihood with which they choose *Co-operation*.

Chapter 2

A model for indefinitely repeated *asymmetric* trust games*

2.1 Introduction

In this chapter, game theoretical models for the indefinitely repeated *symmetric* and *asymmetric* trust games are presented. The model for the *symmetric* trust game is the standard model for indefinitely repeated games, while the approach of Green and Porter (1984) is used to solve the indefinitely repeated *asymmetric* trust game. In their model, firms observe market prices, which imperfectly reflect the output levels of other firms. Both models presented here differ from existing models in the sense that they incorporate different types of trustees.

Folk Theorems (see, for example, Abreu (1988), Fudenberg and Tirole (1991) and Rasmusen (2001)) show that in infinitely repeated games an infinite number of subgame-perfect Nash equilibrium strategies can exist. The same is true for the indefinitely repeated *asymmetric* and *symmetric* trust games. The isolated encounter equilibrium strategy of (*No co-operation*, *Abuse*) is an example of one of the many possible subgame-perfect Nash equilibrium strategies. Moreover, it is the strategy that satisfies the minmax criterion.[1] This strategy also yields the lowest pay-off combination,

*This chapter and the following chapter are based on the working paper *Information asymmetry and checking in indefinitely repeated trust games*, a joint work with Stephanie Rosenkranz and Vincent Buskens.

[1]A strategy is a minmax strategy for player i when it maximises his pay-off, given that

given the assumption made in Chapter 1 on the pay-off structure of the trust game. At the other extreme, with the potential of yielding the maximum pay-off combination, is the grim-trigger strategy.

It is straightforward enough to see that next to the two strategies mentioned above all kinds of pure and/or mixed strategies can be played where different pay-off combinations can be achieved.[2] This chapter will limit itself to presenting the grim-trigger strategy, the main reason for which is that this sanction strategy makes it most likely that control, as mentioned in Sections 1.3, can support co-operation in equilibrium. When the grim-trigger strategy cannot enforce selfish trustees to choose *Honour* in the indefinitely repeated *asymmetric* or *symmetric* trust games, no other strategy can. Secondly, when control is successful, this simple strategy has the potential to enforce the highest level of co-operation throughout the game. Finally, earlier empirical work by Engle-Warnick and Slonim (2004) shows that in indefinitely repeated trust games trustors play the grim-trigger strategy often.

This chapter is structured as follows. In Section 2.2.1, it is made clear how social preferences are operationalised in *asymmetric* and *symmetric* trust games. In the first part of this chapter, the models assume complete information about the trustee's type. In Section 2.2.2, the model for the grim-trigger strategy in the indefinitely repeated *symmetric* trust game is presented. To study the effect of information asymmetry on the effectiveness of control, this model can be compared with the model for the indefinitely repeated *asymmetric* trust game presented in Section 2.2.3. In Section 2.2.4, the previous sections' models are analysed under incomplete information. Here, the situation is studied where the grim-trigger is not an equilibrium strategy for selfish trustees. This section discusses a model which states that given a sufficient belief about the proportion of honest trustees in the population it can still be interesting for trustors to play the grim-trigger strategy. The chapter ends with a conclusion in Section 2.3.

player j is playing a strategy that will minimise player i's pay-off. When the trustee wants to minimise the trustor's pay-off in the trust game, he will choose *Abuse*. When given this decision, the trustor wants to maximise his pay-off so he will choose *No co-operation*.

[2] Given that the minmax criterion is satisfied, trustors are not willing to accept abuse pay-offs for the entire duration of the game.

2.2 The models

2.2.1 Social preferences

To allow for different types of trustees in the models presented in this chapter, we will use a utility function that combines both egoistic and social preferences. The simplest model of this form is the *social orientation model.* In this model, the utility of player i is represented by a linear combination of the monetary outcomes of player i and the monetary outcomes of player j (McClintock (1972)). The utility function of player i reads as follows:

$$U_i(M_i, M_j) = M_i + \rho_i M_j, \tag{2.1}$$

where $0 \leqslant \rho_i \leqslant 1$ is the parameter for social preferences, which reflects how much player i cares about the well-being of player j. Given the restrictions on ρ_i, player i can never care more about the monetary pay-off of player j than about his own monetary pay-off. The weight attached to player i's monetary pay-off M_i is set to 1, assuming that more money results in a higher utility.

In Table 2.1, the utility for the trustee derived from the different strategy combinations can be found, where ρ reflects the social preferences parameter of the trustee.

Table 2.1: Trustee's utility per strategy combination.

	Trustee's pay-off
No co-operation	$N_2 + \rho N_1$
(*Co-operation*, *Honour*)	$C_2 + \rho E(C_1)$
(*Co-operation*, *Abuse*)	$A_2 + \rho E(S1)$

When $\rho > \frac{A_2 - C_2}{E(C_1) - E(S_1)}$, the trustee prefers *Honour* above *Abuse* in an isolated encounter trust game. According to the definitions of Section 1.2.4, a trustee with these preferences is called an honest trustee. In line with these definitions the trustee is called selfish if $\rho = 0$.

Given the repeated nature of the game, an additional group of trustees can be distinguished. These so-called conditional honest trustees are defined by the following condition: $\frac{A_2 - C_2}{E(C_1) - E(S_1)} > \rho > 0$. For conditional honest trustees the grim-trigger strategy might still be an equilibrium strategy, while for selfish trustees this is not the case, as will be made clear in the following two sections.

The trustor's type is based on his decision whether to co-operate or not in the isolated encounter trust game. As argued in Section 1.2.4, one reason for choosing *Co-operation* could be the trustor's belief about the proportion of honest trustees in the population. In the language of this chapter, this is the trustor's belief about the level of ρ.

In the ensuing two sections, the *asymmetric* and *symmetric* trust games are studied under complete information about the trustee's preferences, i.e. the trustor knows the trustee's ρ. In Section 2.2.4, this assumption is relaxed by assuming that the trustor does not know what type of trustee he faces.

2.2.2 Indefinitely repeated *symmetric* trust game

The indefinitely repeated *symmetric* trust game is, in essence, the one-stage *symmetric* trust game where at the end of the game a continuation probability, δ, determines how likely it is that the game is played again. This process continues until the game stops. For each period t, the outcomes of period $t-1$ are observed before stage t begins.

The grim-trigger strategy is a good example of a sanction strategy that allows the control mechanism to support co-operation. This strategy goes as follows: The trustor will start by choosing *Co-operation*, and will continue doing so as long as the trustee responds with *Honour*. As soon as the trustee responds with *Abuse*, the trustor will change his strategy to *No co-operation* in the next and all coming rounds.

Given that we assume that both the trustor and trustee know the value of ρ, three situations can be distinguished.

For selfish trustees it is assumed that $\rho = 0$. The selfish trustee will prefer to choose *Honour* in every period of the game if:

$$\sum_{t=1}^{\infty} C_2 > A_2 + \sum_{t=2}^{\infty} N_2 \delta^{t-1}, \tag{2.2}$$

or if:

$$\delta > \tilde{\delta} \equiv \frac{A_2 - C_2}{A_2 - N_2}. \tag{2.3}$$

When condition (2.3) is satisfied, it is the trustor's best response to choose *Co-operation*, because $C_1 > N_1$. This trigger strategy constitutes a subgame-perfect Nash Equilibrium, because in an indefinitely repeated game each subgame has the same properties as the entire game. This is the standard equilibrium condition that allows for co-operative behaviour in indefinitely repeated games.

When condition (2.3) is not satisfied, the selfish trustee prefers to choose *Abuse*. The trustor will play *No co-operation* in every period of the game. When the grim-trigger strategy is no equilibrium strategy, limited punishment by the trustor will not encourage the trustee to behave more co-operatively. The only reason for the trustee to choose *Abuse* is when the expected pay-off of this strategy outweighs the expected pay-off of choosing *Honour*. Limited punishment will make choosing *Abuse* only more attractive. Hence, when condition (2.3) is not satisfied, the trustee will choose *Abuse* in period 1 and every other period in which he has the chance to do so. A profit-maximising trustor's best response is to choose *No co-operation* throughout the entire game.

Incorporating social preferences into (2.2) yields:

$$\sum_{t=1}^{\infty}\Big(C_2+\rho E(C_1)\Big)\delta^{t-1} > A_2+\rho E(S_1)+\sum_{t=2}^{\infty}(N_2+\rho N_1)\delta^{t-1}, \qquad (2.4)$$

or:

$$\delta > \delta^*(\rho) \equiv \frac{\Big(A_2+\rho E(S_1)\Big)-\Big(C_2+\rho E(C_1)\Big)}{\Big(A_2+\rho E(S_1)\Big)-(N_2+\rho N_1)}. \qquad (2.5)$$

The partial derivative of (2.5), $\frac{\partial \delta^*}{\partial \rho}$, is a decreasing function. In other words, the right-hand side of (2.5) will decrease when ρ becomes larger, making it more likely that the equilibrium condition is met. When for selfish trustees the grim-trigger strategy is not an equilibrium strategy, it might still be for conditional honest trustees, depending on the value of ρ. Solving (2.5) for ρ will yield:

$$\rho > \rho^* \equiv \frac{A_2-C_2+\delta^*(N_2-A_2)}{E(C_1)-E(S_1)+\delta^*\Big(E(S_1)-N_1\Big)}. \qquad (2.6)$$

Although for the conditional honest trustee choosing *Honour* is not an equilibrium strategy in the isolated encounter *symmetric* trust game, it is in the indefinitely repeated *symmetric* trust game for $\frac{A_2-C_2}{E(C_1)-E(S_1)} > \rho > \rho^*$.[3]

For honest trustees, the equilibrium condition of the grim-trigger strategy is irrelevant. By definition they prefer to choose *Honour* in an isolated encounter trust game and, hence, they should prefer the same in the indefinitely repeated *symmetric* trust game. The trustor will prefer to choose

[3] Substituting (2.6) in the interval and rearranging leads to the following inequality $\frac{A_2-C_2}{E(C_1)-E(S_1)} < \frac{A_2-N_2}{N_1-E(S_1)}$. Given condition (1.1), it can be easily shown that this inequality is satisfied for all possible pay-off values.

Co-operation when he knows that $\rho > \frac{A_2 - C_2}{E(C_1) - E(S_1)}$, i.e. he faces an honest trustee.

2.2.3 Indefinitely repeated *asymmetric* trust game

In the indefinitely repeated *asymmetric* trust game, the grim-trigger strategy described above is no longer applicable, since the trustor is unable to observe the trustor's behaviour. The grim-trigger strategy is based on punishment by the trustor as soon as he observes that the trustee has chosen *Abuse*. In the *asymmetric* trust game, it is not immediately clear when this is the case.

The trustor cannot observe the trustee's response, but he does know that it is more likely to observe an L if the trustee has chosen *Abuse*. Therefore, the strictest grim-trigger strategy he can play is not accepting a single L.

In our analysis of the strict grim-trigger strategy, as defined above, we will consider the same conditions on ρ as in the previous section. When $\rho = 0$, the selfish trustee will prefer to choose *Honour* when his expected utility of choosing *Honour* is larger then his expected utility of choosing *Abuse*. Let V_1 be the expected pay-off when the trustee chooses *Honour*:

$$V_1 = C_2 + (1 - p_h)\delta V_1 + p_h \sum_{t=2}^{\infty} N_2 \delta^{t-1}, \tag{2.7}$$

and V_2 the expected pay-off when the trustee chooses *Abuse*:

$$V_2 = A_2 + (1 - p_a)\delta V_2 + p_a \sum_{t=2}^{\infty} N_2 \delta^{t-1}, \tag{2.8}$$

given that the trustor plays the strict grim-trigger strategy.

In order for the strict grim-trigger strategy to be an equilibrium strategy, it must be that $V_1 > V_2$, or:

$$\frac{(1-\delta)C_2 + \delta N_2 p_h}{(1-\delta)\Big(1 - \delta(1-p_h)\Big)} > \frac{(1-\delta)A_2 + \delta N_2 p_a}{(1-\delta)\Big(1 - \delta(1-p_a)\Big)}. \tag{2.9}$$

This leads to:

$$\delta > \bar{\delta} \equiv \frac{(A_2 - C_2)}{(A_2 - C_2) - p_h(A_2 - N_2) + p_a(C_2 - N_2)}. \tag{2.10}$$

When condition (2.10) is satisfied, the selfish trustee will prefer to choose *Honour* in every period of the game. The trustor will play *Co-operation* until he observes an L. When condition (2.10) is not satisfied, the trustor will play *No co-operation* throughout the entire game, knowing that the selfish trustee will prefer to choose *Abuse* as soon as he gets the chance.

Next, we will include social preferences into equations (2.7) and (2.8). Given that the trustor plays the strict grim-trigger strategy, the trustee's expected pay-off from choosing *Honour*, Y_1, is given by:

$$Y_1 = C_2 + \rho E(C_1) + (1 - p_h)\delta V_1 + p_h \sum_{t=2}^{\infty}(N_2 + \rho N_1)\delta^{t-1}, \tag{2.11}$$

and the expected pay-off from choosing *Abuse*, Y_2, equals:

$$Y_2 = A_2 + \rho E(S_1) + (1 - p_a)\delta V_2 + p_a \sum_{t=2}^{\infty}(N_2 + \rho N_1)\delta^{t-1}. \tag{2.12}$$

The trustee will prefer to choose *Honour* when $Y_1 > Y_2$:

$$\frac{\Big(C_2 + \rho E(C_1)\Big)(\delta - 1) - \delta(N_2 + \rho N_1)p_h}{(\delta - 1)\Big(\delta(p_h - 1) + 1\Big)} > \frac{\Big(A_2 + \rho E(S_1)\Big)(\delta - 1) - \delta(N_2 + \rho N_1)p_a}{(\delta - 1)\Big(\delta(p_a - 1) + 1\Big)}, \tag{2.13}$$

or:

$$\delta > \hat{\delta}(\rho) \equiv \frac{(\widetilde{A_2} - \widetilde{C_2})}{(\widetilde{A_2} - \widetilde{C_2}) - p_h(\widetilde{A_2} - \widetilde{N_2}) + p_a(\widetilde{C_2} - \widetilde{N_2})}, \tag{2.14}$$

where, for simplification, we denote $A_2 + \rho E(S_1)$ by $\widetilde{A_2}$, $N_2 + \rho N_1$ by $\widetilde{N_2}$ and $C_2 + \rho E(C_1)$ by $\widetilde{C_2}$. It should be noted that co-operative behaviour is only possible under the condition that $1 > \hat{\delta} > 0$, or:

$$\frac{p_h}{p_a} < \frac{\widetilde{C_2} - \widetilde{N_2}}{\widetilde{A_2} - \widetilde{N_2}}. \tag{2.15}$$

Again, $\frac{\partial \hat{\delta}}{\partial \rho}$ is negative, thus the larger ρ, the more likely it is that conditional honest trustees will choose *Honour*. Solving (2.14) for ρ yields:

$$\rho > \hat{\rho} \equiv \frac{A_2\Big(\hat{\delta}(p_h - 1) + 1\Big) - C_2\Big(\hat{\delta}(p_a - 1) + 1\Big) + \hat{\delta}N_2(p_a - p_h)}{E(C_1)\Big(\hat{\delta}(p_a - 1) + 1\Big) - \hat{\delta}\Big(N_1(p_a - p_h) + E(S_1)(p_h - 1)\Big) - E(S_1)}. \tag{2.16}$$

If pay-off values are such that the interval $\frac{A_2 - C_2}{E(C_1) - E(S_1)} > \rho > \hat{\rho}$ exists, the conditional honest trustee will prefer to choose *Honour* in the indefinitely repeated *asymmetric* trust game.

The honest trustee prefers to choose *Honour* by definition. The trustor will choose *Co-operation* when he faces an honest trustee.

In the literature, it has been suggested that co-operative behaviour can also be established when players use a trigger strategy with limited punishment. In Ivaldi et al. (2003), the trustor punishes the trustee only for a limited number, D, of rounds upon observing an L, and he will return to choosing *Co-operation* until he observes the next L, when he will punish again and afterwards go back to choosing *Co-operation* etc. So, every time he is confronted with an L he will punish for D rounds and return to choosing *Co-operation* thereafter. Let Z_1 represent the trustee's expected pay-off when he chooses *Honour*:

$$Z_1 = C_2 + \rho E(C_1) + (1 - p_h)\delta V_1 + p_h\left(\left(\sum_{t=2}^{D+1}(N_2 + \rho N_1)\delta^{t-1}\right) + \delta^{D+1}V_1\right), \tag{2.17}$$

and Z_2 the expected pay-off when the trustee chooses *Abuse*:

$$Z_2 = A_2 + \rho E(S_1) + (1 - p_a)\delta V_2 + p_a\left(\left(\sum_{t=2}^{D+1}(N_2 + \rho N_1)\delta^{t-1}\right) + \delta^{D+1}V_2\right), \tag{2.18}$$

given that the trustor plays a trigger strategy with D rounds of punishment.

The trustee will only co-operate when $Z_1 > Z_2$, or:

$$\frac{(N_2 + \rho N_1)p_h\delta^{D+1} + \Big(C_2 + \rho E(C_1)\Big)(\delta - 1) - \delta(N_2 + \rho N_1)p_h}{(1 - \delta)\Big(p_h\delta^{D+1} + \delta(1 - p_h) - 1\Big)} >$$

$$\frac{(N_2 + \rho N_1)p_a\delta^{D+1} + \Big(A_2 + \rho E(S_1)\Big)(\delta - 1) - \delta(N_2 + \rho N_1)p_a}{(1 - \delta)\Big(p_a\delta^{D+1} + \delta(1 - p_a) - 1\Big)}. \tag{2.19}$$

The more lenient the trustor becomes, the larger the right-hand side (RHS) of (2.19) will become in comparison to the left-hand side (LHS). This makes intuitive sense, because the trustee will be able to earn more and, by definition, higher pay-offs when he chooses *Abuse*, compared to the pay-offs he will earn when choosing *Honour*. When the equilibrium condition is not satisfied in the case of infinite punishment, it will definitely not be satisfied in the case of punishment for D rounds, because choosing *Abuse* only becomes more rewarding. The interesting group comprises the conditional honest trustees. Here, two effects are at play: when $C_1 + C_2 > S_1 + A_2$, conditional honest trustees become more willing to co-operate for larger ρ's, because the LHS of (2.19) will become larger compared to the RHS. On the other hand, the lower D, the larger the RHS will become. Put another way, for higher values of ρ the equilibrium condition can be met with a small number of punishment rounds, but for lower values of ρ more rounds are required.

2.2.4 Trustor's belief about trustee's type

So far it has been assumed that the trustor knows which type of trustee he is facing, i.e. the *symmetric* and *asymmetric* trust games have been analysed as games of complete information. Now assume that the trustor does not know whether he is playing with an honest, conditional honest or a selfish trustee. In this section, the models will be revisited, assuming incomplete information about the trustee's type.

We will start with the *symmetric* trust game. If $\delta > \tilde{\delta}$ honest, conditional honest and selfish trustees prefer to respond with *Honour*. In this scenario, the trustee type is irrelevant – if the grim-trigger strategy is played, control aligns the behaviour of all trustees.

If $\delta < \tilde{\delta}$, the selfish trustee will prefer to choose *Abuse*. In Section 2.2.2 it was argued that for $\rho > \rho^*$ conditional honest trustees will prefer to choose *Honour*, because $\delta > \delta^*(\rho)$, while honest trustees prefer to choose *Co-operation* by definition.

If $\delta < \tilde{\delta}$, the trustor will nevertheless play the grim-trigger strategy if he believes, with probability θ, that a large part of the trustee population is either honest or conditional honest with $\delta > \delta^*(\rho)$. To be more precise, the trustor will choose *Co-operation* if his expected pay-off from this strategy is larger compared to his expected pay-off from choosing *No co-operation*, or:

$$\theta \sum_{t=1}^{\infty} E(C_1)\delta^{t-1} + (1-\theta)\Big(E(S_1) + \sum_{t=2}^{\infty} N_1\delta^{t-1}\Big) > \sum_{t=1}^{\infty} N_1\delta^{t-1}. \qquad (2.20)$$

Solving for θ yields:

$$\theta > \theta^* \equiv \frac{(N_1 - S_1)(1-\delta)}{(C_1 - S_1) + \delta(S_1 - N_1)}. \qquad (2.21)$$

If $\theta > \theta^*$, the trustor will prefer to play the grim-trigger strategy; hence, as soon as the trustor observes *Abuse*, he will stop choosing *Co-operation*.

In case of the *asymmetric* trust game, a similar argument can be made. If $\delta < \bar{\delta}$, the selfish trustee prefers to choose *Abuse*. The trustor knows that the conditional honest trustee prefers to choose *Honour* if $\delta > \hat{\delta}(\rho)$ and that the honest trustee will always choose *Honour*. Therefore, the trustor might be prepared to play the strict grim-trigger strategy depending on his belief, θ, about the share of honest and conditional honest trustees with $\delta > \hat{\delta}(\rho)$ in the trustee population.

Let W_1 represent the trustor's expected profit if the trustee responds with *Honour*:

$$W_1 = E(C_1) + (1 - p_h)\delta W_1 + p_h \sum_{t=2}^{\infty} N_1 \delta^{t-1}, \tag{2.22}$$

and W_2 represent the trustor's expected profit if the trustee responds with *Abuse*:

$$W_2 = E(S_1) + (1 - p_a)\delta W_2 + p_a \sum_{t=2}^{\infty} N_1 \delta^{t-1}. \tag{2.23}$$

The trustor will play the strict grim-trigger strategy if:

$$\theta W_1 + (1 - \theta) W_2 > \sum_{t=1}^{\infty} N_1 \delta^{t-1}, \tag{2.24}$$

or

$$\theta > \hat{\theta} \equiv \frac{\delta(N_1 + W_2) - W_2}{(W_1 - W_2)(1 - \delta)}. \tag{2.25}$$

As soon as the trustor observes an L he will stop choosing *Co-operation*.

When conditions (2.21) and (2.25) are satisfied, the trustor will play the (strict) grim-trigger strategy in both games. The equilibrium described here is a perfect Bayesian Nash equilibrium, but unlike the sequential equilibrium in the finitely repeated trust game of incomplete information, it is stationary in beliefs.[4] When the (strict) grim-trigger strategy is played, the trustor does not need to update his belief in order to determine what he should do; it is the (strict) grim-trigger strategy that dictates how the trustor should respond to what he observes.

The Folk Theorem tells us that there exist many equilibrium strategies in indefinitely repeated games when the probability that the game will continue is sufficiently large. Adding incomplete information about the trustee's type

[4]Other equilibrium strategies might exist that are not stationary in beliefs.

does not change the validity of the Folk Theorem. Although we assumed that for selfish trustees the probability that the game will continue is too small in order to establish co-operative behaviour, for conditional trustees this is not the case, and, hence, many equilibrium strategies can exist for sufficiently large ρs.[5] Although many more equilibrium strategies can exist, we decided to limit our exposition to the two strategies discussed above, because they are simple in nature and capture the idea of the trust problem.

The equilibrium strategies described in this section allow for co-operative behaviour solely due to control. The trustor's prior belief, about the proportion of honest and conditional honest trustees in the population, is used to determine if in expected terms it is profitable to play the (strict) grim-trigger strategy or not. The (strict) grim-trigger strategy dictates what the trustor must do at every moment in the game. The (strict) grim-trigger strategy does not state that the trustor should use the information he receives to update his prior belief, so the trustor will not use the information in this way. Learning, for which the updating of beliefs is crucial, does not play a role in the setting presented here.

2.3 Conclusion

In this chapter, the (strict) grim-trigger strategy has been described for the indefinitely repeated *symmetric* and *asymmetric* trust games. The models provide the following insights:

The larger the continuation probability, the easier it becomes for the trustor and the trustee to co-operate.

The utility function used in this chapter models the trustee's utility as a linear combination of his own monetary pay-offs and those of the trustor. The coefficient ρ indicates how much the trustee cares about the trustor's pay-offs. Our models show that for larger values of ρ the trustor and the trustee can achieve co-operation for lower values of the continuation probability.

Not being able to observe the trustee's response in the *asymmetric* trust game reduces the effectiveness of control and, hence, makes it more difficult for the trustor and the trustee to co-operate. A comparison of the models

[5]When the (strict) grim-trigger strategy is an equilibrium strategy, other strategies might also allow for co-operative behaviour. Given that these strategies will always be less punitive than the (strict) grim-trigger strategy, it will be more difficult for them to establish co-operation in equilibrium. When ρ becomes larger and approaches $\frac{A_2-C_2}{E(C_1)-E(S_1)}$, it becomes easier to establish co-operation and, consequently, more strategies can be supported in equilibrium.

for the *symmetric* and *asymmetric* trust games highlights that this result is caused by two different effects. First, when comparing the equilibrium conditions, it can be seen that the critical continuation probability for which the trustee prefers to choose *Honour* is lower in the case of the *symmetric* trust game compared to the *asymmetric* trust game. Second, given that the continuation probability is sufficiently high, it is less likely that co-operation can be supported in equilibrium in the *asymmetric* trust game, due to the fact that the grim-trigger strategy is stricter in terms of when the sanction is applied. As soon as the trustor observes an L, he will punish the trustee by choosing *No co-operation* until the end of the game. Furthermore, in the case of the absence of information asymmetry, the trustor will only activate his punishment strategy if he observes *Abuse*. Consequently, when co-operation is supported in both the indefinitely repeated *symmetric* and *asymmetric* trust games, it is more likely to last in the first. A limited number of punishment rounds can only solve this problem partially, because it makes it more difficult to support co-operation in equilibrium in the first place.

In the case of incomplete information, it is possible that the trustor will play the (strict) grim-trigger strategy, even when he knows that this strategy is not an equilibrium strategy for selfish trustees. The larger the trustee's belief about the share of honest or conditional honest trustees in the trustee population, the more likely it becomes that the trustor will play the (strict) grim-trigger strategy. This also makes intuitive sense, because when 100% of the trustee population is of the honest type, the trustor has a strong preference for choosing *Co-operation*. When the trustor's belief is sufficiently high, the (strict) grim-trigger strategy can be played. However, it will remain more difficult to sustain co-operative behaviour when the *asymmetric* trust game is compared to the *symmetric* trust game, due to the strictness of the grim-trigger strategy.

Chapter 3

The role of information asymmetry in indefinitely repeated trust games with checking: An experiment

3.1 Introduction

In this chapter, the experimental results of the influence of checking on co-operative behaviour in indefinitely repeated *asymmetric* trust games are presented. In the previous chapter, it was concluded that, because control is less effective, it is more difficult to establish co-operation in the indefinitely repeated *asymmetric* trust game compared to the indefinitely repeated *symmetric* trust game. This chapter tries to answer the question as to whether the checking option can help trustors to overcome the problem of information asymmetry by turning the *asymmetric* trust game into a *symmetric* trust game. In other words, we will investigate whether the checking option can increase the level of co-operation between trustors and trustees.

Secondly, we will test whether the two types of trustor and trustee behave differently when the (strict) grim-trigger strategy is not an equilibrium strategy compared to the situation where it is.

This chapter is structured as follows. In Section 3.2, the experimental design will be discussed and the hypotheses presented. The results can be found in Section 3.3. To be more precise, in Section 3.3.1, we explain how the experiment was carried out and some basic information on session level is presented. The variables are described in Section 3.3.2, while the main results can be found in Section 3.3.3. For a discussion of the results, see Section 3.4.

3.2 Experimental design and hypotheses

3.2.1 The design

The experiment is computer-based, using z-Tree by Fischbacher (1999) both for programming the treatments and running the experiment.

As mentioned in Chapter 1, the experimental design consists of three treatments. The two main treatments are the No-Checking Treatment (NCT) and the Checking Treatment (CT). The first one is the indefinitely repeated *asymmetric* trust game without the checking option, while the latter is the same game, but this time the trustor has an option to check on the trustee's decision. In this experiment, the checking option is the treatment variable.

In the CT, the trustor needs to decide whether he wants to use the checking option after both he and the trustee have made their decision, but before the pay-offs earned during that period are made known. When the trustor checks, he will still earn either an H or an L due to the role of Nature, but this time it will also be revealed whether the trustee chose *Honour* or *Abuse*. If the trustor decided to check, the trustee is immediately notified of this. Table 3.1 summarises the information that subjects see on their computer screen during the different treatments.

The indefinite character of the game is simulated in the experiment by letting the subjects play a game with a continuation rule. All subjects know that after every round the game continues to the next round with a

Table 3.1: The information that subjects receive per treatment.

	What subjects see in the NCT
Trustor	his choice and his own pay-off
Trustee	his choice, trustor's choice and his own pay-off
	What subjects see in the CT
Trustor	his choice and his own pay-off
Trustor checks	his choice, trustee's choice and his own pay-off
Trustee	his choice, trustor's choice, his own pay-off and whether the trustor used the checking option or not

probability δ.[1] In order to increase the number of observations, the subjects play five rounds for certain. At the end of round 5, the game continues with $\delta = \frac{5}{6} \approx 0.83$. In expected terms the game will last ten rounds.[2] [3] Adding a number of certain rounds does not change the nature of the results presented in section 2.2 (see Appendix A for details).

Given the ratio of five fixed rounds out of ten expected rounds, as has been chosen here, it is assumed that the indefinite nature of the game is still preserved. Keeping the number of expected rounds fixed, a further increase of the number of certain rounds determines that, as a consequence, the continuation probability must be further decreased. This would result in a situation where it is questionable whether the subjects will regard the game as being indefinitely repeated.

Table 3.2: Parameter values for the indefinitely repeated *asymmetric* trust game experiment.

General parameters	
$p_a = \frac{3}{4}$	$p_h = \frac{1}{4}$
$H = 54$	$L = -10$
Trustor's (expected) pay-offs	Trustee's pay-offs
$E(C_1) = 38 = \frac{1}{4}(-10) + \frac{3}{4}(54)$	$C_2 = 22$
$E(S_1) = 6 = \frac{3}{4}(-10) + \frac{1}{4}(54)$	$A_2 = 44$
$N_1 = 15$	$N_2 = 10$

In the experiment, the values mentioned in Table 3.2 are used for the relevant parameters. When the trustor earns -10 or 54 in the NCT, it is not clear to him whether the trustee has responded with *Abuse* or *Honour* during that period. This is also made visible in Figure 3.1, where the game tree of the *asymmetric* trust game is shown with the parameter values of

[1]In this thesis, temporal preference is ignored. It is assumed that subjects will not be sensitive to time in an experiment that has a duration of approximately one hour. The results presented in Section 2.2.2 and Section 2.2.3 will not be affected by this simplification. δ can be replaced by a parameter that combines a discount factor with the continuation probability.

[2]The total expected number of rounds is equal to the number of certain rounds plus $\frac{1}{1-\delta}$ for the uncertain rounds.

[3]This number of expected rounds should yield enough observations to test the hypotheses. It is also similar to the continuation probabilities used in related studies. Engle-Warnick and Slonim (2004) and Slonim et al. (2006) use $\delta = 0.8$, while Aoyagi and Fréchette (2008) use $\delta = 0.9$.

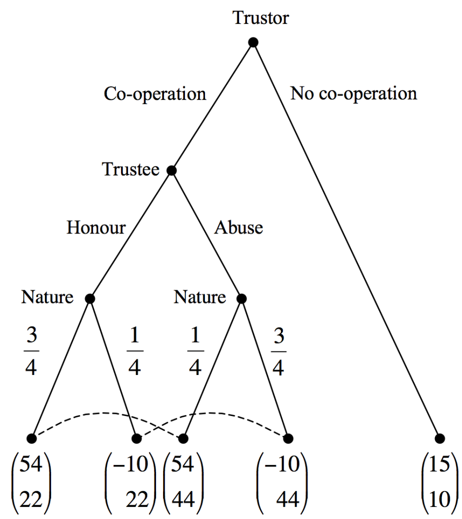


Figure 3.1: *Asymmetric* trust game with pay-offs for the indefinitely repeated *asymmetric* trust game experiment.

Table 3.2.

These particular values were chosen in order to acquire, from a theoretical point of view, discriminating hypotheses with respect to the level of co-operative behaviour between the NCT and the CT (see also Section 3.2.2).

To distinguish selfish from honest trustees in the data, each session starts with a type treatment (TT). In this treatment, the subjects play an adjusted isolated encounter trust game. The TT also uses the pay-off from Table 3.2. Nature has no move in the TT. The trustor either earns 38 if the trustee responds with *Honour* or 6 if the response is *Abuse*.

The adjusted isolated encounter trust game is based on the strategy method (Selten (1967)), and played in the following way: After the structure of the entire game has been explained, all subjects are asked to make the decision of the trustor. Next, they are asked to make the decision of the trustee, under the assumption that the trustor has chosen *Co-operation*. After the subjects have made both decisions, they are randomly assigned to a role and randomly matched with someone with a different role. Subjects' earnings are calculated using the decision made in their own role, linked to the decision made by their partner in his role.

In this way the equilibrium strategy remains the same, but for all subjects both trustor and trustee decisions can be observed.[4] If we had used the standard isolated encounter game, where roles are assigned before the game starts, it would have been impossible to observe the reaction of trustees whose trustor chose *No co-operation*. The advantage of acquiring data on all information sets of a game while using the strategy method has also been mentioned by Roth and Kagel (1995).

According to the definitions presented in Section 1.2.4, a subject in the role of trustor is considered trustful if he chooses *Co-operation* in the TT and a subject in the role of trustee is labelled honest if he choses *Honour*.

In the experiment, we used a 'within subjects' design to control for subject personal idiosyncrasies. To control for the order effect, two different types of sessions where run. In our crossover design, each session started with the TT, while in Type A sessions, the subjects played the NCT before the CT, this order was reversed in the Type B session (see Table 3.3). All sessions ended with a questionnaire, which consisted of a number of basic questions on demographics and questions on trust and trustworthiness attitudes. The latter questions were based on Yamagishi and Yamagishi (1994). Finally, we added some questions on checking.

Table 3.3: Ordering of the treatments.

Type A session
(TT) Type game
(NCT) Asymmetric trust game without checking option (2×)
(CT) Asymmetric trust game with checking option (2×)
Questionnaire
Type B session
(TT) Type game
(CT) Asymmetric trust game with checking option (2×)
(NCT) Asymmetric trust game without checking option (2×)
Questionnaire

[4]Bohnet and Huck (2004) used a different method. They took the decision made by the trustor in the first round of their 20 times-repeated trust game as the type indicator. For trustees they took the average rate at which they chose *Honour* during the first ten periods. This method is not without risk, given the fact that co-operation levels in repeated games are higher compared to isolated encounter game, and might lead to an overestimation of the number of trustful trustors and honest trustees. A more practical reason why we did not choose their method is the fact that the subjects play both roles in this experiment.

To increase the observations of personal characteristics, the NCT and the CT were played twice during a session. When a game was played for the second time, subjects switched roles, so each subject played a treatment once in the role of trustor and once in the role of trustee.

To prevent reputation building between subjects over the duration of a session, the subjects were rematched after each repeated game they played. Including the type game, all subjects were rematched five times with a new partner.

The instructions for all three treatments are enclosed in the Appendices C.1, C.2 and C.3. The questionnaire can be found in Appendix E.

3.2.2 Hypotheses

In the CT, the trustor has an option to check on the behaviour of the trustee. For the reasons mentioned in Section 1.6.2, we made this checking option costless. When the trustor checks every round he will have the same information as in the *symmetric* trust game. Although his pay-offs in the case of *Co-operation* still depend on a probability distribution, the trustor can turn the *asymmetric* trust game into a *symmetric* trust game, the latter of which is strategically equivalent to the isolated encounter trust game.

The equilibrium condition for the CT, under the assumption of constant checking, can be calculated by entering the parameter values from the experiment in equation (2.3). For selfish trustees, $\rho = 0$, $\tilde{\delta} > \frac{11}{17} \approx 0.65$. We implemented $\delta \approx 0.83$ such that $\delta \approx 0.83 > \tilde{\delta} \approx 0.65$; hence, selfish trustees should prefer to choose *Honour* when the trustor plays the grim-trigger strategy. Substituting the parameter values in equation (2.10) yields $\bar{\delta} \approx 0.98$; thus, for selfish trustees, choosing *Honour* is no longer an attractive option in the NCT where the game is asymmetric. Honest trustees prefer to choose *Honour* by definition, regardless of the treatment they are in.[5] In Table 3.4, the trustee's equilibrium strategies for both treatments are summarised.[6]

Given the parameter values used in this experiment, the grim-trigger strategy is an equilibrium strategy for selfish trustees in the CT, while this

[5] As we are not able to observe the true ρ we cannot identify conditional honest trustees. Therefore, the analysis in this chapter limits itself to the distinction between honest and selfish trustees.

[6] Recall that the first five periods are played for certain in the experiment. The actual values of $\tilde{\delta}$ and $\hat{\delta}$ can be found in Appendix A. Note, however, that given these values equilibrium play does not change. As a result, in the case of the CT the grim-trigger strategy is an equilibrium strategy, while in the case of the NCT it is not.

Table 3.4: Trustee's response when trustor plays (strict) grim-trigger strategy

	Equilibrium strategies	
	Honest trustee	Selfish trustee
NCT	*Honour*	*Abuse*
	$\forall\delta$	$0.83 \approx \delta < \bar{\delta} \approx 0.98$
CT	*Honour*	*Honour*
	$\forall\delta$	$0.83 \approx \delta > \tilde{\delta} \approx 0.65$

is not the case for the strict grim-trigger strategy in the NCT. Trustors, who believe they only face selfish trustees, should not choose *Co-operation* in the NCT. However, when the trustor believes that the proportion of honest trustees is sufficiently large, the trustor can still play the strict grim-trigger strategy (see section 2.2.4).

Checking makes it impossible for trustees to hide their untrustworthy behaviour behind the cloud of information asymmetry. Checking thus allows for more efficient punishment and therefore forces the trustee to behave in a more trustworthy way. Given that checking enhances the effectiveness of control, we come to the following two hypotheses:

HYPOTHESIS 1: *Trustors choose* Co-operation *more frequently in the CT compared to the NCT.*

HYPOTHESIS 2: *Trustees choose* Honour *more frequently in the CT compared to the NCT.*

Given the fact that the grim-trigger strategy is an equilibrium strategy for selfish trustees in the case of the CT, non-trustful trustors should be equally prepared to choose *Co-operation* as trustful trustors. For the same reason, both selfish and honest trustees must prefer to choose *Honour*.

HYPOTHESIS 3a: *In the CT, no distinction can be made in the behaviour of the different trustor and trustee types.*

In the NCT, where the strict grim-trigger strategy is no equilibrium strategy for selfish trustees, the different types of trustors and trustees should make different decisions.

HYPOTHESIS 3b: *In the NCT, trustful trustors and honest trustees choose* Co-operation *and* Honour *more frequently compared to non-trustful trustors and selfish trustees.*

The grim-trigger strategy suggests that it should be less likely to observe *Co-operation* and *Honour* later on in the game. When this strategy is played, trustors should stop choosing *Co-operation* as soon as the sanction rule is triggered. This is more likely in case of the NCT, where the grim-trigger strategy is far stricter. This leads to the following hypotheses:

HYPOTHESIS 4a: *Over time, the frequency with which* Co-operation *and* Honour *is chosen decreases in both NCT and CT.*

HYPOTHESIS 4b: *The decrease in the frequency with which* Co-operation *and* Honour *is chosen is stronger in the NCT.*

How should the trustor react given the information he receives? The moment the sanction is triggered should be related to the observation of a low pay-off, L, in the NCT, while in case of the CT it should be related to observing *Abuse*. The trustor's type should not matter for this decision, because the (strict) grim-trigger strategy dictates that both types of trustors should react in the same way.

HYPOTHESIS 5a: *In the NCT, trustors, regardless of type, are less likely to choose* Co-operation *after observing an L.*

HYPOTHESIS 5b: *In the CT, trustors, regardless of type, who have checked and observed* Abuse *are less likely to choose* Co-operation.

3.3 Results

3.3.1 Carrying out the experiment

The experiment was conducted in six sessions during the last two weeks of May 2007 at ELSE.[7] In total, 114 subjects participated in the experiment, all of whom had registered at the ELSE subject pool.[8] Although some

[7]Experimental Laboratory for Sociology and Economics at Utrecht University.

[8]ELSE recruits from among the entire student population of Utrecht University (for more details, see www.elseutrecht.nl).

subjects had participated in other experiments, none had participated in binary trust game experiments.

The maximum number of subjects the computer laboratory of Utrecht University can accommodate is 20. Half of the sessions had 20 subjects, the other half 18 subjects. 85.09% of the subjects have the Dutch nationality, including people with double nationalities. The 17 foreign subjects come from 11 different nationalities, from amongst which no large group can be distinguished. The experiment was bilingual in the sense that the subjects could choose to do the experiment in either Dutch or in English. Not only were the instructions available in both languages, but also the subjects could choose whether the messages that appeared on their computer screens were in Dutch or English. Women were in the majority, 68.42%, versus 31.58% men. 23.94% of the subjects indicated that he or she studied economics.

The subjects were told that they could not communicate with each other during the experiment. Next to this announcement the following active measures were taken, which successfully prevented communication between the subjects. Before the subjects entered the laboratory, numbers were handed out at random. The number a subject received was linked to a computer. By handing out the numbers randomly, the situation where people who might know each other decide to sit closely together was avoided. In addition, dividers separated the computers so that the subjects could not give signals to each other during the experiment.

When subjects entered the laboratory they found the instructions for the TT on their desk. After this treatment they received, depending on the session type, instructions for either the NCT or the CT, combined with a small quiz to test whether they had understood the instructions. After the second treatment they received the instructions for the final treatment. Subjects were given ample time to read the instructions. Before a treatment started, the instructor gave a plenary explanation of the key features of the instructions. This was always done by the same person to maintain uniformity in the explanation between sessions. To avoid framing effects, abstract labels were used for players' names and strategies in the instructions (see Appendices C.1, C.2 and C.3).

The continuation rule led to different game lengths; the shortest game lasted five rounds and the longest 36 rounds.

Each session ended with a questionnaire. In the meantime, the instructors counted out the money earned by the subjects. After each subjects had answered the questionnaire, they received their earnings. Subjects were paid according to the outcomes of the experiment. While playing the trust game, subjects earned points, which were translated into euro cents after-

Table 3.5: Some basic information on the session level for the indefinitely repeated *asymmetric* trust game experiment.

	Session 1	Session 2	Session 3
Date	21-5-07	23-5-07	23-5-07
Session type	type A	type A	type A
Subjects	20	20	20
Economics students	10%	15%	5%
Male subjects	30%	35%	25%
Trustful trustors†	35%	75%	80%
Honest trustees†	30%	40%	20%
Duration*	5/7/9/9	5/8/6/17	7/6/9/7
Average earnings	€11.04	€13.54	€9.29
Maximum earnings	€14.51	€20.73	€11.54
Minimum earnings	€6.89	€8.92	€6.89
	Session 4	Session 5	Session 6
Date	24-5-07	30-5-07	30-5-07
Session type	type B	type B	type B
Subject	18	18	18
Economics students	27.78%	27.78%	50%
Male subjects	27.78%	22.22%	50%
Trustful trustors†	55.56%	44.44%	55.56%
Honest trustees†	33.33%	22.22%	16.67%
Duration*	11/11/29/36	13/6/21/5	6/6/10/14
Average earnings	€26.96	€15.81	€12.52
Maximum earnings	€33.78	€18.78	€15.88
Minimum earnings	€19.44	€10.91	€8.73

*Duration of each game in number of rounds. †Type as derived from the TT.

wards. On average, the subjects earned €14.67 per session.[9] [10] A session

[9]In the questionnaire we asked the subjects about their monthly income. Only 7% refused to answer this question. 4.39% had an income of more then €1000 a month. The other four categories with the scores between brackets are: €0-€250 (31.6%), €250-€500 (31.6%), €500-€750 (17.5%) and €750-€1000 (7.9%). The subjects' earnings seem to be adequate given the average monthly income. Another indication is the wage of student assistants. Student assistants, a popular job among students given the relatively high wage, earn about €12.50 gross an hour. For most students gross becomes net, because when they work a limited number of hours their income taxes do not usually exceed the levy discount.

[10]In experimental economics, monetary rewards are preferred, because money is consid-

lasted between 60 and 75 minutes. Approximately 15 minutes were used for instructional purposes. For a more detailed overview of the information discussed above, see Table 3.5.

3.3.2 Variables and descriptive statistics

In this section, we determine four categories of variables. For an overview of the summery statistics of all the variables, see Table 3.6.

There are two different dependent variables: *ChoiceA* is a dummy variable, which equals 1 if the trustor chooses *Co-operation* and zero otherwise. *ChoiceB* is a similar variable, only this time it is for trustees choosing *Honour*, which is of course only possible after trustors chose *Co-operation*.

Next, we come to the independent variables. The variables that are the focus of our attention are: *Checktreat*, a dummy variable that is 1 if a subject is in the CT. *Period* indicates the period subjects are playing in. The TT gave us a trustor and trustee type-indicator. Table 3.5 lists the percentage of trustful trustors and honest trustees per session. *Trustful* is a dummy variable that is 1 if the trustor is trustful. *Honest* is a similar variable for honest trustees. *OutcomeL−lag1* is a dummy variable that is 1 if the trustor received an *L* in the previous round. *Abuse−lag1* is a dummy variable indicating when the trustor has used his checking option in the previous period and observed that the trustee chose *Abuse*.

Because some features of the experimental design might influence the dependent variables, we generated some (main) control variables. To correct for the difference between Type A and Type B sessions, we created the dummy *Checkfirst*, which is 1 for sessions of Type B. Subjects switched roles several times during the experiment. To see whether their first role influenced subject behaviour, we created the *Trustorfirst* dummy that indicates if a subject was first assigned to the role of trustor. As mentioned in Section 3.2.1, the subjects played five rounds for certain before the continuation rule applied. This change might have an effect on subject behaviour. In order to correct for this we added the *Period6ormore* dummy, which is 1 for all periods after period 5 and zero otherwise.

The questionnaire provided us with basic demographics as well as some

---

ered a neutral reward. According to economic theory, subjects are assumed to maximise their utility. In the case of a monetary reward, this would mean that subjects will try to earn as much money as possible in an experiment. This is closely related to the dominance condition from the induced value theory, which states that changes in subjects' utility from the experiment come predominantly from the monetary reward and other influences are negligible (see Friedman and Sunder (1994)).

Table 3.6: Overview of variable descriptive statistics for the indefinitely repeated *asymmetric* trust game experiment.

Variable name	Obs	*Mean*	Std. Dev.	Min	Max
Dependent variables					
ChoiceA*	3614	0.611	0.488	0	1
ChoiceB*	2208	0.465	0.499	0	1
Independent variables					
Checktreat *	3614	0.516	0.500	0	1
Period†	3614	4.694	2.622	1	10
Trustful*	3614	0.575	0.494	0	1
Honest*	3614	0.273	0.446	0	1
OutcomeL−lag1*	990	0.502	0.500	0	1
Abuse−lag1*	3158	0.132	0.338	0	1
Main control variables					
Checkfirst*	3614	0.513	0.500	0	1
Trustorfirst*	3614	0.500	0.500	0	1
Period6ormore*	3614	0.369	0.483	0	1
Other control variables					
Female*	3614	0.685	0.465	0	1
Economics*	3614	0.227	0.419	0	1
Dutch*	3614	0.848	0.359	0	1
Age	3614	21.756	4.879	16	59
Friends	3614	0.339	0.616	0	2
Work*	3614	0.606	0.489	0	1
Volwork*	3614	0.431	0.495	0	1
Sequence*	3614	0.506	0.500	0	1
Donor*	3614	0.125	0.330	0	1
Organdonor*	3614	0.487	0.500	0	1
Religious*	3614	0.298	0.457	0	1

Note: Obs = Number of observations; Std. Dev. = Standard deviation; Min = minimum value; Max = maximum value and * indicates dummy variables, 1=yes.
† As explained in Section 3.3.3, the analysis presented in this Chapter makes use only of the observations acquired during the first ten rounds. We applied the same selection criteria when making this table.

other control variables. *Female* is a dummy variable that is 1 if the subject is female. *Economics* and *Dutch* are also dummy variables, indicating whether a subject studies economics or has Dutch nationality. *Age* indicates the age of a subject, while *Friends* depicts the number of people present in the laboratory that a subject knows by first name. *Work* and *Volwork* are both dummy variables. The first tells us whether the subject has a (part-time) job, the second whether the subjects does voluntary work. We also asked one question to test the subjects' intuition for event sequences given a specified probability. We asked them the following question: The following two random sequences are the result of tossing a fair coin 8 times, where Heads and Tails are represented by H and T, respectively. Sequence I: HTHTTHTH, Sequence II: HHTTTTTT. Which of the two sequences is more likely to occur: (A) Sequence I, (B) Sequence II or (C) Both sequences are equally likely to occur. The dummy variable *sequence* indicates the subject who answered the sequence question correctly (answer (C)). Next to *Volwork*, the following three variables were also included as possible indicators for the social orientation of the subjects and to serve as a robustness check for the type indicators. The dummies *donor* and *organdonor* tell us whether a subject is a blood donor and whether they have filled in a organ donor card. Finally, the dummy *religious* is 1 if the subject considers himself or herself a religious person.

3.3.3 Data analysis

The fact that in the experiment an indefinitely repeated game has been tested, has as a consequence that the panel is unbalanced. A simple solution to this problem would be to base the estimations only on the balanced subpanel. However, a substantial amount of observations will be lost if the panel is limited to the first five rounds.

The missing observations are, fortunately, missing at random, because the computer randomly determined when a game stopped. In other words, conditioning upon the outcome of the selection process does not affect the conditional distribution of the dependent variable given the independent variables (see Verbeek (2000)). As a result, the entire panel can be used.

To test the hypotheses, a comparison between the treatments will be made. However, for periods 11 through to 36 there are mainly observations for the NCT and almost no observations for the CT. We found that this has an influence on the treatment comparison. In later rounds, subjects co-operate on average less. When the data contain many more NCT observations for the later rounds compared to the CT, the *checktreat* variable suggests a stronger difference between the treatments than is actually present. The panel used in the analysis is, for this reason, limited to the first ten periods.

A first investigation of the data is given in Table 3.7, which presents the average rate of *Co-operation* and *Honour* for both the NCT and CT. Compared to Engle-Warnick and Slonim (2004), who for their first supergame report an average *Co-operation* rate of 0.91 and an average *Honour* rate of 0.77, the average *Co-operation* and *Honour* rates in the CT are relatively low. Engle-Warnick and Slonim (2004) used a continuation probability of 0.8. Experimental results of indefinitely repeated prisoner's dilemma games show lower first round co-operation rates, which corresponds with the *Honour* rates we observe. Dal Bó (2005) reports for a continuation probability of 0.75 a co-operation rate of 0.40, and Duffy and Ochs (2008) report a co-operation rate of 0.48 for a continuation probability of 0.9. In Figure 3.2, the same information is displayed in more detail reporting the rate of *Co-operation* and *Honour* per period. The results of the Wilcoxon-Mann-Whitney (WMW) ranksum test using period averages indicate that the difference between the *Co-operation* and *Honour* rates per period are not significant.[11]

Table 3.7: *Co-operation* and *Honour* rates for both NCT & CT in the indefinitely repeated *asymmetric* trust game experiment.

	Co-operation	*Honour*
NCT	0.610	0.423
CT	0.612	0.504

In Section 3.2.2, it was concluded that the grim-trigger strategy is an equilibrium strategy in the case of the CT for all types of players. In the NCT, the strict grim-trigger strategy is no equilibrium strategy for selfish trustees, but the trustor might still play this strategy if condition (2.25) is satisfied. If this condition is not satisfied, the trustor can always choose to play the isolated encounter equilibrium strategy of *No co-operation*.

In the NCT, in 7 out of the 114 groups[12] that were formed during the experiment, the trustor decided not to choose *Co-operation* in any of the

[11] WMW ranksum test results: *Co-operation* p=0.7913 and *Honour* p=0.3845

[12] During each session, four games were played. In total, 114 subjects participated, which means 57 groups per game and 228 groups in total. Half of these were NCT groups and the other half CT groups.

rounds. In the same treatment, only in 3 groups did trustors play the strict grim-trigger strategy. In the CT, in 6 out of the 114 groups the trustor chose *No co-operation* in every period of the game, while in 9 groups the trustor played the grim-trigger strategy.

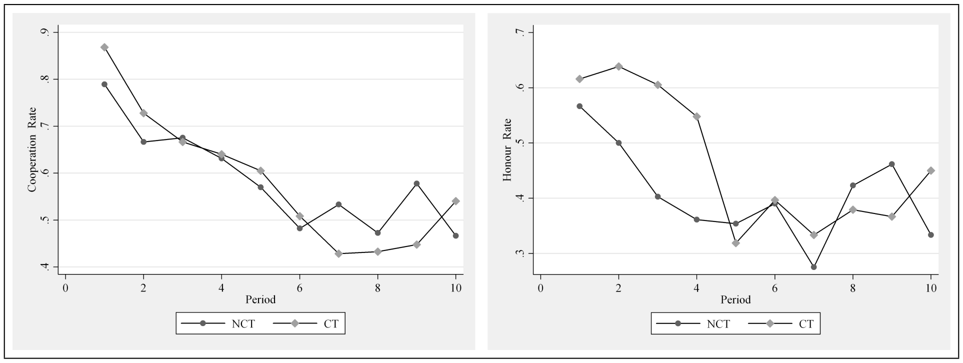


Figure 3.2: *Co-operation* and *Honour* rates over time in the indefinitely repeated *asymmetric* trust game experiment.

The majority of the groups did not play the (strict) grim-trigger strategy throughout the entire game. However, when the set of trustor decisions was observed, it was evident that most trustors violated the (strict) grim-trigger strategy later on in the game. To judge whether trustor behaviour corresponded with the (strict) grim-trigger strategy, the following measurement can be used. We count the proportion of decisions that are in line with the (strict) grim-trigger strategy from the start of a game until the first violation. So, once the (strict) grim-trigger strategy is violated, all decisions made afterwards are counted as violations. Using this stringent approach, 31.84% of all the decisions made by trustors in the NCT are in line with the strict grim-trigger strategy. When the same calculation is made for the CT, 37.30% of the trustors' decisions are in line with the grim-trigger strategy, which is somewhat higher compared to the NCT.

Alternatively, we can look at the proportion of decisions that are in line with the (strict) trigger strategy. When the trustor plays a (strict) trigger strategy, he should punish the trustee when he observes *Abuse* in the CT or an L in the NCT, but he does not need to punish the trustee forever. For this less stringent measurement all violations are counted, but once the decisions are in line again with the (strict) trigger strategy, they are no longer considered to be part of the violation.[13] Using this measurement,

[13]To make the difference between the two ways of measuring clear, we will give an

68.52% of trustors' decisions are in line with the strict trigger strategy in the NCT, which hardly differs from the CT where 69.05% are in line with the trigger strategy.

These results support earlier findings, for example Engle-Warnick (2007) also found evidence for the trigger strategy in indefinitely repeated trust games.

Before the data are studied in more detail, the appropriate statistical model will be discussed briefly. The data can be characterised in the following way: First, the subjects faced discrete decisions in the experiment. Second, the observations were repeated within subjects, i.e. they became experienced. The appropriate statistical model to analyse the experimental data is the panel logit model including random effects or fixed effects at the level of subjects. This model takes the binary nature of the dependent variable and the unobserved individual heterogeneity into account. The random effects method is preferred, because the fixed effect method cannot deal with explanatory variables that are constant over time like, for instance, the dummy *Trustful*. The random effects model assumes that the unobserved effect is uncorrelated with the explanatory variables in all time periods. In the models estimated in this chapter and Chapters 5 and 6, this assumption holds unless stated otherwise, given the experimental nature of the research.

In the results reported here, we do not include the category 'other control variables', as mentioned in Table 3.6. The additional, unreported estimations where we include these control variables did not influence the main results.

Trustors checked 86% of the time they chose *Co-operation* in the CT. It seems that checking is a logical thing to do for most trustors. This is also supported by the answers we got to the following statement from the questionnaire: "When in a business relationship it is unclear to me what my business partner is doing, it is only natural for me to find this out, even when I trust him." Only 14.33% of the subjects disagreed with this statement, 21.45% were neutral about it and 64.21% agreed.[14] We also tested

example: Assume that until round 4 the trustee responded with *Honour* to the trustor's choice of *Co-operation*. In round 5, the trustee, however, responds with *Abuse*. In round 6 the trustor reacts correctly by choosing *No co-operation*. In round 7 the trustor violates the grim-trigger strategy, by choosing *Co-operation*. In rounds 8 until 10 the trustor's decisions are again in line with the trigger strategy. Using the first measurement, four decisions are counted as violations. In the case of the second, less stringent measurement, only one decision counts as a violation.

[14]A seven-point scale was used, going from completely disagree to completely agree. For expositional reasons we combined the first three and the last three categories into two

if control variables such as Female, Age, Trustful, etc. had any predictive power on checking behaviour. We did not find any significant results.

FINDING 1: *Trustors choose* Co-operation *more frequently in the CT compared to the NCT.*

Although the WMW-ranksum test showed a different result, is seems that after correcting for the experimental design characteristics as well as a time trend, both trustors that started with the NCT as well as those who started with the CT choose *Co-operation* significantly more frequently in the CT (see Table 3.8). The coefficient of the last group is a bit larger. Trustors who started with the CT were more likely to choose *Co-operation* in the NCT compared to trustors who started with the NCT, with a p-value of 0.007. This result can also be found in Figure 3.3, which presents the average rate of *Co-operation* per period, only this time differentiated according to the treatment order. Although Figure 3.2 suggests a non-linear effect for the *Co-operation* time trend, we found no proof for this in the additional estimations we ran.

FINDING 2: *Trustees choose* Honour *more frequently in the CT compared to the NCT.*

At the 10% significance level, trustees were more likely to choose *Honour* in the CT, regardless of the treatment they began in. Given the fact that the variable *Honest*, including its interactions, is not significant, we also ran an estimation without these three variables. The significance level of *Checktreat* and the constant improved. In this alternative estimation these variables are significant at the 5% level. For the complete estimation, see Table 3.9.

In the CT, both trustors and trustees are more likely to choose *Co-operation* and *Honour*, which raises the question as to whether checking has a positive effect on the earnings of the trustor and the trustee.[15] In the CT, trustees earned on average 24.23 points per round compared to 25.02 points in case of the NCT – trustees were worse off. Trustors, on the other hand, earned 18.07 points in the NCT compared to 19.76 points in the CT, indicating that checking was beneficial to the trustors. This is not surprising given that checking should make control more effective and, hence,

large categories.

[15]The results presented here are based on the first ten periods.

Table 3.8: Random-effects logit model for the probability of choosing *Co-operation* in the indefinitely repeated *asymmetric* trust game experiment.

Dependent variable: ChoiceA				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Checktreat	0.916	0.293	3.12	0.002
Period	−0.267	0.062	−4.31	0.000
Checktreat×Period	−0.039	0.045	−0.85	0.393
Trustful	1.040	0.355	2.93	0.003
Trustful×Period	−0.076	0.046	−1.65	0.100
Trustful×Checktreat	−0.532	0.230	−2.31	0.021
Checkfirst	0.766	0.283	2.71	0.007
Checkfirst×Checktreat	−0.626	0.237	−2.65	0.008
Trustorfirst	−0.151	0.248	−0.61	0.543
Period6ormore	−1.810	0.526	−3.44	0.001
Period6ormore×Period	0.285	0.083	3.42	0.001
Constant	0.940	0.355	2.65	0.008
Number of observations: 1807, Number of subjects: 114				
LL: -1040.513, σ_u: 1.157, ρ : 0.289⋆				

⋆ LL: Log Likelihood, σ_u: Standard deviation of the error term at the subject level, ρ: proportion of the variance in the composite error term that is at the subject level, or $\rho = \frac{\sigma_u^2}{\sigma_u^2+\sigma_e^2}$.

Table 3.9: Random-effects logit model for the probability of choosing *Honour* in the indefinitely repeated *asymmetric* trust game experiment.

Dependent variable: ChoiceB				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Checktreat	0.550	0.316	1.74	0.081
Period	−0.313	0.069	−4.53	0.000
Checktreat×Period	−0.089	0.056	−1.58	0.115
Honest	0.312	0.428	0.73	0.466
Honest×Period	−0.035	0.062	−0.56	0.573
Honest×Checktreat	0.208	0.315	0.66	0.508
Checkfirst	0.026	0.321	0.08	0.937
Checkfirst×Checktreat	0.175	0.302	0.58	0.562
Trustorfirst	−0.245	0.273	−0.90	0.369
Period6ormore	−2.211	0.751	−2.94	0.003
Period6ormore×Period	0.418	0.113	3.70	0.000
Constant	0.620	0.354	1.75	0.080
Number of observations: 1104, Number of subjects: 114				
LL: -680.842, σ_u: 1.211, ρ: 0.308				

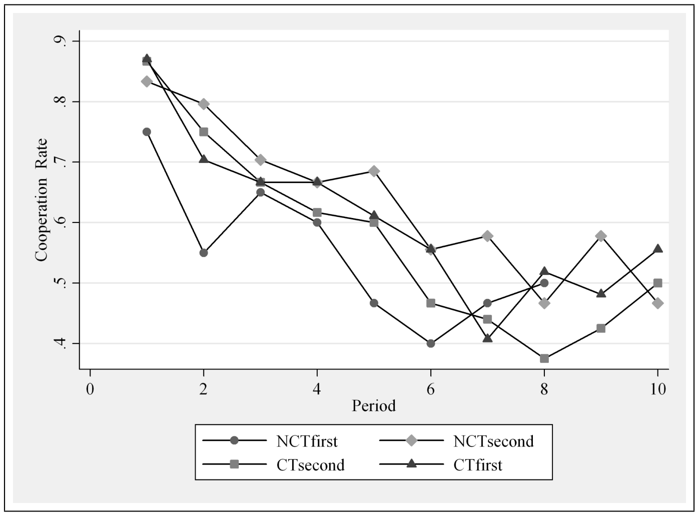


Figure 3.3: *Co-operation* rates over time correcting for treatment order. Trustors that had the NCT as second treatment were more likely to choose *Co-operation* compared to those who started with the NCT.

reduce opportunistic behaviour on the part of trustees. However, according to the WMW-ranksum test, there is no significant difference between the underlying distributions, with p=0.3071 and p=0.2821, respectively.

Next, we look at the total income of a group, the earnings of a trustor and a trustee together. In the NCT, the trustor and the trustee earned on average 331.90 points together in a game, while in the CT they earned on average 356.25 points. This difference is not significant according to the WMW-ranksum, with p=0.1221.

FINDING 3a: *In the CT, no distinction can be made in the behaviour of the different trustor and trustee types.*

FINDING 3b: *Trustful trustors are more likely to choose* Co-operation *in the NCT. Honest trustees, however, are not more likely to choose* Honour *in the NCT.*

Trustful trustors are significantly more likely to choose *Co-operation.* At the 5% significance level the initial increase in the likelihood with which trustors choose *Co-operation* is smaller in the CT. However, a Chi^2-test on the combined effect of *Trustful* and *Trustful*×*Checktreat* showed that trustor's type has no effect in the CT. This is in line with the theoretical prediction that there should be no difference between non-trustful and trustful trustors in this treatment, due to the dominant effect of the control mechanism.

As was already mentioned above, the type indicator for honest trustees, including the interaction terms, are not significant in either treatment. This result is in line with Hypothesis 3a, which states that there is no difference between the two types of trustees in the CT.

Given that in the NCT the strict grim-trigger is no equilibrium strategy for selfish trustees, they should choose *Abuse*, while honest trustees choose *Honour*. While theory suggests a clear difference between the trustee types, the data do not support this view.

FINDING 4a: *Over time, the frequency with which* Co-operation *and* Honour *are chosen by the trustor and the trustee decreases in both the NCT and CT.*

Over time, both trustors and trustees are significantly less likely to co-operate in both treatments. This linear trend is primarily noticeable in the first five rounds. It seems that trustor and trustee behaviour has been influenced by the experimental design, because in the periods where the continuation probability is operational, trustors and trustees are significantly less likely to choose *Co-operation* or *Honour*, with p-values of 0.001 and 0.003, respectively. It seems that period 5 was a focal point for trustors and trustees to decrease the frequency with which they chose *Co-operation* and *Honour*. After the initial drop at the beginning of period 6, the frequency with which *Co-operation* and *Honour* were chosen increased, as a comparison of the coefficients of the variable *Period* and the interaction term *Period6ormore*×*Period* shows.

The result that co-operation decreases over time in indefinitely repeated social dilemma games is not uncommon (see, for example, Engle-Warnick and Slonim (2004) for indefinitely repeated trust games and Dal Bó (2005) for indefinitely repeated prisoner's dilemma games).

FINDING 4b: *In the NCT, co-operation does not decrease stronger over time compared to the CT.*

As can be seen in Tables 3.8 and 3.9, there is no difference between the treatments with respect to co-operation over time; *Period*×*Checktreat* is far from significant. Hypothesis 4b must therefore be rejected.

Table 3.10: Random-effects logit model for the probability of choosing *Cooperation* in the NCT after receiving an L in the previous period in the indefinitely repeated *asymmetric* trust game experiment.

Dependent variable: ChoiceA				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
OutcomeL$-$lag1	-2.168	0.465	-4.66	0.000
OutcomeL$-$lag1$\times$Trustful	-0.074	0.581	-0.13	0.898
Period	-0.110	0.120	-0.92	0.358
Trustful	-0.188	0.853	-0.22	0.826
Trustful$\times$Period	0.056	0.113	0.49	0.622
Checkfirst	0.355	0.343	1.04	0.300
Trustorfirst	-0.135	0.323	-0.42	0.675
Period6ormore	-0.388	0.487	-0.80	0.425
OutcomeLperiod1	-0.613	0.354	-1.73	0.083
OutcomeLperiod1$\times$Trustful	0.117	0.435	0.27	0.788
Constant	2.974	0.757	3.93	0.000
Number of observations: 480, Number of subjects: 107				
LL: -238.057, σ_u: 0.990, ρ: 0.230				

Table 3.11: Random-effects logit model for the probability of choosing *Cooperation* in the CT after observing *Abuse* in the previous period in the indefinitely repeated *asymmetric* trust game experiment.

Dependent variable: ChoiceA				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Abuse$-$lag1	-2.359	0.411	-5.74	0.000
Abuse$-$lag1$\times$Trustful	0.068	0.515	0.13	0.894
Period	-0.067	0.116	-0.58	0.563
Trustful	0.508	0.679	0.75	0.454
Trustful$\times$Period	-0.051	0.107	-0.48	0.631
Checkfirst	0.121	0.273	0.44	0.659
Trustorfirst	0.051	0.272	0.19	0.852
Period6ormore	-0.095	0.436	-0.22	0.827
Abuseperiod1	0.012	0.428	0.03	0.977
Abuseperiod1$\times$Trustful	-0.212	0.494	-0.43	0.668
Constant	2.067	0.581	3.55	0.000
Number of observations: 510, Number of subjects: 108				
LL: -257.287, σ_u: 0.688, ρ: 0.126				

Next, we take a closer look at the NCT and CT separately. The lagged endogenous variables, presented in Tables 3.10 and 3.11, are correlated with the independent effect. This means that the Random Effects (RE) logit model used so far would yield inconsistent estimates. Therefore, we follow the suggestion of Wooldridge (2002) and include the following regressors in the RE logit models presented in Table 3.10: *outcomeLperiod1* and its interaction with *Trustful*, where *outcomeLperiod1* indicates whether the trustor observed an *L* in period 1. *Abuseperiod1* is a similar variable that indicates whether the trustor observed *Abuse* in period 1. This variable and its interaction with *Trustful* are added to the regression presented in Table 3.11. By adding these additional regressors, the RE logit models yield consistent estimates.

FINDING 5a: *Trustors, regardless of type, are less likely to choose* Co-operation *after observing an L in the NCT.*

FINDING 5b: *Trustors, regardless of type, are less likely to choose* Co-operation *after observing* Abuse *in the CT.*

As can be seen in Table 3.10, trustors chose *Co-operation* significantly less frequently if they observed an *L* in the previous period compared to observing an H, with a p-value smaller than 0.001. Although this does not necessarily mean that trustors play a strict (grim) trigger strategy, as described in the previous chapter, observing an *L* does, however, trigger the right reaction.

We found a similar result for the CT, whereby if trustors respond to observing *Abuse* as expected, they are less likely to choose *Co-operation* in the next period (see Table 3.11).

In addition, we also tested whether trustors ignored their pay-off realisation when they used the checking option. When observing *Abuse*, the non-trustful trustors ignored the pay-off realisation, only caring about the fact that their trust had been abused. However, trustful trustors do care about their pay-off realisation. There is a small chance that they chose *Co-operation*, as long as they observed an *H*.

We also found that trustors, regardless of type, were significantly more likely to choose *Co-operation* when they observed *Honour* in the previous round, with a p-value of less than 0.001. In the case of observing *Honour*, trustors do care about their pay-off realisation. We found that both types of trustors were significantly more likely to choose *Co-operation* if they earned an *H* in the previous round, with a p-value of 0.006. Observing *Honour* and receiving an L have a positive effect on the frequency with which trustors choose *Co-operation*, but observing *Honour* and receiving an H have even greater influence.

3.4 Discussion and conclusion

Trustors seem to understand the fundamental difference between *asymmetric* and *symmetric* trust games. When trustors were able to use the checking option, they did so 86% of the time.

The data show that the presence of a checking option increases the frequency with which the trustors choose *Co-operation*. The most logical explanation for this result is that trustors understand that trustees are more likely to choose *Honour* in the CT, given the fact that control is more effective in this treatment. The data support the theoretical prediction that control becomes more effective, because trustees choose *Honour* more frequently in the CT.

The result on trustee types confirms the theoretical predictions, with the exception of the honest trustee in the NCT. The data suggest that honest and selfish trustees act the same, regardless of the treatment they are in, while theory predicts that this should only be true for the CT. Perhaps the subjects did not realise that the strict grim-trigger strategy was not an equilibrium strategy in the case of the NCT. It should be noted that this result is by no means uncommon. Dal Bó and Fréchette (2007) test the evolution of strategies in indefinitely repeated prisoner's dilemma games, comparing treatments where co-operation is supported in equilibrium, with treatments where it is not.

In order to test their hypothesis, subjects need to have experience with playing indefinitely repeated prisoner's dilemma games. Similar to Dal Bó (2005), continuation probabilities of $\frac{1}{2}$ and $\frac{3}{4}$ are used, which allows the subjects to play many, but relatively short, games.

The setting where the continuation probability, δ, determines whether co-operation is an equilibrium strategy can be compared with the setting presented here.[16] In order for the grim-trigger strategy to support co-operation in equilibrium $\delta \geq 0.72$, this condition is satisfied in the treatment where the continuation probability equals $\frac{3}{4}$, but not in the treatment where the continuation probability equals $\frac{1}{2}$. This is roughly similar to the setting used here where the continuation probability is the same in both treatments, but the equilibrium conditions for the grim-trigger strategy differ per treatment.

[16]In other treatments, Dal Bó and Fréchette (2007) change the pay-off structure.

Dal Bó and Fréchette (2007) find that co-operation decreases in treatments where it is not supported in equilibrium. However, in the first repeated game there is no significant difference between the treatments, whether co-operation is supported in equilibrium or not. When subjects believe that the strict grim-trigger strategy is an equilibrium strategy in the case of the NCT, the result that there is no difference between selfish and honest trustees is less surprising.

As theory predicts, trustors and trustees are less likely to co-operate over time. However, given that the (strict) grim-trigger strategy is not played consistently by many groups, this result might alternatively be explained by the fact that subjects have difficulty in co-ordinating their strategies. When one's opponent does not respond as one expects him to do, one might prefer to maximise short-run pay-offs. For example, when a trustee responds with *Honour*, it might be expected that the trustor will choose *Co-operation* again in the next period, but when he reacts with *No co-operation* the trustee might want to go for *Abuse* the next time he has the chance.

We also found that trustors react differently to pay-off realisations. Trustful trustors are still prepared to place trust, even if they have observed that their trust has been abused, as long as they have earned an H. Both types of trustors are even more likely to place trust in this period when they have earned an H and observed that the trustee chose *Honour* in the previous period.

These results can be explained by reinforcement learning models (see, for example, Bush and Mosteller (1955), Roth and Erev (1995) and Erev and Roth (1998)). The idea is that strategies that have been successful are played with greater frequency than their less successful counterparts. When the trustor has obtained an H with a specific action, it is more likely that he will choose that action again in the next period.

The fact that trustors react differently to pay-off realisations can also partially explain why subjects find it difficult to co-ordinate their strategies in both treatments. Trustees might expect to be punished after they choose *Abuse*, but perhaps they are not when a trustful trustor has earned an H. Given the fact that the trustee does not know the trustor's pay-off realisation, he might find this rather confusing.

Chapter 4

Modelling finitely repeated *asymmetric* trust games*

4.1 Introduction

As mentioned in Chapter 1, the *asymmetric* trust game is studied in this thesis under two different horizon perspectives. In this chapter, the *asymmetric* trust game is analysed as a finitely repeated game.

In general, many equilibrium strategies can exist in finitely repeated games of complete information (Benoît and Krishna (1985)). The prisoner's dilemma and the binary trust game, however, have a unique feature – the isolated encounter equilibrium coincides with the minmax strategy.[1] Due to this characteristic, there exists only one equilibrium strategy in finitely repeated trust games of complete information, namely the isolated encounter equilibrium strategy of (*No co-operation*, *Abuse*). When incomplete information about a player's type is added, more equilibrium strategies can exist. Kreps and Wilson showed that next to the isolated encounter equilibrium a co-operative strategy is also feasible. In this Chapter, we will extend the Kreps and Wilson model by incorporating asymmetric information about the trustee's decision.

Experimental results have showed that in isolated encounter trust games trustors choose *Co-operation* and trustees respond with *Honour*, as further exemplified by Berg et al. (1995) and Snijders (1996). As mentioned in Chapter 1, this result is explained in the literature by the suggestion that

*This chapter and the following chapter are based on the working paper *Finitely repeated asymmetric trust games: An experiment*, a joint work with Stephanie Rosenkranz.

[1]The binary trust game can also be described as a one-sided prisoner's dilemma.

next to selfish preferences people can also be motivated by social preferences. On the basis of the decisions made in the isolated encounter trust game we defined two types of trustees, namely the honest and the selfish trustee.

The models presented in this chapter also assume the presence of these two types of trustees. Although the honest trustees' pay-offs differ from the selfish trustees' pay-offs (see Figure 1.2), the analysis presented in this chapter does not require the specification of an underlying utility function that accounts for this difference, as was the case in Chapter 2. It only requires the assumption that honest trustees prefer to choose *Honour* in an isolated encounter trust game.

In Section 1.3, the two mechanisms of temporal embeddedness – learning and control – were discussed. In the finitely repeated trust game, control and learning can support co-operative behaviour between trustors and trustees. Unlike the models from Chapter 2, learning determines when the sanction is activated. In this sense, the models presented in this chapter differ from those presented in Chapter 2 where control was the dominant mechanism that allowed for co-opertative behaviour.

This chapter is structured as follows: In Section 4.2, the standard model for finitely repeated trust games of incomplete information is briefly presented. The analysis focuses on how selfish trustees can build up a reputation as being trustworthy by (temporarily) mimicking the behaviour of honest trustees. This is followed by the newly developed model for finitely repeated *asymmetric* trust games of incomplete information in Section 4.3, while the concluding remarks can be found in Section 4.4.

4.2 Finitely repeated *symmetric* trust game

In finitely repeated games of complete information, the last interaction has the same characteristics as the isolated encounter game. When backward induction is applied, theory states that co-operative behaviour is only possible in finitely repeated games, if it is also an equilibrium in the isolated encounter game. Hence, repeating social dilemma games of complete information has no positive effect on the level of co-operative behaviour in these games.

The situation changes when the dilemma game is transformed into a game of incomplete information. In their ground-breaking work, Kreps et al. (1982) and Kreps and Wilson (1982a) show that rational co-operation is also possible for selfish players in a finitely repeated game, as long as the opponent knows there is a (small) probability that he is interacting with an honest player. The honest player is defined in such a way that he prefers to co-operate in every period.[2]

Later papers describe the sequential equilibrium [3] for the trust game, as seen in Camerer and Weigelt (1988) and Bower et al. (1996). For an extensive proof of the sequential equilibrium in a trust game we refer to Bower et al. (1996). In this section, the intuition behind the equilibrium concept will be briefly explained. Following this, the trustor's threshold for choosing *Co-operation* and the mixing probabilities for both the trustor and trustee will be presented.

The finitely repeated *symmetric* trust game is solved by backward induction, starting with the final period. Whether the trustor will place trust in the final period depends on his belief about the proportion of honest trustees in the population. If the belief is sufficiently high, the trustor will prefer to choose *Co-operation* above *No co-operation.*

Past events have, of course, an influence on the trustor's belief. It is assumed that throughout the game the trustor updates his belief using Bayes's Rule. When both honest and selfish trustees choose *Honour*, it is, however, impossible for the trustor to update his belief.

If the selfish trustee knows that the trustor's belief is insufficiently high enough for the trustor to choose *Co-operation* in the coming period, the trustee can actively change this situation by randomising over his actions in this period. When the selfish trustee mixes between choosing *Honour* and *Abuse*, he gives the trustor the opportunity to update his belief.

In equilibrium it will be the case that the trustee will randomise over his actions by choosing a probability that will make the trustor indifferent between choosing *Co-operation* and *No co-operation* in the next period. In a mixed strategy equilibrium, both players must be indifferent. In order for the trustee to play a mixing strategy in this period, the trustor needs to randomise between *Co-operation* and *No co-operation* in the next period.

Whether the trustor will choose *Co-operation* in a given period depends on his expected pay-off. The expected pay-off is determined by the trustor's belief, as well as by the trustee's mixing probability. Thus, the trustor's threshold for choosing *Co-operation* changes over time, which is given by the following function where P stands for the belief, T for the final period, and k for the number of periods left:

$$P_{T-k} \geq \left(\frac{N_1 - E(S_1)}{E(C_1) - E(S_1)} \right)^{k+1}. \tag{4.1}$$

[2]See Section 1.2.4.

[3]See Kreps and Wilson (1982b) for the general concept of sequential equilibria.

As long as the trustor's prior belief[4] is above the threshold, the trustor will choose *Co-operation*.

As evident in (4.1), the threshold value increases when the number of periods remaining decreases. At first, the trustee will play a pure strategy by choosing *Honour*. The trustor is not able to update his belief. Before the threshold value rises above the trustor's prior belief, the selfish trustee will start to randomise, where the probability of choosing *Honour* is given by $\tilde{\lambda}_{T-k}$:

$$\tilde{\lambda}_{T-k} = \frac{P_{T-k}\Big(E(C_1) - E(S_1)\Big)^k - P_{T-k}\Big(N_1 - E(S_1)\Big)^k}{(1 - P_{T-k})\Big(N_1 - E(S_1)\Big)^k}. \tag{4.2}$$

The trustor will in response continue to choose *Co-operation* with probability $\tilde{\phi}$ in order to make the trustee indifferent:

$$\tilde{\phi} = \frac{A_2 - C_2}{A_2 - N_2}. \tag{4.3}$$

This mixing probability is optimal in every period where (4.1) is met with equality.

As soon as the trustor observes *Abuse* he will choose *No co-operation* for the remainder of the game. This is true if *Abuse* is the result of the mixing strategy of the trustee or if the trustee chooses *Abuse* out-of-equilibrium. Co-operative behaviour can also come to a halt when the trustor chooses *No co-operation* as a result of randomisation. If this happens, the trustor can no longer update his belief. In the following rounds the belief will be below the threshold. Consequently, the trustor prefers to choose *No co-operation* until the end of the game.

When the trustor's prior belief is greater than the threshold in the final round, as given by (4.1), the trustor will choose *Co-operation* in every round. The selfish trustee's best response is to choose *Honour* in every round with the exception of the last round, where he will choose *Abuse*. If the trustor's prior belief is lesser than the threshold in the first round, the trustor will choose *No co-operation* in all periods.

When the trustor only observes *Honour* during the entire game, his updated belief at the end of the final round will equal 1.

[4]This is the belief at the beginning of the game.

4.3 Finitely repeated *asymmetric* trust game

In this section, it will be shown that rational co-operation is also possible in the finitely repeated version of the *asymmetric* trust game.[5] In the game of asymmetric information, learning is less efficient, because when a trustee chooses *Abuse* this does not immediately reveal him as being untrustworthy.

This section starts with an analysis of the two-stage game. To solve the sequential equilibrium we begin with the final period $T = 2$. In this round the trustworthy trustee will always choose *Honour* and the selfish trustee will always choose *Abuse*. The trustor will only choose *Co-operation* when his expected pay-off of this action outweighs his pay-off from choosing *No co-operation*:

$$P_T E(C_1) + (1 - P_T)E(S_1) > N_1. \tag{4.4}$$

The trustor chooses *Co-operation* if:

$$P_T \geq \tilde{P}_T \equiv \frac{N_1 - E(S_1)}{E(C_1) - E(S_1)}. \tag{4.5}$$

Where P_T denotes the probability with which the trustor believes that he is playing with an honest trustee in T, given what he observed in the first period.

For the moment, assume that in the final round the trustor's threshold is larger than his prior belief, or $P_T < \tilde{P}_T$, where $\tilde{P}_T$ stands for the trustor's threshold in the final round, as presented in condition (4.5).

In the first round, $T - 1$, the trustor will either observe an H or an L. The trustor can use this information to update, according to Bayes's Rule, his prior belief P_{T-1} (that he is dealing with an honest trustee (HT)) into his posterior belief P_T. When both honest and selfish trustees choose *Honour*, the trustor cannot update his belief.

If $P_T < \tilde{P}_T$, the trustee might find it beneficial to randomise over his actions so that the trustor will have the chance to update his belief and possibly choose *Co-operation* in the last round. We will use λ_{T-1} to refer to the probability that the selfish trustee chooses *Honour* in the first round. Bayes's Rule, if the trustor observes an H, is then given by:

$$P_T(HT|H) = \frac{P_{T-1}(1 - p_h)}{P_{T-1}(1 - p_h) + (1 - P_{T-1})\Big(\lambda_{T-1}(1 - p_h) + (1 - \lambda_{T-1})(1 - p_a)\Big)}. \tag{4.6}$$

[5] As introduced in Section 1.4.

In the case of an L, the trustor will update his prior belief as follows:

$$P_T(HT|L) = \frac{P_{T-1}p_h}{P_{T-1}p_h + (1-P_{T-1})\Big(\lambda_{T-1}p_h + (1-\lambda_{T-1})p_a\Big)}. \tag{4.7}$$

Given the fact that by definition $p_a > p_h$ and thus $(1-p_h) > (1-p_a)$, it should be clear that if the trustor observes an H, it must be that $P_T > P_{T-1}$ and if he observes an L, it must hold that $P_T < P_{T-1}$. The selfish trustee only needs to randomise in $T-1$ if he knows that $P_T < \tilde{P}_T$. The trustor should have the chance to update his belief such that P_T at least meets this threshold. If the trustor observes an L, it will be the case that $P_T < P_{T-1}$, which will induce the trustor to choose *No co-operation* in the final round. Therefore, the trustee should determine his equilibrium mixing probability λ^*_{T-1} for the situation in which the trustor observes an H. The trustor will be indifferent between choosing *Co-operation* and *No co-operation* in period T when:

$$\frac{N_1 - E(S_1)}{E(C_1) - E(S_1)} =$$
$$\frac{P_{T-1}(1-p_h)}{P_{T-1}(1-p_h) + (1-P_{T-1})\Big(\lambda_{T-1}(1-p_h) + (1-\lambda_{T-1})(1-p_a)\Big)}. \tag{4.8}$$

This leads to the following mixing probability for the trustee in period $T-1$:

$$\lambda^*_{T-1} =$$
$$\frac{\Big(E(C_1) - E(S_1)\Big)P_{T-1}(1-p_h) - \Big(N_1 - E(S_1)\Big)\Big(1 - p_a + P_{T-1}(p_a - p_h)\Big)}{(1-P_{T-1})\Big(N_1 - E(S_1)\Big)(p_a - p_h)}. \tag{4.9}$$

Bower et al. (1996) show by contradiction that λ^*_{T-1} is the equilibrium value of λ_{T-1}. We will show that a similar argumentation can be used to prove that in equilibrium λ_{T-1} must be λ^*_{T-1}. If $\lambda_{T-1} > \lambda^*_{T-1}$, then:

$$P_T(HT|H) < \frac{N_1 - E(S_1)}{E(C_1) - E(S_1)}. \tag{4.10}$$

This means that the trustor will not choose *Co-operation*, even if he observes an H. The selfish trustee's expected pay-off over both rounds equals $\lambda_{T-1}C_2 + (1-\lambda_{T-1})A_2 + N_2$. He now strictly prefers to choose *Abuse* in period $T-1$ and earn an expected pay-off of $A_2 + (1-p_a)A_2 + p_a N_2$. In other words, $\lambda_{T-1} = 0$, which is a contradiction. If $\lambda_{T-1} < \lambda^*_{T-1}$, then:

$$P_T(HT|H) > \frac{N_1 - E(S_1)}{E(C_1) - E(S_1)}. \tag{4.11}$$

The trustor will always choose *Co-operation* in both periods. The selfish trustee strictly prefers choosing *Honour* in the first round if:

$$C_2 + (1-p_h)A_2 + p_h N_2 > A_2 + (1-p_a)A_2 + p_a N_2. \tag{4.12}$$

In this case, $\lambda_{T-1} = 1$ – again a contradiction. When inequality (4.12) is not satisfied, the trustee will prefer to choose *Abuse*. The trustor will update his belief, given the pay-off he earns.

λ^*_{T-1} is the equilibrium mixing probability for parameters that satisfy inequality (4.12). This is different compared to the infinitely repeated *symmetric* trust game where $\tilde{\lambda}_{T-1}$ is the equilibrium mixing probability for all parameters that satisfy (1.1). The reason for this difference lies in the fact that in case of the *symmetric* trust game the trustee knows for certain what pay-off he will earn in the final period when the trustor prefers to play *Co-operation* as a pure strategy. In the case of the *asymmetric* trust game, the trustee is uncertain about the pay-off he will earn in the final period, because nature's move determines whether the trustor earns an H or an L and, consequently, how the trustor will update his belief. If the trustor observes an H, he will prefer to choose *Co-operation*, while if he observes an L, he will choose *No co-operation*. Hence, the trustee is more demanding on the parameters when it comes to the conditions for which a mixing equilibrium can exist.

When the trustee randomises between choosing *Honour* and *Abuse* using the mixing probability from (4.9), $P_T = \tilde{P}_T$ if the trustor observes an H. In other words, the trustor will be indifferent between choosing *Co-operation* and *No co-operation* in round T. However, the trustor needs to make the trustee indifferent as well in his choice between *Honour* and *Abuse* in period $T-1$. The trustor will achieve this by choosing *Co-operation* with a probability ϕ. He will choose a value for ϕ that will make the trustee indifferent in the following expression:

$$\Big(C_2 + (1-p_h)\Big(\phi A_2 + (1-\phi)N_2\Big) + p_h N_2\Big) =$$

$$\Big(A_2 + (1-p_a)\Big(\phi A_2 + (1-\phi)N_2\Big) + p_a N_2\Big). \tag{4.13}$$

The trustee is indifferent if the trustor randomises with probability:

$$\phi^* = \frac{A_2 - C_2}{(A_2 - N_2)(p_a - p_h)}. \tag{4.14}$$

To make sure that ϕ^* lies in the interval [0,1], condition (1.4) must be satisfied and the following inequality should hold:

$$p_a - p_h > \frac{A_2 - C_2}{A_2 - N_2}. \tag{4.15}$$

Inequality (4.15) can be rearranged such that it is equal to inequality (4.12). Therefore, for both the trustor and trustee, the same condition must be satisfied in order for randomisation to be part of the equilibrium strategy. If inequality (4.15) is not satisfied, the trustor will choose *Co-operation* in period T.

In $T-1$ the trustor will only choose *Co-operation* if his expected pay-off from this strategy outweighs his expected pay-off from choosing *No co-operation*, where p_{T-1} is the belief of the trustor:

$$P_{T-1}E(C_1) + (1 - P_{T-1})\Big(\lambda_{T-1}E(C_1) + (1 - \lambda_{T-1})E(S_1)\Big) \geq N_1. \tag{4.16}$$

Solving for λ_{T-1} and substitution of (4.9) yields:

$$\frac{\Big(E(C_1) - E(S_1)\Big)P_{T-1}(1 - p_h) - \Big(N_1 - E(S_1)\Big)\Big(1 - p_a + P_{T-1}(p_a - p_h)\Big)}{(1 - P_{T-1})\Big(N_1 - E(S_1)\Big)(p_a - p_h)}$$

$$\geq \frac{\Big(N_1 - E(S_1)\Big) - P_{T-1}\Big(E(C_1) - E(S_1)\Big)}{(1 - P_{T-1})\Big(E(C_1) - E(S_1)\Big)}. \tag{4.17}$$

Solving for P_{T-1} leads to:

$$P_{T-1} \geq \tilde{P}_{T-1}$$

$$\equiv \frac{\Big(N_1 - E(S_1)\Big)\Big(E(C_1)(1 - p_a) - E(S_1)(1 - p_h) + N_1(p_a - p_h)\Big)}{\Big(E(C_1) - E(S_1)\Big)^2(1 - p_h)}. \tag{4.18}$$

A brief summary of the exposition so far: If $P_{T-1} > \tilde{P}_{T-1}$ and $P_{T-1} < \tilde{P}_T$, the trustor will choose *Co-operation* in the first period. The trustee will randomise in the first round under the condition that inequality (4.12) is satisfied; if not, he will choose *Abuse*. When the trustor observes an H, he will update his belief and $P_{T-1} = \tilde{P}_T$. The trustor will play a mixing strategy in the second period, when inequality (4.15) holds. The trustor will choose *No co-operation* either due to his own mixing or when he has observed an L, i.e. $P_{T-1} < \tilde{P}_T$. If the trustor's mixing leads to *Co-operation*, the selfish trustee will choose *Abuse* in the final period.

When $P_{T-1} = \tilde{P}_{T-1}$, the trustor is indifferent, and we assume that he will randomise between *Co-operation* and *No co-operation* in the first period. When the trustor chooses *Co-operation*, the mixed strategy equilibrium as described above will be played by both the trustor and trustee.

Next to the equilibrium strategy described above, other equilibrium strategies can exist, depending on P_{T-1}. For example, if $P_{T-1} < \tilde{P}_{T-1}$, the trustor will choose *No co-operation* in both periods of the game.

If $P_{T-1} = \tilde{P}_T$, the trustor will choose *Co-operation* in the first period and be indifferent in the final period. The trustee will prefer to choose *Honour* in the first period when the following two conditions are satisfied, where ξ indicates the trustor's randomisation probability:

$$C_2 + \xi A_2 + (1-\xi)N_2 > A_2 + p_a N_2 + (1-p_a)A_2,$$

or:

$$\xi > \frac{A_2(2-p_a) - C_2 + N_2(p_a - 1)}{A_2 - N_2}. \tag{4.19}$$

To make sure that $\xi \leq 1$, it must be that:

$$p_a \geq \frac{A_2 - C_2}{A_2 - N_2}. \tag{4.20}$$

Now assume that $P_{T-1} > \tilde{P}_T$. When the selfish trustee mimics the behaviour of the honest trustee in $T-1$, the trustor cannot update his belief and will choose *Co-operation* in both periods.

Nevertheless, in contrast to the finitely repeated *symmetric* trust game, the trustor cannot observe *Abuse*, which gives the trustee the opportunity to gain an additional *Abuse* pay-off, depending on P_{T-1}. When the trustor's prior belief is exceptionally high, it can be the case that even after observing an L it might be that $P_T > \tilde{P}_T$. If this occurs, the trustee will choose *Abuse* in both rounds ($\lambda_{T-1} = 0$). To be more precise, this will be the case if:

$$\frac{P_{T-1}p_h}{P_{T-1}p_h + (1-P_{T-1})p_a} > \frac{N_1 - E(S_1)}{E(C_1) - E(S_1)}, \tag{4.21}$$

or

$$P_{T-1} > \frac{p_a(N_1 - E(S_1)}{p_h\Big(E(C_1) - N_1\Big) + p_a\Big(N_1 - E(S_1)\Big)}. \tag{4.22}$$

Condition (4.22) is a stricter condition than (4.5), given our assumption that $p_a > p_h$.

If condition (4.22) is not satisfied, the question should be raised whether the trustee might be better off when he randomises over his strategies compared to playing a pure strategy. Assume the trustee can randomise over his strategies in such a way that when the trustor observes an L, the result will be that $P_T > \tilde{P}_T$. In this case, the trustee can earn $\lambda C_2 + (1-\lambda)A_2 + A_2$. If he chooses *Honour* in $T-1$, the trustee will earn $C_2 + A_2$, which is clearly less.

Appendix B shows that under certain, limited conditions such an equilibrium can exist. Given the parameter values used in Chapter 5, only an equilibrium in pure strategies exists if the situation arises that $P_{T-1} > \tilde{P}_T$ and condition (4.22) is not satisfied.

We can now generalise the result from the two-stage game to a T-stage game. The trustee's mixing probability in stage $T-k$ is given by the following equation, where k again stands for the number of rounds left until the end of the game:

$$\lambda^*_{T-k} = \frac{P_{T-k}\Big((1-p_h)\Big(E(C_1)-E(S_1)\Big)\Big)^k - \Big(N_1-E(S_1)\Big)\Big(1-p_a+P_{T-k}(p_a-p_h)\Big)\sigma^{k-1}}{(1-P_{T-k})\Big(N_1-E(S_1)\Big)(p_a-p_h)\sigma^{k-1}}, \tag{4.23}$$

with

$$\sigma = \Big(E(C_1)(1-p_a) - E(S_1)(1-p_h) + N_1(p_a-p_h)\Big).$$

The trustor's threshold under the condition that inequality (4.12) is satisfied is:

$$P_{T-k} \geq \frac{\Big(N_1-E(S_1)\Big)\Big(E(C_1)(1-p_a)-E(S_1)(1-p_h)+N_1(p_a-p_h)\Big)^k}{\Big(E(C_1)-E(S_1)\Big)\Big(\Big(E(C_1)-E(S_1)\Big)(1-p_h)\Big)^k}. \tag{4.24}$$

This threshold increases over time – the nearer the end, the larger the threshold. If mixing is not part of the equilibrium strategy, the trustor's threshold, given by condition (4.5), remains constant over time.

When the trustor is made indifferent, he will use the mixing probability from (4.14), which is constant over time.

In order to establish the sequential equilibrium in the T-stage game, we need to extend the trustor's strategy. Depending on the prior belief P_{T-k} and the threshold path based on (4.24), the trustee should start mixing in a certain round, say in $T-n$. The trustor must force the trustee to choose *Honour* before the mixing phase begins by formulating a punishment strategy for the first phase of the game. Unlike the finitely repeated *symmetric* trust game, the trustor cannot observe *Abuse* in the finitely repeated *asymmetric* trust game. If the trustor cannot punish a trustee who abuses his trust, the trustee has no reason to refrain from doing so. The trustor can force the trustee to choose *Honour* by formulating the following strategy: As soon as the trustor observes more Ls than would be expected if the trustee chose *Honour*, the trustor will choose *No co-operation* for the remainder of the game. The trustor will continue choosing *Co-operation* as long as:

$$\Big(T-(n+1)\Big)p_h \geq \sum_{t=1}^{t=T-(n+1)} L_t. \tag{4.25}$$

Where L_t equals 1 when an L is observed in period t and zero otherwise. Because the trustee only knows with which probability a trustor earns an H or an L, and he does not know the actual realisation (see Section 1.4), choosing *Honour* becomes a dominant strategy for the selfish trustees in the first $T-(n+1)$ periods of the finitely repeated *asymmetric* trust game.

In the finitely repeated *asymmetric* trust game, the trustor's belief will not become 1 at the end of the game when the trustor has observed an H in all rounds, because observing an H can also be the result of *Abuse* on the side of the trustee.

4.4 Conclusion

This chapter presented a model for the finitely repeated *asymmetric* trust game.[6] In this model, the trustor learns about the type of trustee he is facing, given the information he receives. The ambiguous character of the information slows down the learning process, in the sense that the trustor updates his belief less effectively.

Given the fact that the trustor has less information he can use to learn about the trustee's type, he wants a higher assurance that he is facing an honest trustee. In reaction, the probability with which the trustee chooses

[6]The models presented in this chapter ignore risk preferences and/or altruistic behaviour on the side of trustors. The main insights of the models remain unaffected when these differences in preferences are included. Camerer and Weigelt (1988) show that when trustors have risk tastes or are altruistic this can lead to shifts in the trustor's threshold for choosing *Co-operation*: a higher threshold for risk-averse and/or envious trustors and a lower threshold for risk-seeking and/or altruistic trustors.

Abuse must escalate in order for the trustor's belief to be equal to the higher threshold. Because the learning process is hampered by information asymmetry, it is highly likely that the sanction of playing *No co-operation* will be triggered sooner by the randomisation process.

Another way in which learning affects the moment the sanction becomes operational becomes clear when comparing the models for the finitely repeated *symmetric* trust game with those from the finitely repeated *asymmetric* trust game. This comparison makes clear that the trustor's threshold increases earlier in the presence of asymmetric information.[7] Put another way, the trustee needs to randomise over his actions sooner, given a fixed prior belief.

The pure strategy phase of the sequential equilibrium also favours co-operative behaviour less in the presence of asymmetric information. When, due to chance, the trustor observes more Ls than expected, he will stop choosing *Co-operation* altogether.

On a more positive note, the trustor's mixing probability favours *Co-operation* more in the presence of asymmetric information compared to symmetric information.

Overall, given that learning is less effective in the *asymmetric* trust game, the trustor will very likely sanction the trustee earlier compared to the *symmetric* trust game. Hence, the control mechanism becomes less effective in sustaining co-operation.

[7] It is not easy to directly compare the different conditions for the *symmetric* and *asymmetric* trust game. In Section 5.2.2, a graphical representation of the equilibrium paths for both trust games can be found, which will help to interpret the differences between the models presented in this chapter.

Chapter 5

Testing finitely repeated *asymmetric* trust games

5.1 Introduction

The theoretical models presented in the previous chapter suggest that it should be more difficult for trustors and trustees to co-operate in the finitely repeated *asymmetric* trust game, where the trustor cannot observe the trustee's action, compared to the finitely repeated *symmetric* trust game, where the trustee's choice is observable. This chapter presents the experimental result on whether the checking option can stimulate co-operation in finitely repeated *asymmetric* trust games.

The sequential equilibrium assumes the presence of two types of trustees, each of which behaves in a strategically different manner. Whether or not the trustor chooses *co-operation* depends on his prior belief about the proportion of honest trustees in the population and his observation of the trustee's behaviour. Whether the selfish trustee will choose *Honour* depends on the effectiveness of learning and control. In this chapter, the relevance of the theoretical distinction between trustor and trustee types in finitely repeated trust games will be empirically tested.

The experimental design used in this chapter is very similar to the experimental design of Chapter 3. In Section 5.2.1, the differences between the two experiments will be pointed out. In Section 5.2.2, the hypotheses are formulated, while the data analysis is presented in Section 5.3. The chapter ends with a discussion in Section 5.4.

5.2 Experimental design and hypotheses

5.2.1 The design

For the finitely repeated *asymmetric* trust game experiment, the experimental design of Chapter 3 is used, but with one obvious exception – the isolated encounter *asymmetric* trust game is finitely repeated instead of indefinitely. To be more precise, the game is repeated ten times.[1]

As mentioned in Chapter 1, it was impossible to find one set of parameter values that, on theoretical grounds, would yield a significant difference between the treatments of both experiments. In order to maintain a clear difference between the NCT and the CT in the finitely repeated *asymmetric* trust game experiment, a different set of parameter values is used in this experiment (see Table 5.1).

Table 5.1: Parameter values for the finitely repeated *asymmetric* trust game experiment.

Parameters	
$p_a = \frac{3}{4}$	$p_h = \frac{1}{4}$
$H = 35$	$L = -5$
Trustor's pay-offs	Trustee's pay-offs
$E(C_1) = 25 = \frac{1}{4}(-5) + \frac{3}{4}(35)$	$C_2 = 25$
$E(S_1) = 5 = \frac{3}{4}(-5) + \frac{1}{4}(35)$	$A_2 = 38$
$N_1 = 20$	$N_2 = 10$

When the trustor earns -5 or 35 in the *asymmetric* trust game, it is not clear to him whether the trustee has responded with *Abuse* or *Honour* that period, as is clearly depicted in Figure 5.1.

In the TT, the trustor's pay-offs equal his expected values from the NCT and CT. Therefore, when the trustor chooses *Co-operation* and the trustee responds with *Honour*, the trustor earns 25, whereas he earns 5 if the trustee responds with *Abuse*.

Although the TT provides us with the proportion of honest trustees in the population, this information was not made available to the subjects. In other words, the subjects played the game without a common prior. If such a prior had been introduced, this might have compromised the comparison

[1] Recall that the expected duration of the indefinitely repeated game is also ten periods. The reason behind this is that a comparison will be made in Chapter 6.

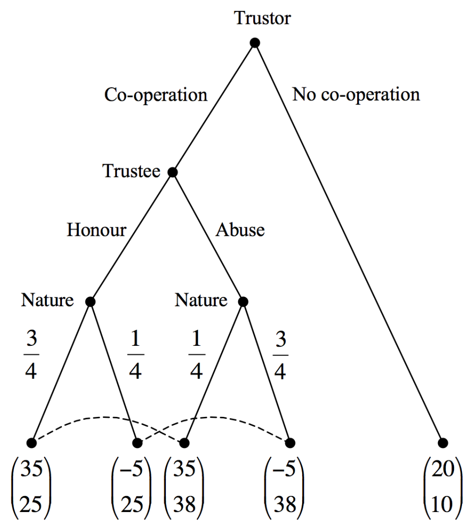


Figure 5.1: *Asymmetric* trust game with pay-offs for the finitely repeated *asymmetric* trust game experiment.

between the NCT and the CT. Given the fact that we only had nine sessions, the variation in the priors would have been limited and might have driven the results. Now the trustor's private prior is an unobserved personal characteristic. Given that the experimental data is a panel, the unobserved personal prior is taken into account by the random effects estimator.

The instructions for all treatments can be found in Appendices D.1, D.2 and D.3. The questionnaire questions were not altered for this experiment, so for these we refer again to Appendix E.

5.2.2 Hypotheses

Given the pay-offs, probabilities and the number of repetitions used in the experiment, it is possible to calculate the trustor's mixing probabilities, which are the same for all periods where (4.1) and (4.24) are equalities. Substituting the parameter values in conditions (4.3) and (4.14) yields $\tilde{\phi} = 0.46$ and $\phi^* = 0.93$, respectively. The pay-offs and probabilities used in the experiment are chosen such that inequalities (4.12) = (4.15) are satisfied.

In Figure 5.2, the equilibrium paths for P_{T-k} are depicted for both the CT and the NCT. It is evident that the trustor's threshold from the NCT is above the threshold from the CT at the beginning of the first nine rounds.

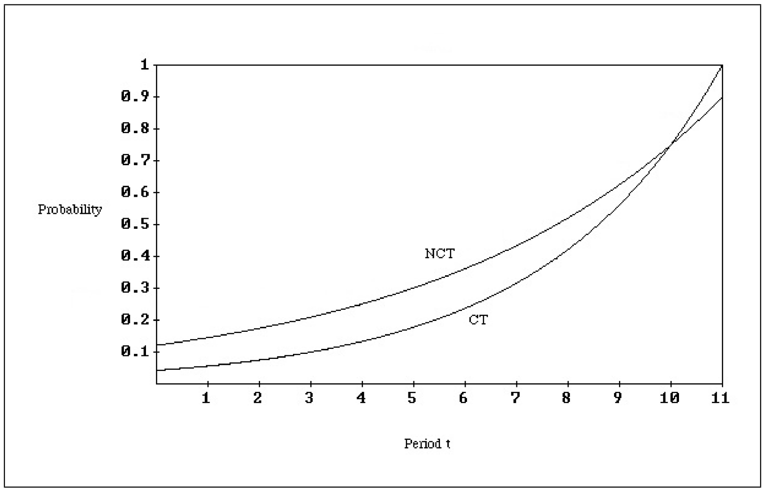


Figure 5.2: Trustors' thresholds for the NCT and the CT. The upper curve is the threshold for the NCT the lower curve for the CT.

As mentioned before, the trustor's belief will become equal to 1 at the end of the final round in the *symmetric* trust game when the trustor observes *Honour* throughout the game. This is not true in the *asymmetric* trust game, given the parameter values used in this experiment it only becomes 0.9.

Depending on the prior belief, the difference between the period in which mixing begins can vary. The trustee's mixing probabilities also depend on the prior belief. To get an idea how things might look, we assume that the trustor can meet an honest trustee with a prior probability of 0.25, which is graphically displayed in Figure 5.3. In order to make sure that the trustor's threshold is met in the next round when he observes an H, the selfish trustee should start mixing in round 3 in the case of the NCT and in round 6 in the case of the CT for a prior of 0.25. Comparing the NCT with the CT, it can be seen that the probability with which the trustee chooses *Honour* is the same in the first two periods. Starting from the third period, it is, nonetheless, always smaller in the case of the NCT. In other words, $\lambda^*_{T-k} \leq \tilde{\lambda}_{T-k}$ throughout the repeated game. As can be seen in the NCT, the mixing probability decreases very quickly. Given our example prior of 0.25, the mixing probability drops so fast that selfish trustees should choose *Abuse* in the final four rounds of the NCT compared to only the final round in case of the CT.

Due to the additional uncertainty generated by information asymmetry,

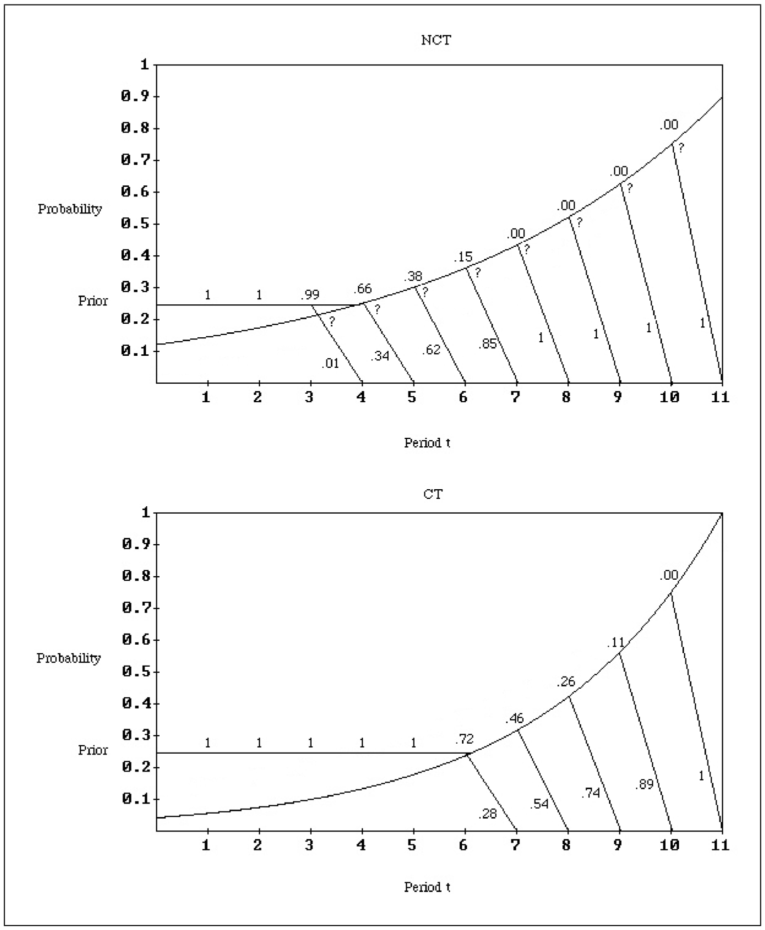


Figure 5.3: An example of when the trustor has a prior probability of 0.25 to meet an honest trustee. The probabilities with which the trustee should choose *Honour* in the CT can be found in the lower graph, and are placed along the trustor's threshold. The vertical lines reflect the situation in which the trustee chooses *Abuse*, where the trustor immediately updates his belief to 0. In the upper graph, the trustee's mixing probabilities for the NCT can be found. The question marks reflect the fact that it is not the trustee's strategy but the trustor's observation that determines how the trustor will update his belief. The latter is, of course, determined by the trustee's strategy. Choosing *Honour* makes it more likely that the trustor will observe an H, but this is not certain. The trustee can also be lucky when due to mixing he chooses *Abuse*, and the trustor can still earn an H and the trustee will have the opportunity to randomise over his strategies again in the next round. When the trustor earns an L, he will stop choosing *Co-operation* in the mixing phase of the equilibrium.

the trustor needs more assurance that he is dealing with an honest trustee. In other words, he is more demanding on P_T. The trustee in the NTC will find that it is more difficult to convince the trustor that he is honest. To make sure that the trustor can update his belief sufficiently, he needs to choose *Honour* with a lower probability. In Figure 5.3, it can be seen that in the NCT, the trustor wants to be convinced in an earlier stage of the game that he is facing an honest trustee. The selfish trustee needs to mix sooner to give the trustor the chance to update his belief. In the NCT, the selfish trustee does this by using a mixing probability that favours choosing *Abuse* above choosing *Honour* compared to the mixing probability of the CT. Conversely, the trustor's mixing probability favours choosing *Co-operation* more in the case of the NCT compared to the CT. This does not, however, outweigh the overall negative effect due to the trustee's mixing. In addition, the trustor stops choosing *Co-operation* as soon as he sees an L in the mixing phase of the NCT, which might also be the case when the trustor chooses *Honour*. In the NCT, it will be more likely that the trustor updates his belief to zero, and stops choosing *Co-operation* altogether. We can thus formulate the following hypothesis:

HYPOTHESIS 1: *Trustors choose* Co-operation *more frequently in the CT compared to the NCT.*

Given the fact that trustees are required to mix sooner in the NCT and choose *Honour* with a lower probability, it should be less likely that trustees choose *Honour* in the NCT.

HYPOTHESIS 2: *Trustees choose* Honour *more frequently in the CT compared to the NCT.*

The equilibrium described in Chapter 4 is based on the assumption that there is a (small) probability that the trustor is confronted with an honest trustee. Trustors should have some belief about the proportion of the honest type in the trustee population. Because trustful trustors choose *Co-operation*, even in an isolated encounter trust game, they should have a higher prior belief compared to non-trustful trustors. Note that the threshold for the final stage of the finitely repeated trust game is the same as the threshold of the isolated encounter trust game. Consequently, trustful trustors should choose *Co-operation* more frequently compared to non-trustful trustors. The two trustee types should behave differently in order for the trustor to update his belief. Hence, the following hypotheses:

HYPOTHESIS 3a: *Trustful trustors choose* Co-operation *more frequently in both NCT and CT compared to non-trustful trustors.*

HYPOTHESIS 3b: *Honest trustees choose* Honour *more frequently in both NCT and CT compared to selfish trustees.*

The theoretical models suggest that trustors and selfish trustees are less likely to co-operate over time.

HYPOTHESIS 4a: *Over time, trustors are less likely to choose* Co-operation *and selfish trustees are less likely to choose* Honour *in both the CT and the NCT.*

HYPOTHESIS 4b: *Honest trustees remain honest over time.*

Honest trustees should, however, remain honest during the entire game by definition, while both non-trustful and trustful trustors should become less trustful over time. The reason for the latter is that at the aggregate level it is more likely that most trustors will face a selfish trustee. Each period the trustor will receive more information, which he can use to learn about the trustee's type. Most trustors will find that they face a selfish trustee.

While the first four hypotheses focus on the treatment comparison, the last two will be about the role of checking in the CT. Intuitively, it should be expected that trustful trustors are less eager to check on the behaviour of trustees after they choose *Co-operation*, compared to non-trustful trustors, who will only choose *Co-operation* when they know that choosing *Honour* is the best response for the trustee.

HYPOTHESIS 5: *After choosing* Co-operation*, trustful trustors will check less frequently compared to non-trustful trustors.*

The equilibrium strategy dictates that whenever *Abuse* is observed, either out-of-equilibrium in the pure strategy phase or due to mixing in the mixing phase, it should be responded to by choosing *No co-operation* in the next period and all succeeding periods.

HYPOTHESIS 6a: *Non-trustful trustors who have checked and observed* Abuse *in the CT will lower the frequency with which they choose* Co-operation*.*

HYPOTHESIS 6b: *Trustful trustors who have checked and observed* Abuse *in the CT will lower the frequency with which they choose* Co-operation*, even more compared to non-trustful trustors.*

Trustful trustors have a higher prior belief, which indicates that when they observe *Abuse*, a larger reduction should be visible in the likelihood with which they choose *Co-operation*. After observing *Abuse*, both types of trustors should know that they face a selfish trustee. Nevertheless, trustful trustors have to make a bigger adjustment given their higher prior belief.

5.3 Results

5.3.1 Carrying out the experiment and descriptive statistics

The experiment [2] was held at ELSE and consisted of a total of nine sessions. The first six sessions were conducted in the last week of May and the first week of June 2007. The last three sessions were held in September of the same year.

The finitely repeated *asymmetric* trust game experiment was carried out in the same way as the indefinitely repeated *asymmetric* trust game experiment, presented in Chapter 3.

In total, 164 subjects participated. All subjects were registered with the ELSE subject pool. Table 5.2 provides some basic information about the experiment on the session level.

The number of subjects hardly differs per session. 60.98% of the subjects are women. As can be seen, the number of women differs per session. In five sessions, female subjects are in the majority. 85.96% of the subjects are Dutch, which also includes subjects with double nationalities. The group of foreign subjects can be divided into 18 different nationalities. The largest groups are Italian and German students, representing 2.44% and 1.83% of the subject population, respectively. 14.63% of the subjects indicate that they have studied economics. On average, the subjects earned €12.65 per session. The average session took about one hour and 5 minutes.
For an overview of the summery statistics of all variables, see Table 5.3. A description of the majority of these variables can be found in Section 3.3.2. This chapter introduces three new variables. *Checkyn* is a dummy variable that indicates that the trustor has used his checking option. *Period9* and

[2] z-Tree by Fischbacher (1999) was used for programming the treatments and to run the experiment.

Table 5.2: Some basic information on the session level for the finitely repeated *asymmetric* trust game experiment.

	Session 1	Session 2	Session 3
Date	22-5-07	31-5-07	31-5-07
Session type*	type A	type A	type A
Subjects	14	20	18
Economics students	14.29%	10%	5.56%
Male subjects	50%	30%	11.11%
Trustful trustors†	42.86%	35%	55.56%
Honest trustees†	50%	30%	33.33%
Average earnings	€15.58	€12.07	€12.67
Maximum earnings	€19.43	€13.85	€14.97
Minimum earnings	€12.48	€10.14	€10.29
	Session 4	Session 5	Session 6
Date	4-6-07	4-6-07	5-6-07
Session type*	type B	type B	type B
Subject	18	20	14
Economics students	16.67%	25%	7.14%
Male subjects	33.33%	15%	21.43%
Trustful trustors†	38.89%	50%	57.14%
Honest trustees†	44.44%	50%	42.86%
Average earnings	€12.43	€12.38	€12.58
Maximum earnings	€15.19	€15	€10.71
Minimum earnings	€10.17	€9.26	€15.09
	Session 7	Session 8	Session 9
Date	26-9-07	27-9-07	27-9-07
Session type*	type A	type A	type B
Subject	20	20	20
Economics students	10%	0%	40%
Male subjects	50%	60%	75%
Trustful trustors†	55%	50%	25%
Honest trustees†	35%	60%	45%
Average earnings	€12.30	€12.20	€12.46
Maximum earnings	€15.96	€14.78	€14.58
Minimum earnings	€9.68	€10.45	€9.05

*For the definition of session type, see Table 3.3 †Type as derived from the TT.

Table 5.3: Overview of variable descriptive statistics for the finitely repeated *asymmetric* trust game experiment.

Variable name	Obs	*Mean*	Std. Dev.	Min	Max
Dependent variables					
ChoiceA*	6560	0.571	0.495	0	1
ChoiceB*	3748	0.709	0.454	0	1
Checkyn*	2084	0.775	0.417	0	1
Independent variables					
Checktreat*	6560	0.500	0.500	0	1
Period	6560	5.500	2.873	1	10
Period9*	6560	0.100	0.300	0	1
Period10*	6560	0.100	0.300	0	1
Trustful*	6560	0.451	0.498	0	1
Honest*	6560	0.433	0.496	0	1
Abuse−lag1*	5904	0.052	0.222	0	1
Main control variables					
Checkfirst*	6560	0.439	0.496	0	1
Trustorfirst*	6560	0.500	0.500	0	1
Other control variables					
Female*	6560	0.610	0.488	0	1
Economics*	6560	0.146	0.353	0	1
Dutch*	6560	0.860	0.347	0	1
Age	6560	21.195	2.669	17	34
Friends	6560	0.427	1.380	0	16
Work*	6560	0.598	0.490	0	1
Volwork*	6560	0.402	0.490	0	1
Donor*	6560	0.165	0.371	0	1
Organdonor*	6560	0.549	0.498	0	1
Religious*	6560	0.317	0.465	0	1

Note: Obs = Number of observations; Std. Dev. = Standard deviation; Min = minimum value; Max = maximum value and * indicates dummy variables, 1=yes.

Period10 are dummy variables that indicate the one before last and the final period of a game. The results of the Type Treatment are presented in Table 5.2 and, where used, to create the dummy variables *Trustful* and *Honest*.

5.3.2 Data analysis

To get a feeling for the data, Table 5.4 presents the average rate of *Co-operation* and *Honour* per treatment. As seen, trustors choose *Co-operation* more in the CT compared to the NCT. The same is true for trustees with respect to choosing *Honour*. The CT averages reported in Table 5.4 fall in the range of results found by other studies on finitely repeated trust games. Bohnet and Huck (2004) found an average rate of *Co-operation* of 0.59 and an average rate of *Honour* of 0.61. In the first supergame of their finitely repeated trust game experiment, Engle-Warnick and Slonim (2004) found an average *Co-operation* rate of 0.82 and an average *Honour* rate of 0.81.

Table 5.4: *Co-operation* and *Honour* rates for both NCT & CT in the finitely repeated *asymmetric* trust game experiment.

	Co-operation	*Honour*
NCT	0.505	0.644
CT	0.638	0.760

Figure 5.4 gives a clear picture of the average rates of *Co-operation* and *Honour*, which are higher in the CT compared to the NCT. According to the WMW ranksum test, this difference is significant.[3]

The checking option was used 77.54% of the time by trustors who placed trust in the CT, which is slightly less than in the indefinitely repeated *asymmetric* trust game experiment.

In order to test the hypotheses, we have estimated four panel logit models including random effects on the level of subjects. We also ran regressions including the group 'other control variables', as specified in Table 5.3. These control variables did not alter the main results qualitatively, and are therefore not reported. In five sessions, female subjects were in the majority. We created a dummy variable to test if this might have affected the results. The additional estimations we ran showed that this was not the case.

FINDING 1: *Trustors choose* Co-operation *more frequently in the CT compared to the NCT.*

Trustors, regardless of the treatment they start in, choose *Co-operation* more

[3]Wilcoxon-Mann-Whitney (WMW) ranksum test using period averages: *Co-operation* p=0.0256 and *Honour* p=0.0126.

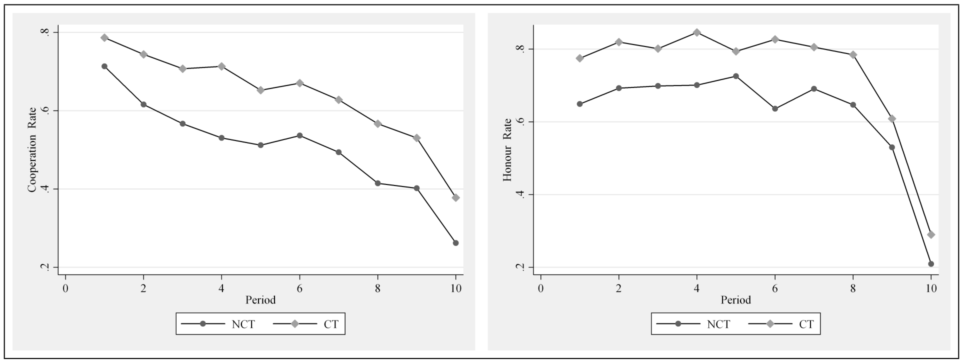


Figure 5.4: *Co-operation* and *Honour* rates over time in the finitely repeated *asymmetric* trust game experiment.

frequently in the CT, with a p-value of less than 0.001, although the coefficient of trustors who start with the CT is lower than those who start with the NCT. There is some evidence at the 10% significance level that trustors who start with the CT are more likely to choose *Co-operation* in the NCT. A clear end-round effect is visible for the last period, as well as a decreasing time trend, both with p-values of less than 0.001. See Table 5.6 for more details.

FINDING 2: *Trustees choose* Honour *more frequently in the CT compared to the NCT.*

As can be seen in Table 5.7, trustees choose *Honour* significantly more often in the CT. The combined coefficient shows that for trustees who start with the CT this effect is less strong. A negative time trend can be observed for both types of trustees, as well as a clear end-round effect for periods 9 and 10.[4]

That trustors and trustees co-operate more in the CT compared to the NCT is also reflected in the average earnings. Both trustors and trustees earn more points in the CT compared to the NCT, 6.6% and 8.3%, respec-

[4]In some of the additional estimations that were run, we found that the interaction between the variables *checktreat* and *female* was significant at the 1% significance level, and had a negative sign for trustor decisions. We found a similar result at the 5% significance level for trustee decisions. So, both female trustors and trustees were less likely to choose *Co-operation* and *Honour*, respectively, in the CT. Adding this interaction term did not influence the results as they are presented in Tables 5.6 and 5.7.

tively. See Table 5.5 for an overview. The WMW-ranksum test shows that the underlying distributions are not the same. Both trustors and trustees earn significantly more in the CT with p=0.0660 and p=0.0065 respectively.

Table 5.5: Subjects earnings in points per treatment.

		Trustor		
	Mean	Std. Err.	Min	Max
NCT	187.68	48.70	65	310
CT	200.06	47.48	80	350
		Trustee		
	Mean	Std. Err.	Min	Max
NCT	199.12	59.16	100	324
CT	215.57	51.72	100	339

FINDING 3a: *Trustful trustors choose* Co-operation *more frequently in both NCT and CT compared to non-trustful trustors.*

FINDING 3b: *Honest trustees choose* Honour *more frequently in both NCT and CT compared to selfish trustees.*

Trustful trustors choose *Co-operation* significantly more often, regardless of the treatment they are in. Honest trustees are also significantly more likely to choose *Honour* in the NCT, with a p-value of less than 0.001. Compared to selfish trustees, honest trustees choose *Honour* more often in the CT; however, the positive effect of being honest on the frequency with which *Honour* is chosen is smaller in this treatment relative to the NCT. A Chi^2-test on the combined effect of *Honest* and $Honest \times Checktreat$ confirms this. *Honest* has a positive effect on the frequency with which *Honour* is chosen in the CT, but only with a p-value of 0.062.

The result that in finitely repeated trust games trustful trustors are more likely to choose *Co-operation* and honest trustees are more likely to choose *Honour* supports the finding of Bohnet and Huck (2004).

FINDING 4a: *Both types of trustors and selfish trustees are less likely to respectively choose* Co-operation *and* Honour *over time in both treatments.*

Table 5.6: Random-effects logit model for the probability of choosing *Co-operation* in the finitely repeated *asymmetric* trust game experiment.

Dependent variable: ChoiceA				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Checktreat	1.411	0.229	6.15	0.000
Period	−0.141	0.027	−5.24	0.000
Checktreat×Period	−0.007	0.031	−0.22	0.824
Period10	−0.684	0.171	−4.00	0.000
Trustful	1.720	0.329	5.23	0.000
Trustful×Period	−0.120	0.033	−3.65	0.000
Trustful×Checktreat	−0.285	0.182	−1.57	0.117
Checkfirst	0.478	0.272	1.76	0.079
Checkfirst×Checktreat	−0.983	0.180	−5.46	0.000
Trustorfirst	−0.019	0.255	−0.08	0.940
Constant	0.166	0.290	0.57	0.566
Number of observations: 3280, Number of subjects: 164				
LL: -1755.692, σ_u: 1.513, ρ : 0.410⋆				

⋆ LL: Log Likelihood, σ_u: Standard deviation of the error term at the subject level, ρ: proportion of the variance in the composite error term that is at the subject level, or $\rho = \frac{\sigma_u^2}{\sigma_u^2+\sigma_e^2}$.

Table 5.7: Random-effects logit model for the probability of choosing *Honour* in the finitely repeated *asymmetric* trust game experiment.

Dependent variable: ChoiceB				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Checktreat	1.575	0.327	4.81	0.000
Period	−0.138	0.046	−3.04	0.002
Checktreat×Period	0.022	0.049	0.44	0.658
Period9	−1.244	0.266	−4.67	0.000
Period10	−3.078	0.365	−8.44	0.000
Honest	1.760	0.447	3.94	0.000
Honest×Period	0.081	0.051	1.57	0.116
Honest×Checktreat	−0.911	0.296	−3.07	0.002
Checkfirst	0.123	0.359	0.34	0.731
Checkfirst×Checktreat	−0.672	0.289	−2.33	0.020
Trustorfirst	0.252	0.323	0.78	0.434
Constant	0.294	0.376	0.78	0.435
Number of observations: 1874, Number of subjects: 163				
LL: -859.748, σ_u: 1.796, ρ : 0.495				

FINDING 4b: *Honest trustees remain honest over time.*

As seen in Table 5.6, non-trustful trustors choose *Co-operation* significantly less often every next round of the game, with a p-value of less than 0.001. Although trustful trustors initially choose *Co-operation* more frequently, the frequency with which they do so also decreases faster over time. *Trustful*× *Period* is significant with a p-value lower than 0.001 and has a negative sign. These results support the theoretical assumption that trustors learn over time about the type of trustee they are facing. Given the fact that, on average, the majority of the trustees is selfish, trustors become less trustful over time.

Although trustors might have different prior beliefs about the trustee population, when they get the impression that they face a selfish trustee their belief should be updated in the correct manner. This means that the likelihood with which trustful trustors choose *Honour* should drop more sharply when they obtain more information over time compared to non-trustful trustors. The data support this theoretical prediction. This finding points in the same direction as Finding 6b, which states that trustful trustors react in a stronger way when observing *Abuse* than non-trustful trustors.

Selfish trustees choose *Honour* significantly less often over time (see Table 5.7). A Chi^2-test on the combined effect of *Period* and *Honest*×*Period* shows that there is no significant effect for honest trustees. It seems that the theoretical assumption also holds in the laboratory, because honest trustees remain honest over time.

The next two findings relate to trustor behaviour in the CT.

FINDING 5: *When* Co-operation *has been chosen, trustful trustors in type A sessions use the checking option less frequently than non-trustful trustors.*

At the 1% significance level, we found evidence that trustful trustors who start with the NCT are less likely to check on the trustee compared to non-trustful trustors. A Chi^2- test on the combined effect of *Trustful* and *Trustful*×*Checkfirst* showed that there is no significant effect of trustful trustors on the decision to use the checking option when they start with the CT. Next, it seems that both types of trustors who start the first game they play as trustor are less likely to use the checking option compared to the trustors who start their first game as a trustee. See Table 5.8 for the details.

FINDING 6a: *Non-trustful trustors who have checked and observed* Abuse *in the CT lower the frequency with which they choose* Co-operation*.*

Table 5.8: Random-effects logit model for the probability of checking in the CT in the finitely repeated *asymmetric* trust game experiment.

Dependent variable: Checkyn				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Trustful	−2.178	0.838	−2.60	0.009
Period	−0.088	0.060	−1.48	0.139
Trustful×Period	0.062	0.079	0.79	0.432
Checkfirst	−0.511	0.791	−0.65	0.519
Trustful×Checkfirst	2.111	1.135	1.86	0.063
Trustorfirst	−1.173	0.559	−2.10	0.036
Constant	4.490	0.720	6.24	0.000
Number of observations: 1042, Number of subjects: 156				
LL: -379.538, σ_u: 2.805, ρ : 0.705				

Table 5.9: Random-effects logit model for the probability of choosing *Co-operation* in the CT after observing *Abuse* in the previous period in the finitely repeated *asymmetric* trust game experiment.

Dependent variable: ChoiceA				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Abuse−lag1	−2.582	0.365	−7.07	0.000
Trustful	0.834	0.718	1.16	0.245
Abuse−lag1×Trustful	−2.032	0.698	−2.91	0.004
Period	−0.198	0.063	−3.13	0.002
Period10	−1.068	0.387	−2.76	0.006
Trustful×Period	−0.078	0.088	−0.89	0.375
Checkfirst	−0.379	0.336	−1.13	0.260
Trustorfirst	−0.356	0.330	−1.08	0.280
Abuseperiod1	−0.959	0.473	−2.03	0.043
Abuseperiod1×Trustful	0.951	0.707	1.35	0.178
Constant	3.719	0.550	6.76	0.000
Number of observations: 980, Number of subjects: 155				
LL: -372.094, σ_u: 1.487, ρ : 0.402				

FINDING 6b: *Trustful trustors who have checked and observed* Abuse *in the CT do lower the frequency with which they choose* Co-operation *even more compared to non-trustful trustors.*

As evidenced in Table (5.9), selfish trustors in the CT who have used the checking option are significantly less likely to choose *Co-operation* when they discover that the trustee has chosen *Abuse* in the previous period.[5] Trustful trustors who have observed that their trust has been abused in the previous period are even less likely to choose *Co-operation* compared to non-trustful trustors, with a p-value of 0.004. This result differs from the result found in Chapter 3, where both types of trustors react in the same way to observing *Abuse*, just like theory predicted.

Given the results of Chapter 3, where we found that trustors reacted differently to observing an H or an L, we decided to test if a similar reaction could be found in the finitely repeated trust game experiment. We found that the pay-off realisation of the previous period has an influence on the trustor's decision to choose *Co-operation* in the next period.

When the trustor's trust is being abused, Table 5.9 shows that trustors choose *Co-operation* less frequently afterwards. However, differentiating between the two realisable pay-offs, we found that for both types of trustors there is a small chance that they will choose *Co-operation*, as long as they have earned an H.

If the trustor's trust is honoured in the previous period, trustors are more likely to choose *Co-operation*, with a p-value of less than 0.001. When they have earned an H in the last period, it is even more likely that they will choose *Co-operation*.

5.4 Discussion and conclusion

The main empirical findings confirm the theoretical predictions. The reduction of information asymmetry by checking has a positive effect on both the frequency with which trustors choose *Co-operation* and the frequency with which trustees choose *Honour*. Checking makes learning and control more effective, allowing trustors and trustees to reach a higher level of income.

Trustors used the checking option often, although trustful trustors in the type A sessions where the NCT was played before the CT, seemed to

[5]For the same reason mentioned in Chapter 3, the variable *Abuseperiod1* and its interaction with *Trustful* are included in the regression model.

feel there was less need to check out the behaviour of the trustee they were facing.

Furthermore, the type indicators inferred from the TT pointed in the expected direction. Trustful trustors chose *Co-operation* more frequently compared to non-trustful trustors, and honest trustees were more likely to choose *Honour* compared to selfish trustees. More important is the finding that honest trustees do not change their behaviour over time. The key assumption on which the learning is based is reflected in the experimental data.

That learning plays a role in finitely repeated games is also supported by the finding that both non-trustful and trustful trustors are less likely to choose *Co-operation* in later rounds of the game. Even differences in the way trustors should update their beliefs are reflected in the data by the finding that trustful trustors react stronger when they observe *Abuse* compared to non-trustful trustors.

Less obvious is the result that honest trustees choose *Honour* less frequently in the CT compared to the NCT. One possible explanation could be the crowding out effect. Crowding out, first mentioned by Titmuss (1970) and mainly developed by cognitive social psychologists (see, for example, Deci and Ryan (1985) and Lepper and Greene (1978)), is mostly known in economics through the work of Bruno Frey, for example Frey (1997a). Frey (1997b) makes a distinction between two psychological processes that can account for crowding out:

1. "When individuals perceive the external intervention to be 'controlling' in the sense of reducing the extent to which they can determine actions themselves, intrinsic motivation is substituted by extrinsic control.

2. An intervention from the outside undermines the actor's intrinsic motivation if it carries the notion that the actor's intrinsic motivation is not acknowledged".

It has been argued that trustworthiness itself can also be crowded out, exemplified by Fehr and Gächter (1998), Bohnet et al. (2001) and Falk and Kosfeld (2006).

Checking can be described as a weak form of 'control', and unlike monitoring and other stronger forms of 'control' that can crowd out trustworthiness by replacing it with external force, it could be argued that checking can crowd out trustworthiness by not acknowledging it; checking could be perceived as a contradicting signal. The trustor indicates that he trusts the trustee by choosing *Co-operation*, but later he signals distrust by checking the trustee. The honest trustee will respond with choosing *Honour*. When the trustor checks, the honest trustee's kind response, honouring placed trust, is not acknowledged.

We ran a regression to test how honest trustees reacted when faced with the decision to choose *Honour* or not, after being checked in the previous round. However, no significant effect was found for checking in the previous period on the behaviour of the honest trustees in next period. The same result was obtained when the effect of several checks was taken into account. Hence, the crowding out argument cannot be supported.

Camerer and Weigelt (1988) found that after many repetitions of the finitely repeated game, actual play roughly corresponded to the sequential equilibrium. In this study, it was not the goal to test the sequential equilibrium. Nevertheless, the results show, in line with the models of Chapter 4, that co-operative play decreases over time in both treatments. This result itself is not unique and can be found in several papers where finitely repeated games are experimentally tested. See, for example, Engle-Warnick and Slonim (2004) for the trust game, Dal Bó (2005) for the prisoner's dilemma and Bornhorst et al. (2004) for the investment game.

As seen in the indefinitely repeated *asymmetric* trust game experiment, pay-off realisations do influence trustor behaviour in the CT. This result can be explained by reinforcement learning, as mentioned in Chapter 3.

Chapter 6

Repeated *asymmetric* trust games: does uncertainty about the end of the game matter?

6.1 Introduction

As mentioned in Chapter 1, two classes of repeated games can be distinguished: games with a certain end and games with an uncertain end. In Chapters 2 and 4, it is has been made clear that from a theoretical perspective there is an important difference between both types of games regarding the mechanisms of temporal embeddedness. In the case of finitely repeated trust games, learning and control can support co-operative behaviour. In these games, learning determines the moment the trustor will use his sanction of choosing *No co-operation*. In the case of indefinitely repeated trust games, control is the dominant mechanism that can enforce co-operation. The moment the sanction will be implemented is determined by the strategy itself. In Chapter 2, this was the (strict) grim-trigger strategy.

With the *symmetric* trust game there is a clear theoretical and empirical difference while comparing co-operative behaviour in finitely and indefinitely repeated games. This chapter explores whether similar differences can be found when a comparison is made between the finitely and indefinitely repeated *asymmetric* trust game.

This chapter has the following outline: In Section 6.2, the difference in the horizon perspective is discussed in more detail. The experimental design used in this chapter is presented in Section 6.3.1, the hypotheses can be found in Section 6.3.2 and the experimental results in Section 6.4. This chapter will conclude with a discussion in Section 6.5.

6.2 Horizon perspectives: a comparison

The difference in the horizon perspective becomes most clear when studying repeated binary (or the strategically equivalent *symmetric*) trust games of complete information. In the finitely repeated trust games, the last interaction has the same characteristics as the isolated encounter game. When backward induction is applied, theory predicts that co-operative behaviour can only be supported in finitely repeated trust games if it is an equilibrium strategy in the isolated encounter trust game. Hence, in the finitely repeated trust game of complete information, players do not co-operate at all, while in indefinitely repeated trust games of complete information, if the grim-trigger strategy is an equilibrium strategy, players co-operate in every round. Therefore, a great difference exists in the possible level of co-operation between games with certain and uncertain endings when they are studied under complete information.

In this thesis, incomplete information about trustee type is explicitly assumed. The introduction of incomplete information does not alter the behaviour of trustors and trustees in the indefinitely repeated trust game, as long as the grim-trigger strategy is an equilibrium strategy. When the finitely repeated trust games is turned into a game of incomplete information, alternative equilibrium strategies become possible. When a sequential equilibrium exists, the trustor and the trustee can co-operate to a certain degree, depending on the prior belief. Given the prior belief, the trustee starts randomising between *Honour* and *Abuse*, one period before the trustor's prior belief will be lower than his threshold. One period after the trustee starts mixing, the trustor will start mixing between *Co-operation* and *No co-operation* in response. If the trustor's prior belief is relatively low, co-operation will very likely stop before the end of the game. When the grim-trigger strategy is an equilibrium strategy in the indefinitely repeated trust game – and a sequential equilibrium exists in the finitely repeated trust game – the difference in co-operative behaviour observed will mainly depend on the level of the trustor's prior. Compared to the situation of complete information, the difference in the horizon perspective becomes smaller in the presence of incomplete information.

Several authors performed experimental studies to analyse the difference between finitely and indefinitely repeated social dilemma games.

Engle-Warnick and Slonim (2004) and Slonim et al. (2006) make a comparison between finitely repeated and indefinitely repeated trust games. Both studies found that, initially, there is no significant difference in player behaviour between finitely and indefinitely repeated trust games. However, when players get more experienced in playing these repeated games, the level of co-operation is higher in indefinitely repeated games. Engle-Warnick and Slonim (2004) also found that trustors play different strategies under the two horizon perspectives.

Dal Bó (2005) reports a similar study for the prisoner's dilemma game.[1] He finds that players are more likely to co-operate in indefinitely repeated games compared to finitely repeated games. This difference becomes more clearly visible in later repetitions of the repeated game compared to the first time the game is played.

On both theoretical and empirical grounds, a clear difference exists between finitely and indefinitely repeated trust games. This difference is, however, less clear when the theoretical models of the indefinitely and finitely repeated *asymmetric* trust games of incomplete information are compared. In the indefinitely repeated *asymmetric* trust game, the strict grim-trigger strategy dictates that co-operation should be terminated as soon as the trustor observes an L. In the finite version of the *asymmetric* trust game, co-operative behaviour is also short-lived because the trustor's belief is inefficiently updated. The experimental data will be used to investigate whether the difference in the horizon perspective is also visible in repeated *asymmetric* trust games.

6.3 Experimental design and hypotheses

6.3.1 The design

The three studies mentioned in Section 6.2 show that it is preferable to allow subjects in repeated social dilemma game experiments to become more experienced in order to acquire a clear difference in co-operation levels between finitely and indefinitely repeated social dilemma games. In order to gain an insight into the working of the two mechanisms of temporal embeddedness, it is not necessary that the experimental design allows for cross-sequence learning. The different mechanisms of temporal embeddedness that are at

[1] For an overview of earlier experimental results by social psychologists on repeated prisoner's dilemma games, see Colman (1982). For a discussion on economic experiments on repeated prisoner's dilemma games, see Roth and Kagel (1995).

work in repeated games with a certain end (learning and control) versus games with an uncertain end (control) should be visible, regardless of how many times they are repeated. Given the overall set-up of the experiments in this thesis, we have chosen to increase the number of observations regarding the individual characteristics (see Section 3.2.1) rather than the number of repeated games. The first is essential for estimations involving subjects' types.

The experimental design consists of three treatments. The first is, again, the Type Treatment (TT), as described in Chapter 3. The second treatment is the finitely repeated version of the *asymmetric* trust game (FT); subjects play ten periods in the FT. The third treatment is the indefinite version of the *asymmetric* trust game (IT). In the IT, subjects play five rounds for certain, but after the fifth round the game continues with a probability of $\frac{5}{6}$. In expected terms, the IT also lasts ten periods.

The analysis in Chapter 3 showed a clear drop in the frequency with which trustors and trustees choose *Co-operation* and *Honour*, respectively, after round 5. It is doubtful whether, in this instance, the subjects perceived the game as indefinitely repeated. On the one hand, it is desirable to prevent a similar outcome in this experiment, while on the other, it is undesirable to lose too many observations. Therefore we kept the experimental design from Chapter 3, but in the instruction it was stated that the game would continue at the end of each round with a probability of $\frac{9}{10}$. The subjects would play a game that would last an expected number of ten periods; like the version used in Chapter 3, but the advantage of more usable observations was maintained.

In Chapters 3 and 5, different parameter values are used. In this experiment, the parameter values of Chapter 5 are used for both treatments. See Table 5.1 for the exact values. The reason behind this choice is the fact that given these parameter values co-operative behaviour is supported in equilibrium in both treatments.

This experiment uses the 'between subjects' design (see Table 6.1). Depending on the session type, subjects were either confronted with the FT or with the IT. Subjects played the base treatment twice, once in the role of trustor and once in the role of trustee. Subjects were rematched every time they played a new game, so, including the TT, all subjects where rematched three times. Every session ended with a questionnaire.

Table 6.1: Session types as used in the comparison experiment.

Type I session:	(TT) Type game
	(FT) Finitely repeated *asymmetric* trust game (2×)
	Questionnaire
Type II session:	(TT) Type game
	(IT) Indefinitely repeated *asymmetric* trust game (2×)
	Questionnaire

6.3.2 Hypotheses

Using the theoretical framework of Chapter 2, it is easy to calculate the continuation probability $\bar{\delta}$ for which the selfish trustee will prefer to choose *Honour*. Substituting the parameter values used in this experiment in condition (2.10) leads to $\bar{\delta} \approx 0.75$. Note that $0.9 = \delta > \bar{\delta} \approx 0.75$. In other words, given the pay-offs used in this experiment, the strict grim-trigger strategy is an equilibrium strategy for selfish trustees.[2] Therefore, both honest and selfish trustees should choose honour trust in the IT and, consequently, both types of trustors should choose *Co-operation*.

However, the strict grim-trigger strategy formulated in Chapter 2 states that as soon as the trustor observes an L he should punish the trustee until the end of the game. The trustor might be less vengeful and perhaps choose to play a strict trigger strategy where he only punishes the trustee for a number of rounds, after which he will return to co-operation. From a theoretical perspective this is possible as long as $\delta > \bar{\delta}$. However, $\bar{\delta}$ will increase the more forgiving the trustor becomes.

In the IT, about 25% of the trustors should stop co-operating every round when they play the strict grim-trigger strategy, because they will observe an L. This means that even in later periods we should observe some co-operation in the IT. For example, at the beginning of period 7 around 18% of the trustors should still be prepared to choose *Co-operation*. This can be

[2] The experimental design chosen in this chapter has a disadvantage in the sense that it does not allow for a comparison between the two indefinitely repeated *asymmetric* trust game experiments. It would have been interesting to compare the two experiments, particularly given the fact that in the experiment of Chapter 3 the grim-trigger is not an equilibrium strategy, while it is in this chapter. As will become clear when Figures 3.2 and 6.4 are compared, the experimental design had a great impact on subject behaviour, hence making a useful comparison impossible.

even higher when trustors only punish for a limited number of rounds and then return to co-operative play.

In the FT, the mixing probabilities for both trustor and trustee are the same as in Chapter 5, given the fact that the same parameter values are used. In the FT, the lower the trustor's prior belief, the sooner co-operation will stop. In Chapter 5, an example is given where trustors believe that 25% of the trustees are honest. In this example, trustees should not honour trust in the final four rounds.

The models suggest that co-operation is more likely to be seen in the final rounds of the IT compared to the FT. In Chapter 3, we found that only a very limited number of trustors actually played the strict grim-trigger strategy, but about 68.52% of trustors' decisions were in line with a strict trigger strategy, suggesting that trustors punish only for a limited number of periods. This also makes it more likely to observe higher levels of co-operation in the IT compared to the FT.

Hence, we arrive at the following two hypotheses:

HYPOTHESIS 1: *Trustors will choose* Co-operation *more frequently in the IT compared to the FT.*

HYPOTHESIS 2: *Trustees will choose* Honour *more frequently in the IT compared to the FT.*

If trustors play as theoretically predicted, a difference should be observable in the trustor's reaction upon observing an L in the IT compared to the FT. The strict grim-trigger strategy, as well as other strict trigger strategies, requires a reaction as soon as an L is observed. In the FT, the trustor can ignore an L in the pure strategy phase of the game, as long as he does not observe more Ls than he should reasonably expect.

HYPOTHESIS 3: *Trustors are less likely to choose* Co-operation *in the IT compared to the FT if they have observed an L in the previous period.*

Because in this experiment the strict grim-trigger strategy is an equilibrium strategy, there should be no difference in the behaviour of non-trustful and trustful trustors in the IT. Both types of trustor should prefer to choose *Co-operation*, because control enforces the trustee to choose *Honour*. In the FT, on the other hand, trustful trustors should find it easier to choose *Co-operation*, due to their higher prior belief. This is not only true because it is easier for them to meet the threshold in the first period, but also given the fact that for them the mixing phase will start later.

HYPOTHESIS 4a: *In the FT, trustful trustors are more likely to choose* Co-operation *compared to non-trustful trustors.*

HYPOTHESIS 4b: *In the IT, there is no difference in the likelihood with which trustful trustors and non-trustful trustors choose* Co-operation.

The models of Chapters 2 and 4 suggest that, over time, trustors will choose *Co-operation* less frequently. In the IT, this decrease should be the same for both types of trustors, while in the FT the decrease should be larger for trustful trustors. Given their higher prior belief it is more likely that trustful trustors will choose *Co-operation*. When they are faced with negative evidence and no longer believe that they face an honest trustee, they should update their behaviour accordingly.

HYPOTHESIS 5a: *In the IT, the gradual decrease in the likelihood with which* Co-operation *is chosen is the same for both types of trustors.*

HYPOTHESIS 5a: *In the FT, the gradual decrease in the likelihood with which* Co-operation *is chosen is larger for trustful trustors.*

6.4 Results

6.4.1 Carrying out the experiment and descriptive statistics

In Table 6.2, basic information on the session level is presented. The experiment consisted of nine sessions. A first series of three sessions was conducted in May 2007. A second series of six sessions was conducted during September and October of the same year. A total of 166 subjects participated in this experiment, which was held at ELSE. The experiment was computer-based using z-Tree by Fischbacher (1999).

66.34% of the subjects where female and 85.96% Dutch, including double nationalities.[3] The second largest group, 3.01%, consisted of German nationals. In total, 12.16% of the subjects studied economics.

[3] The experiment was bilingual; subjects could choose to do the experiment either in Dutch or in English. There were two versions of the instructions, and before the experiment started the subjects could choose their preferred language, in which the messages would appear on their computer screen.

Table 6.2: Some basic information on the session level for the comparison experiment.

	Session 1	Session 2	Session 3
Date	22-5-07	31-5-07	31-5-07
Type	Type I	Type I	Type I
Duration*	10/10	10/10	10/10
Subjects	14	20	18
Economics students	14.29%	10%	5.56%
Male subjects	50%	30%	11.11%
Trustful trustors†	42.86%	35%	55.56%
Honest trustees†	50%	30%	33.33%
Average earnings	€8.26	€6.50	€7.01
Maximum earnings	€10.24	€7.64	€8.68
Minimum earnings	€6.60	€4.86	€5.31
	Session 4	Session 5	Session 6
Date	26-9-07	27-9-07	10-10-07
Type	Type I	Type I	Type II
Duration*	10/10	10/10	9/10
Subject	20	20	18
Economics students	10%	0%	22.22%
Male subjects	50%	60%	11.11%
Trustful trustors†	55%	50%	55.56%
Honest trustees†	35%	60%	44.44%
Average earnings	€6.80	€6.31	€7.14
Maximum earnings	€9.22	€7.52	€8.30
Minimum earnings	€4.70	€5.06	€5.15
	Session 7	Session 8	Session 9
Date	11-10-07	31-10-07	31-10-07
Type	Type II	Type II	Type II
Duration*	7/7	7/9	7/9
Subject	18	18	20
Economics students	16.67%	22.22%	15%
Male subjects	27.78%	22.22%	35%
Trustful trustors†	33.33%	50%	25%
Honest trustees†	44.44%	61.11%	30%
Average earnings	€5.97	€8.02	€7.51
Maximum earnings	€8.28	€10.69	€9.09
Minimum earnings	€4.73	€4.86	€5.67

*Duration of each game in number of rounds. †Type as derived from the TT.

Table 6.3: Overview of variable descriptive statistics for the comparison experiment.

Variable name	Obs	*Mean*	Std. Dev.	Min	Max
Dependent variables					
ChoiceA*	3042	0.502	0.500	0	1
ChoiceB*	1526	0.638	0.481	0	1
Independent variables					
Indef*	3042	0.395	0.489	0	1
Period	3042	5.162	2.753	1	10
Period10*	3042	0.066	0.249	0	1
OutcomeL−lag1*	703	0.465	0.499	0	1
Trustful*	3042	0.453	0.498	0	1
Honest*	3042	0.426	0.495	0	1
Trustorfirst*	3042	0.500	0.500	0	1
Control variables					
Female*	3042	0.663	0.473	0	1
Economics*	3042	0.122	0.327	0	1
Dutch*	3042	0.860	0.347	0	1
Age	3042	21.011	2.664	17	30
Friends	3042	0.429	0.715	0	4
Work*	3042	0.586	0.493	0	1
Volwork*	3042	0.424	0.494	0	1
Sequence*	3042	0.158	0.365	0	1
Donor*	3042	0.559	0.497	0	1
Organdonor*	3042	0.438	0.496	0	1
Religious*	3042	0.253	0.435	0	1

Note: Obs = Number of observations; Std. Dev. = Standard deviation; Min = minimum value; Max = maximum value and * indicates dummy variables, 1=yes.

When the subjects entered the laboratory they found the instructions for the TT on their desk. After this treatment, they received instructions for either the FT or the IT, depending on whether they were in a Type I or Type II session. Next to the instructions the subjects also received a small quiz, which was used to test if they had understood the instructions. They were given enough time to read the instructions, and then the instructor gave a plenary explanation of the key features of the instructions. This was always done by the same person to maintain uniformity in the explanation between sessions. After this, the instructor checked the answers given to the multiple-choice quiz questions. To avoid framing effects, abstract labels were used in the instructions for players' names and strategies.

As a consequence of the continuation rule, we have different game lengths for the Type II Sessions; the shortest game lasted seven rounds and the longest ten.

At the end of each session, the subjects were asked to fill in a questionnaire, which can be found in Appendix E. This gave the instructors time to count out the money earned by the subjects. After all the subjects had answered the questionnaire, they received their earnings. They were paid according to the outcomes of the experiment. On average, subjects earned €7.02 per session (for the details per session, see Table 6.2). A session took about 45 minutes.

In Table 6.3, an overview is given of the dependent and independent variables. Compared to Chapters 3 and 5, there is one new variable – *Indef*, a dummy variable, which is 1 if a subject is in the IT.

6.4.2 Data analysis

Table 6.4 presents the averages of *Co-operation* and *Honour* chosen for both the FT and IT. It is evidenced that, on average, *Co-operation* and *Honour* are chosen more frequently in the IT compared to the FT. In Figure 6.1, the same information is provided per period. The probability that *Co-operation* is chosen decreases over time. The probability that trustees respond with *Honour* remains reasonably stable, although a clear end round effect is visible for the FT. The final three periods can be compared less easily, because there are less observations for the IT. According to the Wilcoxon-Mann-Whitney (WMW) ranksum test, the difference between the period averages

Table 6.4: *Co-operation* and *Honour* rates for both FT & IT in the comparison experiment.

	Co-operation	*Honour*
FT	0.477	0.620
IT	0.539	0.664

in trust placed and trust honoured is not significant.[4]

In the IT, in 4 out of the 74 groups the trustor played the strict grim-trigger strategy and in six groups the trustor never chose *Co-operation.* Most trustors violate the strict grim-trigger strategy later on in the game. If all decisions after and including the moment trustors violated the strict grim-trigger strategy are ignored, we find that 40.10% of the decisions made by trustors are in line with the strict grim-trigger strategy.[5] When we simply count all the decisions that are in line with the strict trigger strategy, regardless of the moment in the game they are made, then 76.54% of the decisions are in line with the strict trigger strategy.

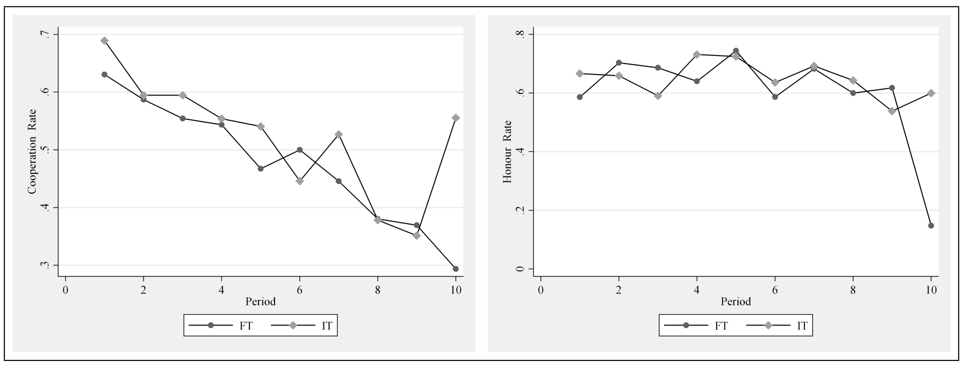


Figure 6.1: *Co-operation* and *Honour* rates over time for the comparison experiment.

In order to test the hypotheses, panel logit models including random effects at the level of subjects were estimated. As a robustness check we also ran estimations using the balanced panel, which did not alter the findings. Variables from the category 'control variables', as presented in Table 6.3, are only mentioned in the regression result when they have a significant influence on the results.

FINDING 1: *Non-trustful trustors do not choose* Co-operation *more frequently in the IT compared to the FT. Trustful trustors are more likely to choose* Co-operation *in the FT compared to the IT.*

Table 6.5 shows that there is no significant difference in the frequency with

[4]For the Co-operation rate WMW-ranksum test reports p=0.3642 and for the Honour rate p=0.6230

[5]In Section 3.3.3, in 'line with the strict grim-trigger strategy' is explained.

which non-trustful trustors choose *Co-operation* between the IT and FT, regardless of whether the subject was first assigned the role of trustor in the experiment. Hence, hypothesis 1 should be rejected.

A significant negative time trend is visible. Given the development of the *Co-operation* averages, reported in Figure 6.1, this result is not wholly surprising. For a discussion on the variables *Trustful* and *Female*, see finding 4.

FINDING 2: *Trustees do not choose* Honour *more frequently in the IT compared to the FT.*

Trustees do not choose *Honour* more often in the IT, regardless of the type and order of the role assignment. Hypothesis 2 must therefore be rejected. Trustees in the IT do choose *Honour* more often in period 10 compared to trustees in the FT. Given the limited number of observations for trustees in period 10 in the IT, this result should be carefully interpreted. For trustees in the FT, a clear end round effect is visible, with a p-value of less than 0.001. Furthermore, a negative time trend is visible, so it is significantly less likely that trustees will choose *Honour* later on in the game.

Finally, the type dummy is significant. Honest trustees do choose *Honour* more frequently compared to selfish trustees, regardless of the treatment they are in. This result is in line with theoretical predictions as far as the FT is concerned, but in the case of the IT no type effect is to be expected. Although theory predicts that in the case the strict grim-trigger strategy is an equilibrium strategy, honest trustees should not choose *Honour* more frequently compared to selfish trustee, no support can be found for this in the data. See Table 6.6 for the complete estimation.

FINDING 3: *Trustors are less likely to choose* Co-operation *in the IT compared to the FT if they have observed an L in the previous period.*

Table 6.7 presents the result on the trustors' likelihood to choose *Co-operation* after they have observed an L in the previous round. [6] We find that it is less likely that trustors will choose *Co-operation* in the FT after they have observed an L, with a p-value of less than 0.001; however, trustors in the IT are even less likely to do so, with p=0.012.

[6] In order to yield consistent estimators, the variable *OutcomeLperiod1* and its interaction with *Indef* is included in the regression model presented in this table. See Chapter3 for an explanation.

Table 6.5: Random-effects logit model for the probability of choosing *Co-operation* in the comparison experiment.

Dependent variable: ChoiceA				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Indef	0.551	0.721	0.77	0.444
Period	−0.097	0.045	−2.15	0.031
Indef×Period	−0.103	0.074	−1.40	0.161
Period10	−0.288	0.335	−0.86	0.391
Indef×Period10	1.000	0.906	1.10	0.269
Trustful×Period	−0.205	0.061	−3.37	0.001
Trustful×Period×Indef	0.182	0.106	1.72	0.086
Trustful	2.104	0.512	4.10	0.000
Trustful×Indef	−1.674	0.796	−2.10	0.035
Trustorfirst	0.076	0.384	0.20	0.843
Indef×Trustorfirst	−0.412	0.589	−0.70	0.484
Female	0.545	0.393	1.39	0.165
Female×Indef	0.852	0.652	1.31	0.191
Constant	−0.439	0.448	−0.98	0.328
Number of observations: 1521, Number of subject: 166				
LL: -859.349, σ_u: 1.608, ρ : 0.440⋆				

⋆ LL: Log Likelihood, σ_u: Standard deviation of the error term at the subject level, ρ: proportion of the variance in the composite error term that is at the subject level, or $\rho = \frac{\sigma_u^2}{\sigma_u^2+\sigma_e^2}$.

Table 6.6: Random-effects logit model for the probability of choosing *Honour* in the comparison experiment.

Dependent variable: ChoiceB				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Indef	−0.775	0.742	−1.04	0.297
Period	−0.176	0.067	−2.63	0.009
Indef×Period	0.072	0.091	0.80	0.426
Period10	−3.014	0.769	−3.92	0.000
Indef×Period10	3.910	1.543	2.53	0.011
Honest	2.792	0.705	3.96	0.000
Honest×Period	−0.028	0.086	−0.32	0.745
Honest×Indef	−0.304	0.780	−0.39	0.697
Trustorfirst	−0.611	0.518	−1.18	0.237
Indef×Trustorfirst	1.072	0.767	1.40	0.162
Constant	0.634	0.515	1.23	0.218
Number of observations: 763, Number of subject: 148				
LL: -386.574, σ_u: 1.784, ρ : 0.492				

Table 6.7: Random-effects logit model for the probability of choosing *Co-operation* after observing an L in the previous period in the comparison experiment.

Dependent variable: ChoiceA				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Indef	0.016	0.852	0.02	0.985
Period	−0.188	0.087	−2.16	0.031
Indef×Period	0.113	0.107	1.06	0.291
Period10	−0.428	0.537	−0.80	0.426
Indef×Period10	−0.969	1.592	−0.61	0.543
OutcomeL−lag1	−2.071	0.304	−6.82	0.000
OutcomeL−lag1×Indef	−1.289	0.516	−2.50	0.012
Trustful	0.247	0.661	0.37	0.708
Trustful×Period	0.029	0.091	0.32	0.751
Trustful×Indef	−0.481	0.601	−0.80	0.423
Trustorfirst	−0.226	0.381	−0.60	0.552
Indef×Trustorfirst	0.593	0.594	1.00	0.318
OutcomeLperiod1	−0.951	0.430	−2.21	0.027
OutcomeLperiod1×Indef	0.312	0.636	0.49	0.623
Constant	3.205	0.632	5.07	0.000
Number of observations: 703, Number of subject: 147				
LL: -332.674, σ_u: 1.084, $\rho : 0.263$				

Table 6.8: Random-effects logit model for the probability of choosing *Co-operation* in the FT from the comparison experiment.

Dependent variable: ChoiceA				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Period	−0.100	0.046	−2.19	0.029
Period10	−0.294	0.339	−0.87	0.387
Trustful	2.215	0.540	4.10	0.000
Trustful×Period	−0.207	0.062	−3.35	0.001
Trustorfirst	0.095	0.412	0.23	0.817
Female	0.515	0.422	1.22	0.223
Economics	−1.463	0.835	−1.75	0.080
Constant	−0.371	0.477	−0.78	0.437
Number of observations: 920, Number of subject: 92				
LL: -502.711, σ_u: 1.738, $\rho : 0.479$				

FINDING 4a: *In the FT, trustful trustors are more likely to choose* Co-operation *compared to non-trustful trustors.*

FINDING 4b: *In the IT, there is no difference in the likelihood with which trustful trustors and non-trustful trustors choose* Co-operation.

FINDING 5a: *In the IT, the gradual decrease in the likelihood with which* Co-operation *is chosen is the same for both types of trustors.*

FINDING 5b: *In the FT, the gradual decrease in the likelihood with which* Co-operation *is chosen is larger for trustful trustors.*

As can be clearly seen in Table 6.8, trustful trustors choose *Co-operation* more frequently in the FT, as theory suggests. Over time, when the trustors update their belief about whether they are facing an honest trustee or not, the frequency with which they choose *Co-operation* drops. As can be seen from the interaction of *Trustful*×*Period*, this effect is stronger for trustful trustors. [7] At the 10% significance level it is less likely that economic students will choose *Co-operation* in the FT.

In the IT, trustful trustors behave no differently than non-trustful trustors as theory suggests (see Table 6.9). For both types of trustors a negative time trend is visible, which is to be expected when the subjects play a trigger strategy. These results can also be found in Table 6.5. A Chi^2-test on *Trustful* and *Trustful*×*Indef* shows that there is no significant influence of trustor type in the IT. A similar test on *Trustful*×*Period* and *Trustful*×*Period*×*Indef* does not show a significant effect either.

Female subjects choose *Co-operation* more frequently in the IT compared to the male subjects, which is also supported by the estimation presented in Table 6.5. A Chi^2-test on *Female* and *Female*×*Indef* shows that female subjects are significantly more likely to choose *Co-operation* in the IT.

6.5 Discussion and conclusion

In this chapter, finitely and indefinitely repeated *asymmetric* trust games were compared. The theoretical mechanisms that allow for co-operative

[7] This result should not come as a surprise, especially given that the data from the FT come from the experiment discussed in Chapter 5, where in Table 5.6 the same result can be found.

Table 6.9: Random-effects logit model for the probability of choosing *Co-operation* in the IT from the comparison experiment.

Dependent variable: ChoiceA				
Independent variables	Coef.	Std. Err.	z	$P > \|z\|$
Period	−0.193	0.057	−3.36	0.001
Period10	0.697	0.831	0.84	0.401
Trustful	0.467	0.579	0.81	0.420
Trustful×Period	−0.025	0.085	−0.29	0.770
Trustorfirst	−0.279	0.415	−0.67	0.501
Female	1.284	0.479	2.68	0.007
Economics	−0.218	0.526	−0.41	0.678
Constant	0.160	0.521	0.31	0.759
Number of observations: 601, Number of subject: 74				
LL: -354.168, σ_u: 1.397, ρ : 0.372				

behaviour in these two types of games are quite different.

We found that the average rate with which trustors choose *Co-operation* and trustees choose *Honour* is higher in the IT compared to the FT. The difference is, however, not significant. The RE logit estimations do not show a significant difference for the treatment effect either, which is not very surprising. First of all, this is because the theoretical models show that the difference between finitely and indefinitely repeated games becomes smaller in the presence of asymmetric information. The models suggest that co-operation might be more easily sustained in the indefinitely repeated settings, but the theoretical difference is small to begin with. Secondly, earlier findings from Engle-Warnick and Slonim (2004) and Dal Bó (2005) show that a significant difference in the level of co-operation is only found when subjects get more experienced playing repeated social dilemma games.

Hence, the data support earlier findings that there is no difference in the level of co-operation between finitely and indefinitely repeated dilemma games when subjects are inexperienced in these games. It could be interesting to test the finitelyand indefinitely repeated *asymmetric* trust games in an experimental setting that allows for cross-sequence learning, in order to test whether a difference between the two settings is visible when the subjects are more experienced.

The data showed a significant difference between the treatments when analysing how trustors react to observing a low pay-off in the previous period. This finding seems to indicate that subjects know intuitively how control works differently in finitely and indefinitely repeated trust games, in the sense that either learning or the sanction strategy determines the moment the trustor starts choosing *No co-operation*. This result supports the earlier finding of Engle-Warnick and Slonim (2004) that trustors play different strategies in finitelyand indefinitely repeated trust games. Trigger strategies require a strong reaction the moment they are violated. Reputation models require a strong reaction from the trustor when he no longer believes that he is playing with an honest trustee. When the coefficients of Table 6.7 are studied, it can be seen that the combined effect of the constant and *OutcomeL−lag1* variable is positive for trustors in the FT. However, for trustors in the IT the combined effect is negative. It seems that trustors in the IT are prepared to punish immediately, while for trustors in the FT observing one L is not enough proof for them that they are dealing with a selfish trustee.

The data support most of the theoretical predictions about how the different types of trustors and trustees should behave, and what is especially interesting is the result that trustful trustors act no differently compared to non-trustful trustors in the IT when the strict grim-trigger strategy is an equilibrium strategy. This result is the opposite of the result found in Chapter 3, where the strict grim-trigger strategy was not an equilibrium strategy in the NCT. Trustful trustors choose *Co-operation* in the FT more frequently compared to non-trustful trustors. Over time, both types of trustors choose *Co-operation* less frequently, but, as expected, this effect is stronger for trustful trustors.

Honest trustees are more likely to choose *Honour* compared to selfish trustees, which is in line with theoretical insights when it comes to the FT; nevertheless, in the case of the IT, both types of trustees should behave in the same way. The data do not support the latter.

Chapter 7

Summary and conclusion

In Chapter 1, we presented a short overview of the broader literature on trust and indicated to which sub-branch of the literature this thesis contributes. In this sub-branch of the economic literature on trust, trust is usually seen as a desirable 'commodity', because it can facilitate co-operation between trustors and trustees. The trustor faces a dilemma when he wants to trust the trustee. The trustee's interest might not be perfectly aligned with his own and the trustor lacks knowledge about the trustee's intentions. The economic consequence is that a trustor runs the risk of suffering losses that would not result if the trustor and trustee's interests were aligned. As long at we can assume that voluntary co-operation allows for a Pareto improvement, it is interesting to study how trustors and trustees can overcome this trust problem.

In the literature, it is suggested that the trust problem can be overcome when trust is socially embedded, either through networks, institutions and/or time. We indicted that trust is studied in this thesis under temporal embeddedness. The literature on indefinitely and finitely repeated games shows that, when a trustor and trustee interact with each other over a longer period of time, the mechanisms of temporal embeddedness – control and learning – can support co-operation in equilibrium.

In this thesis, trust is studied in a setting where the trustor faces not only incomplete information about the trustee's preferences, but also he is ill-informed about the decision taken by the trustee. Recall from our example of John and William that John does not know whether he should trust William, because he does not know whether William will do a good job when it comes to investing John's money. Secondly, if John decides to trust William, he faces the problem that he cannot observe whether the returns on the investment reported by William are mainly driven by general market conditions or by William's expert knowledge and efforts. We labelled the situation where the trustor cannot directly observe the trustee's action, because it is hidden by fluctuations in external conditions as a situation of asymmetric information. The few available studies on asymmetric information in repeated social dilemma games (Green and Porter (1984) and Aoyagi and Fréchette (2008)) suggest that co-operation is less easily established and sustained in the presence of information asymmetry.

Many real-life situations can be described by repeated *asymmetric* trust games. It is important to gain an insight into how control and learning become less effective in sustaining co-operation in these games, and whether the checking option introduced in Section 1.5.2 can help trustors to overcome the problem of information asymmetry. In order to do so, we extended the theoretical work on asymmetric information in repeated trust games and ran several experiments to test our models. In Section 7.1, we will present our results, which will be reflected upon in Section 7.2. This chapter ends with an epilogue.

7.1 Results

7.1.1 Theoretical insights

In line with what, intuitively, was to be expected, the theoretical models from Chapters 2 and 4 show that co-operation is less easily established and sustained in the repeated *asymmetric* trust game compared to its symmetric counterpart.

To analyse the indefinitely repeated *asymmetric* trust game, we adapted the model from Green and Porter (1984). In this model, asymmetric information forces the trustor to use a less effective strict grim-trigger strategy. Information asymmetry reduces the effectiveness of the control mechanism in establishing and sustaining co-operation in two ways.

First of all, it makes it less likely that the strict grim-trigger strategy will be an equilibrium strategy. Because the equilibrium threshold increases, a higher continuation probability is required to support co-operation in equilibrium. The source of the increase lies in the fact that in the case of asymmetric information the strict grim-trigger strategy punishes on the basis of observed pay-offs. The more asymmetric the information partition becomes, the more this strategy allows for the unjust punishment of trustees who respond with *Honour*, while at the same time trustees who respond with *Abuse* will be more likely go unpunished.[1]

Secondly, when the grim-trigger strategy is an equilibrium strategy, it is highly unlikely that it will support co-operation throughout the entire game. The higher the probability that the trustor observes a low pay-off when the trustee chooses *Honour*, the more likely it becomes that the sanction of choosing *No co-operation* will be activated by the strict grim-trigger strategy.

We extended the model of Green and Porter (1984) by including social preferences. In our model, the trustee's utility function reflects not only his own material well-being, but also his social orientation towards the trustor's well-being. On the basis of this utility function, three types of trustees can be distinguished. Selfish trustees prefer to choose *Abuse* in the isolated encounter trust game, while honest trustees prefer to choose *Honour* (see Chapter 1). In indefinitely repeated trust games, an additional group of trustees can be distinguished – the so-called conditional honest trustee. The models show that the more conditional honest trustees care about the well-being of trustors, the more likely it becomes that they will prefer to choose *Honour* in the indefinitely repeated *asymmetric* and *symmetric* trust game.

We also studied the indefinitely repeated *(a)symmetric* trust game under incomplete information. Even when the (strict) grim-trigger strategy is not an equilibrium strategy for selfish trustees, co-operation might still be supported in equilibrium when the trustor believes that a significant group of trustees can be labelled as being honest or conditional honest.

In Chapter 4, we extended the literature by analysing the finitely repeated *asymmetric* trust game under incomplete information. The models from this chapter show that information asymmetry influences equilibrium behaviour in a very different way compared to the existing models on indefinitely repeated games of asymmetric information. In the case of the finitely repeated *asymmetric* trust game, asymmetric information has a negative influence on the trustor's capability to update his prior belief, i.e. to learn about the type of trustee he is facing. The models of Chapter 4 showed that the trustor's threshold for choosing *Co-operation* in presence of asymmetric information increases compared to the same threshold under symmetric information. This influences co-operation in three ways.

[1]The information partition becomes more asymmetric when the probabilities that determine how likely it is that a low pay-off or a high pay-off is drawn in the *asymmetric* trust game move in the direction of 0.5. For example, the probability of observing a high pay-off is close to 1 if the trustee responds with *Honour*, and the probability of observing a low pay-off is close to 1 if the trustee responds with*Abuse*. When these probabilities move in the direction of 0.5, the information partition becomes more asymmetric.

First, the trustor's prior belief might not be high enough to meet the trustor's threshold to choose *Co-operation* in the first round, which is especially true for shorter games. The longer the game is repeated, the lower the trustor's threshold will be in the first round.

Second, if the trustor's prior belief is high enough to meet the trustor's threshold in the first period, co-operation will last less time compared to the *symmetric* trust game. The selfish trustee will be forced to start randomising over his actions sooner, given the fact that the trustor's threshold to choose *Co-operation* rises faster.

Third, when randomising begins, the selfish trustee's mixing probability must favour *Abuse* more in order to make the trustor indifferent about choosing between *Co-operation* and *No co-operation* in equilibrium. The selfish trustee needs to choose *Abuse* with a larger probability, because with asymmetric information the trustor requires additional assurance that he faces an honest trustee.

Although the trustor's mixing probability favours *Honour* more with asymmetric information compared to symmetric information, this does not offset the overall negative effect of asymmetric information on the level of co-operation in the finitely repeated *asymmetric* trust game. From the models we can conclude that because learning is less effective under information asymmetry, the trustor starts to sanction the trustee sooner and, hence, control becomes less successful in sustaining co-operation in finitely repeated trust games.

7.1.2 Empirical results

The experimental data show that a vast majority of the trustors – 86% in the indefinitely repeated *asymmetric* trust game and 77.54% in the finitely repeated *asymmetric* trust game – used the checking option when they had the opportunity. The checking option has a positive effect on the frequency with which trustors choose *Co-operation* and trustees choose *Honour*. Checking increases the effectiveness of the mechanisms of temporal embeddedness and therefore has a positive influence on the level of co-operation in repeated *asymmetric* trust games.

In line with the results from the studies mentioned in Section 1.3, we found that the overall level of co-operation in repeated games decreases over time.

The direct comparison of finitely and indefinitely repeated *asymmetric* trust games did not show a clear difference in the frequency with which trustors and trustees choose *Co-operation* and *Honour*. This was to be expected, especially given the earlier results from Engle-Warnick and Slonim (2004) and Dal Bó (2005) that show that the difference between repeated games with certain and uncertain ends becomes clearer when the experimental design allows for ‘cross-sequence learning’.

Control works differently in finitely and indefinitely repeated *asymmetric* trust games. In finitely repeated *asymmetric* trust games, the trustor’s sanction of choosing *No co-operation* is triggered by the trustor’s learning process. In indefinitely repeated *asymmetric* trust games, this sanction is triggered by the strict grim-trigger strategy; we found support in the data for this distinction. Trustors are less likely to choose *Co-operation* after they have observed a low pay-off. In line with theoretical predictions, we found that this effect is more profound in the indefinitely compared to the finitely repeated *asymmetric* trust game. This finding supports the idea that the strict grim-trigger strategy requires a direct reaction as soon as a low pay-off is observed. In relation to learning, the trustor should stop choosing *Co-operation* when he no longer believes that he faces an honest trustee. When the trustor’s prior belief is relatively high, observing one low pay-off might not induce the trustor to stop choosing *Co-operation* immediately. This result supports the earlier finding of Engle-Warnick and Slonim (2004) that trustors play different strategies in indefinitely and finitely repeated games.

The models from Chapters 2 and 4 suggest that non-trustful trustors and selfish trustees should behave differently compared to trustful trustors and honest trustees, depending on the equilibrium conditions. With some exceptions, mostly in the case of honest trustees in indefinitely repeated games, the data support these theoretical predictions.

In both the Checking Treatment from Chapter 3 and the indefinitely repeated *asymmetric* trust game treatment from Chapter 6, the (strict) grim-trigger strategy is an equilibrium strategy. When the (strict) grim-trigger strategy is an equilibrium strategy, it aligns the behaviour for all types of trustees. Hence, the trustor’s belief about the proportion of honest trustees should have no influence on his decision to choose *Co-operation*. The data support the prediction that trustful trustors should not act differently compared to non-trustful trustors.

Selfish trustees and honest trustees should both choose *Honour* when the grim-trigger strategy is an equilibrium strategy; the data are ambiguous on this point. The data from the Checking Treatment of Chapter 3 show that trustee type has no influence on the decision to choose *Honour*, as was to be expected. However, in the indefinitely repeated *asymmetric* trust game treatment from Chapter 6, honest trustees choose *Honour* more frequently compared to selfish trustees. The most obvious explanation can be found in the fact that control is far less effective in the treatment from Chapter 6. Though the strict grim-trigger strategy is an equilibrium strategy, it is a less effective strategy than the 'normal' grim-trigger strategy that can be played in the Checking Treatment from Chapter 3. Our findings may indicate that control in the form of the strict grim-trigger strategy is not strong enough to align the behaviour of all trustees in the indefinitely repeated *asymmetric* trust game.

In the No-Checking Treatment from Chapter 3, where the grim-trigger strategy is not an equilibrium strategy, trustful trustors are more likely to choose *Co-operation* compared to non-trustful trustors. Honest trustees, however, are no more likely to choose *Honour* in this treatment. While the first finding is supported by the theory, the second is not. In Chapter 3, we argued that this result might be explained by the fact that it takes subjects some time to find out whether a strategy is supported in equilibrium. Dal Bó and Fréchette (2007) compared prisoner's dilemma games where the grim-trigger strategy does and does not support co-operation. They found that subjects do co-operate more when co-operation is supported by the grim-trigger strategy. This difference becomes, though, only visible when subjects are experienced with playing repeated prisoner's dilemma games.

The sequential equilibrium assumes that honest trustees always choose *Honour*. The results presented in Chapter 5 show that honest trustees choose *Honour* more frequently and that they remain honest over time. The only deviation from theory that we found is that the effect of being honest is smaller in the Checking Treatment compared to the No-Checking Treatment. At first sight, crowding out theory might provide an intuitive explanation for this latter result. However, in the additional estimations we ran, no negative relation was found between the frequency with which honest trustees choose *Honour* and whether they had been checked or not. The data do not, in this instance, support a crowding out hypothesis. The theoretical suggestion that a trustful trustor should choose *Co-operation* more frequently in the finitely repeated symmetric and *asymmetric* trust games is supported by the experimental data. The result that trustful trustors and honest trustees choose *Co-operation* and *Honour*, respectively, more frequently in finitely repeated trust games compared to non-trustful and selfish trustees confirms a similar earlier result found by Bohnet and Huck (2004).

Theory suggests that in indefinitely and finitely repeated trust games, trustors should decrease the frequency with which they choose *Co-operation* over time. In the case of indefinitely repeated trust games, a similar decrease should be visible for both non-trustful and trustful trustors, because the grim-trigger strategy determines when they should choose *No co-operation*. However, in the case of finitely repeated trust games, a stronger decrease should be visible for trustful trustors. The reason behind this lies in the fact that trustful trustors start with a higher prior belief. When trustful trustors receive evidence that they face a selfish trustee, i.e. they observe *Abuse*, they need to update their belief at the appropriate downward rate compared to the non-trustful trustor who had a lower prior belief to start with. This difference in the behaviour of trustful trustors over time between indefinitely and finitely repeated trust games is confirmed by the data.

7.1.3 Answering the research questions

This thesis is the first study to investigate the possibility of countering asymmetric information in repeated trust games by using the checking option. The theoretical models from Chapters 2 and 4 show that asymmetric information has a negative impact on the level of co-operation in indefinitely and finitely repeated trust games. The checking option can turn the *asymmetric* trust game into a *symmetric* trust game. From a theoretical point, it should be easier to achieve a higher level of co-operation in the *symmetric* trust game, which is supported by the experimental data. Subjects often used the checking option to turn the *asymmetric* trust game into a more symmetric trust game. As a result, higher levels of co-operation were reached in the treatment with the checking option, which themselves had the potential to increase the income of the subjects. To answer the first research question, we can conclude that, from both a theoretical and an empirical perspective, the checking option has a positive effect on the level of co-operation in indefinitely and finitely repeated *asymmetric* trust games of incomplete information.

Whether a trust game of complete information is finitely or indefinitely repeated can have a serious impact on the level of co-operation that can be reached, both from theoretical and empirical perspectives (Engle-Warnick and Slonim (2004) and Slonim et al. (2006)). We raised the question as to whether a similar difference could be found for *asymmetric* trust games of incomplete information. From a theoretical perspective, some arguments could be made that suggest that in the case of an indefinite repetition a higher level of co-operation could be expected. The experimental data showed no significant difference between the two horizon perspectives on the level of co-operation, which is in line with results from earlier experimental studies that show that a difference in horizon perspectives becomes visible only when the experiment allows for 'cross-sequence learning'.

The answer to the second research question is more ambiguous. At first sight the difference in the horizon perspective does not seem to matter in repeated *asymmetric* trust games. We found, however, that trustors play different strategies under the two horizon perspectives. Although no difference in the level of co-operation was found, on a more detailed level differences were visible while comparing the two horizon perspectives. This brings us to the third and final research question, which is about the role of trustor and trustee types in repeated games. The differences in the behaviour of these types are mainly determined by the mechanisms of temporal embeddedness and how they operate under the two horizon perspectives.

To the best of our knowledge, this is the first study to use type indicators derived from an isolated encounter treatment to test whether trustor and trustee types behave in repeated trust games, as theory suggests they should do. Theory provides us with some clear predictions on how trustful and non-trustful trustors, on the one hand, and selfish and honest trustees, on the other hand, should behave. The majority of the results on the type indicators support the theoretical predictions, although there are some deviations, mainly with respect to honest trustees. On the basis of the data, we can conclude that the distinction between trustful and non-trustful trustors and honest and selfish trustees in repeated trust games is not merely theoretical.

7.2 Reflection

7.2.1 Shortcomings and future research

Designing an experiment means making choices. We are aware of the fact that these choices can affect the outcome of an experiment and, hence, we will reflect on some of these choices.

In the experimental design for the indefinitely repeated trust game, we tried to get the best of two worlds: a game that subjects would perceive as being open-ended, and at the same time a game that would yield enough observations to allow for meaningful statistical inference. In Chapter 3, we went for an experimental design where the subjects would play five rounds for certain before a continuation probability determined whether the game should continue into the next round or not. The models in Appendix A show that the equilibrium strategies should not be affected by the introduction of a number of certain rounds. Figure 3.2 indicates a clear drop in the rate with which trustees choose *Honour*. In the experiment from Chapter 6, subjects played the same game, but in the instructions we mentioned that from the end of round 1 a continuation probability would determine whether the game would continue or not. Comparing Figures 3.2 and 6.1 highlights that this adjustment of the instruction had a great influence on how trustees played the game. When subjects are informed about the fact that they will play a number of certain rounds, it seems that they do not consider the game to be indefinitely repeated, whereas when they are not informed about this, they do.

The original design failed in delivering the best of two worlds. The alternative design, used in Chapter 6, has a significant downside in the sense that it requires the experimenter to not fully inform the subjects. Increasing observations while maintaining the perception of an indefinitely repeated game might prove to be a difficult combination.

In order to incorporate asymmetric information in the trust game, we came up with the *asymmetric* trust game, wherein nature determines which pay-off the trustor will earn and the trustee determines, by his own choice, the probability distribution used by nature. In Chapter 1, we argued that the isolated encounter *asymmetric* trust game is strategically equivalent to the isolated encounter (binary) trust game. When the *asymmetric* trust game is repeated, the effect of information asymmetry becomes apparent, as clearly shown by the models from Chapters 2 and 4. The trustor's decision is determined by the pay-offs he observes. However, when trustors use the checking option they can observe the trustee's response, just like in the standard trust game. Given the rationality assumption, it is expected that trustors will use the checking option, and when they do so they should prefer to use the additional information they receive, because it enables them to play a strategy that is more effective in sustaining co-operation. The data showed, though, that pay-off realisations do have an influence on the trustor's decision to choose *Co-operation*, which contradicts our assumption that rational people should only care about expected pay-offs. The trustor's expected pay-offs are higher when the trustee responds with *Honour* compared to *Abuse*; therefore, the trustor should only care about the trustee's decision and not about his pay-off realisations.

Reinforcement learning models can explain why pay-off realisations matter. According to these models, strategies that prove to be successful are played more frequently than those that are less successful. In this context, earning a high pay-off should be interpreted as a successful outcome and, hence, trustors should make the same decision again, regardless of the trustee's underlying response.

With respect to future research, it might be interesting to test whether pay-off realisations continue to be a driving force behind the trustor's decision when he has the chance of gaining more experience playing repeated *asymmetric* trust games. When pay-off realisations do have a lasting impact on the trustor's strategy, a logical follow-up would be the integration of this insight into the models presented in this thesis.

There are several other interesting directions for future research. In most real-life situations, checking will be costly. It would therefore be interesting to study the effect that the cost of checking can have on the trustor's decision to use the checking option.

Alternatively, it might also be interesting to change the moment the checking option becomes available. In the *asymmetric* trust game with the checking option, the trustor needs to decide whether he wants to use this option before he observes nature's move. When they can decide to use the checking option after nature's move, trustors might limit their checking behaviour in the sense that they might only check when they observe a low pay-off.

Another interesting direction for future research would be the situation where trustors and trustees can gain more experience with repeated trust games. The results of Engle-Warnick and Slonim (2004) and Dal Bó (2005) show that the difference between finitely and indefinitely repeated games is stronger when subjects become more experienced. It is likely that the few irregularities that we found regarding the behaviour of honest trustees will disappear when subjects are given the opportunity to gain more experience with repeated trust games.

A fourth direction of research could be to investigate the role of subject anonymity. In experimental economics, subject anonymity is one of the benchmark conditions for running experiments. Especially in this setting, it might be interesting to test if subject anonymity affects the results. The experimental results presented in this thesis show that checking increases the frequency with which trustees choose *Honour*, but we have to consider that trustees might possibly react aversely when they are checked by someone they know. In other words, can checking crowd out trustworthiness when the setting is no longer anonymous?

In line with this question it might be interesting to compare the *asymmetric* trust game – with a checking option – with the *symmetric* trust game. Theoretically, there should be no difference between these settings, but when crowding out plays a role, trustees might be less prepared to co-operate.

Finally, additional experiments can also be used to test the sensitivity of the type indicators and the equilibrium conditions for changes in the pay-off values, as well as the continuation probability, to further investigate the robustness of our findings.

7.2.2 Trust in experiments

In Chapter 1, we introduced the following definition of trust: *Trust is the trustor's decision to choose* Co-operation *in the case of a trust game, given his expectation that the trustee will respond with* Honour. We argued that trust in this definition is a combination of behaviour and disposition. Behaviour is the act of choosing *Co-operation* and disposition is the subjective probability generating the expectation that the trustee will act in a trustworthy manner.

The experimental data show primarily behaviour. We observed the decisions made by our subjects, but did not observe their state of mind while they made their decisions. The question could therefore be raised as to whether we truly observe trust. It could be that most trustors think that their trustee will most likely behave in an untrustworthy way; however, if we assume that subjects are rational beings, they should only choose *Co-operation* when they expect that the trustee will respond with *Honour*. In this sense, the disposition criterion is satisfied as long as we assume a reasonable degree of rationality from our subject. Although we have no conclusive evidence, the answers we received to one statement from the questionnaire do support the presence of an expectation that the outcome will be positive. On a 7-point scale the subjects filled in how much they agreed or disagreed with the following statement: "Most people will respond kindly when they are trusted by others". The results testify that 91.21% of the subjects agreed, while 6.51% reacted with a neutral response.

7.3 Epilogue

In this thesis, we presented economic research on trust. We investigated under what condition trust can enhance and facilitate co-operation between trustors and trustees in repeated *asymmetric* trust games. We found that information asymmetry lowers the frequency with which trustors and trustees are prepared to choose *Co-operation* and *Honour*. The checking option is an instrument that has a positive effect on the effectiveness of learning and control, and in the presence of the checking option trustors and trustees reach higher levels of co-operation in finitely and indefinitely repeated *asymmetric* trust games.

We argued that many economic settings can be described by the repeated *asymmetric* trust game, for example the relation between John and William, voters and their representatives, doctors and patients, and deposit holders and banks. Developments in 2008 on the global financial markets show that trust is essential for the efficient functioning of these markets, while information asymmetry about the real value of financial products and/or the solvability of financial institutions can destroy existing trust relations and seriously slow down the development of new ones.

Translating the checking option to real-life economic situations is beyond the scope of this thesis. In order to undertake this task, detailed information is required about specific real-life trust relations and the way they are socially embedded. For the application of the research presented in this dissertation on real-life trust relations, one should take the following into account:

The checking option can take on a broad range of shapes. In this thesis, the checking option is presented as an instrument used by an individual trustor, which can be realistic when the costs of the checking option are relatively low or the benefits are relatively high. Think, for example, about the doctor-patient relationship, where the patient can get a second opinion from a different doctor. When the existence of an information gap due to information asymmetry is considered a collective problem, for example within the relationship between deposit holders and banks, it might be cost effective and more practical to opt for a collective checking option in the form of government-appointed supervisors and/or regulatory agencies. The frequency with which information asymmetry arises, the cost of the checking option, as well as moral outcries against information asymmetry will probably determine whether the checking option will be a collective or an individual instrument.

It is also important to understand that a real-life checking option might not be able to close the information gap completely. The decision to use the checking option in a specific situation should be a trade-off between the effectiveness of the checking option and its price.

Overall, we can conclude that information asymmetry in repeated trust games, together with the checking option, is a relevant and interesting field of research, and deserves more attention in order to improve our fundamental understanding and to get a better insight into real-life applications.

Appendix A

Indefinitely repeated *asymmetric* trust games with $K+1$ certain rounds

In the experiment discussed in Chapter 3, we let the subjects play an indefinitely repeated *asymmetric* trust game, which differs from the game described in section 2.2. In the original setting, discussed in Chapter 2, at the end of a period the continuation probability determines whether a next period will be played. Hence, both players will always play one period for certain, namely the first period.

In the new setting, used in the experiment discussed in Chapter 3, subjects first play K rounds without a continuation probability. This means that each subject will play $K+1$ number of rounds for certain. After $K+1$ rounds, the continuation probability δ will determine whether the game will continue into the next round. In section 3.2.1 we mentioned that the subjects play five rounds for certain, i.e. $K=4$. If conditions (2.3) and (2.10) are satisfied, all types of trustees prefer to choose *Honest* in the CT and NCT, respectively. Adding certain rounds should not affect the trustee's strategy, because adding certain rounds only lowers $\tilde{\bar{\delta}}$ and $\bar{\bar{\delta}}$. This can be seen in the following equations where k represents the number of certain rounds left before the continuation rule is applied. To present this result in the most general form, we include social preferences in the equations.

Indefinitely repeated *asymmetric* trust game with checking.

$$\Big(C_2 + \rho E(C_1)\Big) + k\Big(C_2 + \rho E(C_1)\Big) + \sum_{t=k+1}^{\infty} \delta^{t-k}\Big(C_2 + \rho E(C_1)\Big)$$

$$> \Big(A_2 + \rho E(S_1)\Big) + k(N_2 + \rho N_1) + \sum_{t=k+1}^{\infty} \delta^{t-k}(N_2 + \rho N_1). \tag{A.1}$$

Solving for δ yields:

$$\delta > \check{\delta} \equiv \frac{\Big(A_2 + \rho E(S_1)\Big) - (1+k)\Big(C_2 + \rho E(C_1)\Big) + k(N_2 + \rho N_1)}{\Big(A_2 + \rho E(S_1)\Big) + (k-1)(N_2 + \rho N_1) - k\Big(C_2 + \rho E(C_1)\Big)}. \tag{A.2}$$

Given the pay-offs from Chapter 3 and assuming $\rho = 0$, the following can be shown; $\check{\delta} = 0.65$ if $k = 0$, $\check{\delta} = 0.45$ if $k = 1$ and $\check{\delta} = 0$ if $k \geq 2$.

Indefinitely repeated *asymmetric* trust game without checking

The trustee's expected pay-off from playing honour:

$$\widetilde{Y_1} = (1-p_h)^k Y_1 + \sum_{t=1}^{k}(1-p_h)^{t-1}\Big(C_2 + \rho E(C_1)\Big) +$$

$$\sum_{t=2}^{k+1} p_h(1-p_h)^{t-2}(N_2 + \rho N_1)\Big(k-(t-1)\Big) + \sum_{t=2}^{k+1} p_h(1-p_h)^{t-2}\frac{(N_2 + \rho N_1)}{1-\delta}, \tag{A.3}$$

where:

$$Y_1 = \frac{\Big(C_2 + \rho E(C_1)\Big)(\delta - 1) - \delta(N_2 + \rho N_1)p_h}{(\delta - 1)\Big(\delta(p_h - 1) + 1\Big)}. \tag{A.4}$$

The trustee's expected pay-off from playing abuse:

$$\widetilde{Y_2} = (1-p_a)^k Y_2 + \sum_{t=1}^{k}(1-p_a)^{t-1}\Big(A_2 + \rho E(S_1)\Big) +$$

$$\sum_{t=2}^{k+1} p_a(1-p_a)^{t-2}(N_2 + \rho N_1)\Big(k-(t-1)\Big) + \sum_{t=2}^{k+1} p_a(1-p_a)^{t-2}\frac{(N_2 + \rho N_1)}{1-\delta}, \tag{A.5}$$

where:

$$Y_2 = \frac{\Big(A_2 + \rho E(S_1)\Big)(\delta - 1) - \delta(N_2 + \rho N_1)p_a}{(\delta - 1)\Big(\delta(p_a - 1) + 1\Big)}. \tag{A.6}$$

When $\widetilde{Y_1} > \widetilde{Y_2}$, the trustee will co-operate.

By substituting the parameters' values used in the experiment, assuming $\rho = 0$ and given $k = 4$, the smallest continuation probability for which the trustee will prefer to choose *Honour* above *Abuse* is $\delta = 0.93$. Adding certain rounds makes it more likely that the trustee will prefer to choose *Honour*. However, given the parameter values chosen for the experiment discussed in Chapter 3, the trustee's equilibrium strategy remains unaltered by adding four certain rounds ($0.83 < 0.93$).

Appendix B

Finitely repeated *asymmetric* trust game equilibrium strategy when $P_{T-1} > \tilde{P}_T$.

The idea behind randomising is that the trustee has the chance of earning an additional *Abuse* pay-off, as long as $P_{T-1} > \tilde{P}_T$. We already know that when the trustor observes an L his posterior belief will be lower than his prior belief, while when he observes an H his posterior belief will be larger than his prior belief. The trustee should randomise between choosing *Honour* and *Abuse* in such a way that the trustor is still prepared to choose *Co-operation* in T after observing an L. Hence:

$$\frac{P_{T-1}p_h}{P_{T-1}p_h + (1-P_{T-1})\Big(\lambda_{T-1}p_h + (1-\lambda_{T-1})p_a\Big)} \geq \frac{N_1 - E(S_1)}{E(C_1) - E(S_1)}, \quad \text{(B.1)}$$

or

$$\lambda_{T-1} \geq \frac{p_h\Big(E(C_1) - N_1\Big)P_{T-1} - p_a\Big(N_1 - E(S_1)\Big)(1 - P_{T-1})}{(P_h - P_a)\Big(N_1 - E(S_1)\Big)(1 - P_{T-1})}. \quad \text{(B.2)}$$

Should the trustee make the trustor indifferent, i.e. should (B.2) be an equality? If the trustor is made indifferent, he will make the trustee indifferent as well, choosing *Co-operation* with probability γ. The trustee is made indifferent when:

$$C_2 + (1-p_h)A_2 + p_h\Big(\gamma A_2 + (1-\gamma)N_2\Big) =$$

$$A_2 + (1 - p_a)A_2 + p_a\Big(\gamma A_2 + (1-\gamma)N_2\Big), \tag{B.3}$$

or when

$$\gamma^* = \frac{(A_2 - N_2)(p_a - P_h) + C_2 - A_2}{(A_2 - N_2)(p_a - p_h)}. \tag{B.4}$$

Sequential rationality dictates that the trustee will only make the trustor indifferent when he earns more from doing so. Substitution of (B.4) in (B.3) yields the trustee's expected pay-off from making the trustor indifferent, which should be larger than his pay-off from choosing *Honour*:

$$\frac{A_2(p_a - 2p_h) + C_2 p_a}{p_a - p_h} > C_2 + A_2, \tag{B.5}$$

or

$$A_2 < C_2. \tag{B.6}$$

Given condition (1.1), inequality (B.6) will never be satisfied; hence, the trustee should not make the trustor indifferent.

The trustee should pick λ_{T-1} in such a way that $P_T > \tilde{P}_T$, but only just. If $\hat{\lambda}_{T-1}$ were the mixing probability for which (B.2) would have been an equality, the trustee would choose *Honour* with $\hat{\lambda}_{T-1} + \epsilon_1$, where ϵ_1 is almost zero. The trustor will now prefer to choose *Co-operation*, regardless of his observation.

The trustor will choose *Co-operation* in $T-1$ where ϵ_2 indicates that the trustor is not indifferent in the final period:

$$P_{T-1}E(C_1) + (1 - P_{T-1})\Big(\lambda_{T-1}E(C_1) + (1 - \lambda_{T-1})E(S_1)\Big) + N_1 + \epsilon_2 \geq 2N_1. \tag{B.7}$$

Substituting λ with (B.2) and solving for P_{T-1} yields

$$P_{T-1} > \hat{P}_{T-1} \equiv$$

$$\frac{(N_1 - S_1)\Big(\Big(E(C_1) - N_1\Big)p_a + \Big(N_1 - E(S_1)\Big)p_h\Big) + \epsilon_2(p_a - p_h)}{p_h\Big(E(C_1) - E(S_1)\Big)^2}. \tag{B.8}$$

Given conditions (1.1) and (1.4), it can be easily derived that $\hat{P}_{T-1} > \tilde{P}_T$. When $P_{T-1} < \hat{P}_{T-1}$ there exists no equilibrium involving mixing on the side of the trustee. That the threshold increases makes intuitive sense, because the trustor runs a serious risk that the trustee will respond twice with *Abuse*. The trustee is better off choosing *No co-operation* in both rounds. When $P_{T-1} > \hat{P}_{T-1}$ an equilibrium exists where the trustor will choose *Co-operation* in both rounds, while the trustee will randomise over his actions in $T-1$ and choose *Abuse* in T. Whether $P_{T-1} > \hat{P}_{T-1}$ depends on the parameter values. $P_{T-1} > 1$ given the values chosen in Chapter 5. Hence, in the finitely repeated games studied in this thesis, only an equilibrium in pure strategy exists for the situation when $P_{T-1} > \tilde{P}_T$.

Appendix C

Instructions for the indefinitely repeated trust game experiment

C.1 Introduction and type game

Before the experiment starts please read the following instructions.

Welcome to this research project! Thank you for participating.

In this experiment you will encounter three different game situations. For each game you will receive a new set of instructions. The instructions for the first game are on the back of this instruction form. Before the first game is explained we will give you some basic rules, which apply to all three games.

The basic rules:

1. **The experiment is conducted anonymously, participants will only be known as a number. Neither the researchers nor other participants will know what you have decided.**

2. **It is not allowed to communicate with the other participants during the entire experiment.**

3. **In this experiment you will be asked to make decisions. In the instructions we will explain which options you will face. These options will appear on your computer screen. You will be asked to make a choice for one of the options.**

4. **If the instructions are unclear, your computer doesn't work properly or you want to ask a question, raise your hand and one of the instructors will attend to you.**

5. **During this experiment you will earn points. How many points you earn depends on the decisions you make. For every game the pay-off structure will be explained. At the end of the experiment the total number of points you earned will be translated into a monetary reward. We use the following exchange rate: 1 point = 1.51 euro cents.**

6. **When the experiment is finished, remain in your seat until you are given the signal that you can come forward to claim your reward. You can claim your reward by handing in the number given to you when entering the laboratory, together with all three complete instruction sets.**

7. **This is one session of a larger experiment. Do not spread information about the content of this study to people who will participate in further sessions of our experiment.**

8. **Please turn off your mobile phone and do not use an MP3-player.**

9. **It is not allowed to leave the experimental software unless you are told to do so.**

Game 1 (For a graphical representation, see Figure 1 on the next page.):

This game has two players: person A and person B.

Possible moves: Person A moves first and has two choices:

1. Left
2. Right

If person A decides to play Right, person B can do nothing and the game ends.

If person A decides to play Left, person B will have the following two choices:

1. Left
2. Right

After person B has made his decision, the game ends.

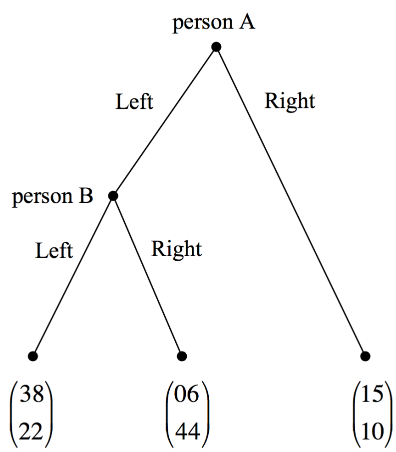
This game will be played in the following way: All participants present in the laboratory will first be given the role of person A. Next, all participants will be given the role of person B, under the assumption that person A chose Left. In other words, you are twice asked to make a decision between Left and Right once in the role of person A and once in the role of person B.

This is how we calculated your pay-off: After all participants have made both decisions, the computer will randomly assign to half of the participants the role of person A and to the other half the role of person B. To calculate the pay-offs the decision made by a person A in the first stage of this game will be randomly matched with the decision made by a person B in the second stage. In other words, you will earn points depending on the role assigned to you and the other participant you are matched with and the decisions you and the other participant made in this role. For a detailed overview of the pay-off structure, see this table:

Assume you are assigned the role of person A	**Pay-offs:**
You chose Right	You receive 15 point and person B receives 10 points
You chose Left and person B chose also Left	You receive 38 points and person B receives 22 points
You chose Left and person B chose Right	You receive 6 and person B receives 44 points
Assume you are assigned the role of person B	**Pay-offs:**
Person A chose Right	You receive 10 point and person A receives 15 points
Person A chose Left and you chose also Left	You receive 22 points and person A receives 38 points
Person A chose Left and person B chose Right	You receive 44 and person A receives 6 points

Table 1. Pay-off structure.

Figure 1. The game tree for game 1.



Pay-offs:
The upper number is person A's pay-off.
The lower number is person B's pay-off.

C.2 Trust game without checking

Before the experiment continues please read the following instructions.

Game 2 (For a graphical representation, see Figure 2.):

This game will be repeated, however, the exact number of repetitions is unknown.

This game has two players: person A and person B.

Possible moves in one round: Person A moves first and has two choices:

1. Left
2. Right

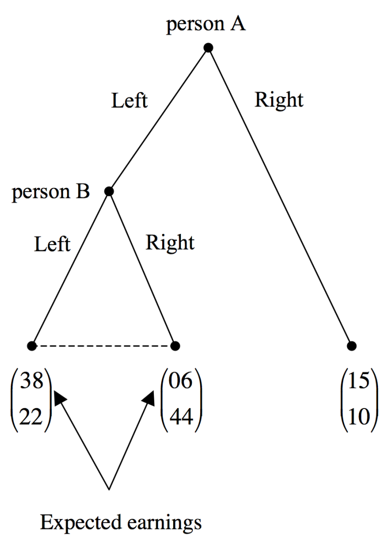
If person A decides to play Right, person B can do nothing and the round ends.

If person A decides to play Left, person B will have the following two choices:

1. Left
2. Right

After person B has made his or her decision the round ends. At the end of each round, both players will learn the number of points they have earned.

Figure 2: The game tree for game 2.



Although game 2 looks similar to game 1, there are some major differences. So read the remainder of these instructions carefully.

Duration of the game: As mentioned above, the game will be repeated. This is done in the following way: The first 5 rounds you play for certain. Beginning at the end of the fifth round the game will continue with a probability of $\frac{5}{6}$ (≈0.83). In other words, starting from round 5, the chance that the game will stop at the end of a round is 17%. The stopping chance is constant over time; this means that, regardless of how many rounds you have played, the chance that there will be a next round is still 83%.

Role and Matching: Before game 2 starts, the computer will randomly assign you the role of either person A or person B. Next, the computer will randomly match each person A to a person B. Once matched, you will play with this participant for the entire duration of the game.

The pay-off structure: If person A decides to choose Left, person B can earn either 22 points if he/she chooses Left, or 44 points when he/she chooses Right.

When person A chooses Right, he/she earns 15 points and person B earns 10 points.

Person A's pay-offs from choosing Left can either be -10 or 54 points. Person B's decision will determine which probabilities the computer will use to generate person A's pay-off. When person B chooses Left, person A will earn -10 with a probability of $\frac{1}{4}$, and 54 with a probability of $\frac{3}{4}$. If person B chooses Right, person A will earn -10 with a probability of $\frac{3}{4}$, and 54 with a probability of $\frac{1}{4}$. This implies that person A will earn an expected pay-off of 38 points when person B chooses Left, or 6 points when person B chooses Right. The probabilities and expected earnings are also given in Table 2.

At the end of every round each player will know what he/she has earned. Person A, however, will not know if person B has chosen Left or Right, because his/her pay-off does not reveal this. This is reflected by the dotted line in Figure 2.

Table 2: The pay-off structure of game 2.

Pay-offs Game 2	Person A	Person B
Left, Left	$\frac{1}{4} \times -10 + \frac{3}{4} \times 54 = 38$ points	22 points
Left, Right	$\frac{3}{4} \times -10 + \frac{1}{4} \times 54 = 6$ points	44 points
Right	15 points	10 points

A final remark: After an unknown amount of rounds the game stops. Next, we will play the game again for another unknown amount of rounds. However, before we play the game again, the following two changes will be made:

First, you will switch roles. For example, if you were person A the first time, you will be person B the second time and vice versa.

Second, you will be rematched with another participant. This means that you will play the game with a new partner. The computer is programmed in such a way that after rematching you never meet a previous partner again.

C.3 Trust game with checking

Before the experiment continues please read the following instructions.

Game 3 (For a graphical representation, see Figure 3 on the final page.):

This game will be repeated, however, the number of repetitions is unknown.

This game has two players: person A and person B.

Possible moves in one round: Person A moves first and has two choices:

1. Left
2. Right

If person A decides to play Right, person B can do nothing and the round ends.

If person A decides to play Left, person B will have the following two choices:

1. Left
2. Right

After person B has chosen either Left or Right, person A has an additional decision moment; he or she can decide to check on the behaviour of person B. Person A has two choices:

1. Check (Check on the behaviour of person B)
2. Don't check (Do not check on the behaviour of person B)

Checking is free of cost. Person B will see if person A has checked or not. After the final move of person A, the round ends. At the end of each round, both players will learn the number of points they have earned.

Again, game 3 looks similar to game 2, but please take into account the minor differences.

Duration of the game: As mentioned above, the game will be repeated. This is done in exactly the same way as in game 2.

Role and Matching: Before game 3 starts, the computer will randomly assign you the role of either person A or person B. Next, the computer will randomly match each person A to a person B. Once matched, you will play with this partner for the entire duration of the game. This is the third time you are matched to another person. Rematching is done in such a manner that you will never play with the same partner; also not a previous partner from game 2.

The pay-off structure: The pay-off structure is exactly the same as in game 2. However, note the following information difference:

At the end of each round, each player will know how many points he/she has earned. If person A did not check on the behaviour of person B, he/she will not know if person B has chosen Left or Right, because his/her pay-off does not reveal this. If person A decides to check on the behaviour of person B, he/she will also know what decision person B has made. This piece of information will appear in the lower-right corner of your screen. If person A does not check, he/she will only know his/her own pay-off. The dotted lines in Figure 3 depict the situations in game 3 when person A does not know what person B has chosen.

A final remark: After an unknown amount of rounds, the game stops. Next, we will play the game again for another unknown amount of rounds. However, before we play the game again, the following two changes will be made:

First, you will switch roles. For example, when you were person A the first time, you will be person B the second time and vice versa.

Second, you will be rematched with another participant. This means that you will play the game with someone other than you did the first time. The computer is programmed in such a way that after rematching you never meet a previous partner again.

Figure 3: The game tree for game 3.

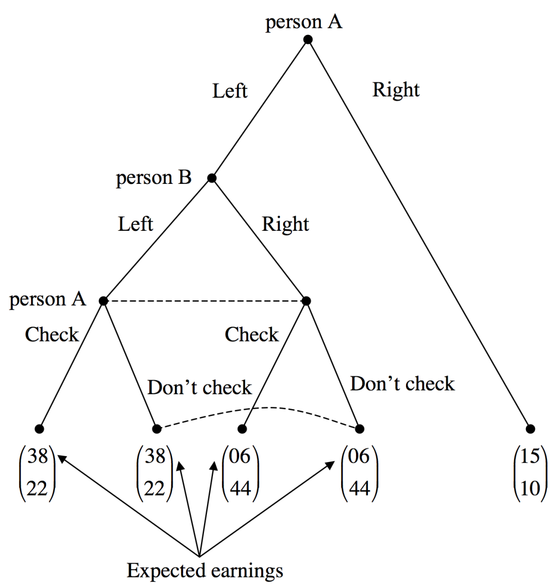


Table 3: The pay-off structure for game 3.

Pay-offs Game 3	Person A	Person B
Left, Left, Check or Don't Check	$\frac{1}{4} \times -10 + \frac{3}{4} \times 54 = 38$ points	22 points
Left, Right, Check or Don't Check	$\frac{3}{4} \times -10 + \frac{1}{4} \times 54 = 6$ points	44 points
Right	15 points	10 points

Appendix D

Instructions for the finitely repeated trust game experiment

D.1 Introduction and type game

Before the experiment starts please read the following instructions.

Welcome to this research project! Thank you for participating.

In this experiment you will encounter three different game situations. For each game you will receive a new set of instructions. The instructions for the first game are on the back of this instruction form. Before the first game is explained we will give you some basic rules, which apply to all three games.

The basic rules:

1. **The experiment is conducted anonymously, participants will only be known as a number. Neither the researchers nor other participants will know what you have decided.**

2. **It is not allowed to communicate with the other participants during the entire experiment.**

3. **In this experiment you will be asked to make decisions. In the instructions we will explain which options you will face. These options will appear on your computer screen. You will be asked to make a choice for one of the options.**

4. **If the instructions are unclear, your computer doesn't work properly or you want to ask a question, raise your hand and one of the instructors will attend to you.**

5. **During this experiment you will earn points. How many points you earn depends on the decisions you make. For every game the pay-off structure will be explained. At the end of the experiment the total number of points you earned will be translated into a monetary reward. We use the following exchange rate: 1 point = 1.35 euro cents.**

6. **When the experiment is finished, remain in your seat until you are given the signal that you can come forward to claim your reward. You can claim your reward by handing in the number given to you when entering the laboratory, together with all three complete instruction sets.**

7. **This is one session of a larger experiment. Do not spread information about the content of this study to people who will participate in further sessions of our experiment.**

8. **Please turn off your mobile phone and do not use an MP3-player.**

9. **It is not allowed to leave the experimental software unless you are told to do so.**

Game 1 (For a graphical representation, see figure 1 on the next page.):

This game has two players: person A and person B.

Possible moves: Person A moves first and has two choices:

1. Left
2. Right

If person A decides to play Right, person B can do nothing and the game ends.

If person A decides to play Left, person B will have the following two choices:

1. Left
2. Right

After person B has made his decision, the game ends.

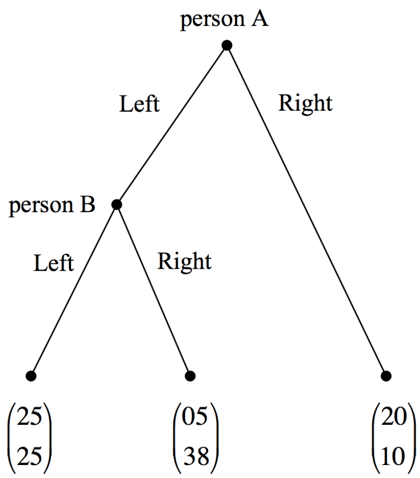
This game will be played in the following way: All participants present in the laboratory will first be given the role of person A. Next, all participants will be given the role of person B, under the assumption that person A chose Left. In other words, you are twice asked to make a decision between Left and Right once in the role of person A and once in the role of person B.

This is how we calculated your pay-off: After all participants have made both decisions, the computer will randomly assign to half of the participants the role of person A and to the other half the role of person B. To calculate the pay-offs the decision made by a person A in the first stage of this game will be randomly matched with the decision made by a person B in the second stage. In other words, you will earn points depending on the role assigned to you and the other participant you are matched with and the decisions you and the other participant made in this role. For a detailed overview of the pay-off structure, see this table:

Assume you are assigned the role of person A	**Pay-offs:**
You chose Right	You receive 20 point and person B receives 10 points
You chose Left and person B chose also Left	You receive 25 points and person B receives 25 points
You chose Left and person B chose Right	You receive 5 and person B receives 38 points
Assume you are assigned the role of person B	**Pay-offs:**
Person A chose Right	You receive 10 point and person A receives 20 points
Person A chose Left and you chose also Left	You receive 25 points and person A receives 25 points
Person A chose Left and person B chose Right	You receive 38 and person A receives 5 points

Table 1. Pay-off structure.

Figure 1. The game tree for game 1.



Pay-offs:
The upper number is person A's pay-off.
The lower number is person B's pay-off.

D.2 Trust game without checking

Before the experiment continues please read the following instructions.

Game 2 (For a graphical representation, see Figure 2.):

This game will be repeated 10 times.

This game has two players: person A and person B.

Possible moves in one round:
Person A moves first and has two choices:

1. Left
2. Right

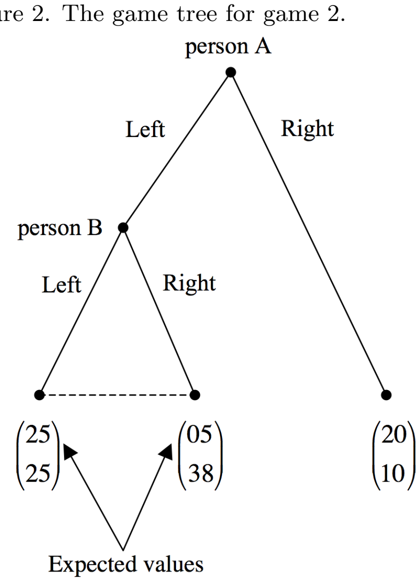
If person A decides to play Right, person B can do nothing and the round ends.

If person A decides to play Left, person B will have the following two choices:

1. Left
2. Right

After person B has made his or her decision the round ends. At the end of each round, both players will learn the number of points they have earned.

Figure 2. The game tree for game 2.



Although game 2 looks similar to game 1, there are some major differences. So read the remainder of these instructions carefully.

Duration of the game:
As mentioned above, the game will be repeated. In total you will play 10 rounds.

Role and Matching:
Before game 2 starts, the computer will randomly assign you the role of either person A or person B. Next, the computer will randomly match each person A to a person B. Once matched, you will play with this participant for the entire duration of the game.

The pay-off structure:
If person A decides to choose Left, person B can earn either 25 points if he/she chooses Left, or 38 points when he/she chooses Right.

When person A chooses Right, he/she earns 20 points and person B earns 10 points.

Person A's pay-offs from choosing Left can either be -5 or 35 points. Person B's decision will determine which probabilities the computer will use to generate person A's pay-off. When person B chooses Left, person A will earn -5 with a probability of $\frac{1}{4}$, and 35 with a probability of $\frac{3}{4}$. If person B chooses Right, person A will earn -5 with a probability of $\frac{3}{4}$, and 35 with a probability of $\frac{1}{4}$. This implies that person A will earn an expected pay-off of 25 points when person B chooses Left, or 5 points when person B chooses Right. The probabilities and expected earnings are also given in Table 2.

At the end of every round, each player will know what he/she has earned. Person A, however, will not know if person B has chosen Left or Right, because his/her pay-off does not reveal this. This is reflected by the dotted line in Figure 2.

Table 2: The pay-off structure of game 2.

Pay-offs Game 2	Person A	Person B
Left, Left	$\frac{1}{4} \times -5 + \frac{3}{4} \times 35 = 25$ points	25 points
Left, Right	$\frac{3}{4} \times -5 + \frac{1}{4} \times 35 = 5$ points	38 points
Right	20 points	10 points

A final remark:
After 10 rounds the game stops. Next, we will play the game again for another 10 rounds. However, before we play the game again, the following two changes will be made:

First, you will switch roles. For example, if you were person A the first time, you will be person B the second time and vice versa.

Second, you will be rematched with another participant. This means that you will play the game with a new partner. The computer is programmed in such a way that after rematching you never meet a previous partner again.

D.3 Trust game with checking

Before the experiment continues please read the following instructions.

Game 3 (For a graphical representation, see Figure 3 on the final page.):

This game will be repeated 10 times.

This game has two players: person A and person B.

Possible moves in one round: Person A moves first and has two choices:

1. Left
2. Right

If person A decides to play Right, person B can do nothing and the round ends.

If person A decides to play Left, person B will have the following two choices:

1. Left
2. Right

After person B has chosen either Left or Right, person A has an additional decision moment; he or she can decide to check on the behaviour of person B. Person A has two choices:

1. Check (Check on the behaviour of person B)
2. Don't check (Do not check on the behaviour of person B)

Checking is free of cost. Person B will see if person A has checked or not. After the final move of person A, the round ends. At the end of each round, both players will learn the number of points they have earned.

Again, game 3 looks similar to game 2, but please take into account the minor differences.

Duration of the game:
As mentioned above, the game will be repeated. In total you will play 10 rounds.

Role and Matching:
Before game 3 starts, the computer will randomly assign you the role of either person A or person B. Next, the computer will randomly match each person A to a person B. Once matched, you will play with this partner for the entire duration of the game. This is the third time you are matched to another person. Rematching is done in such a manner that you will never play with the same partner, also not a previous partner from game 2.

The pay-off structure:
The pay-off structure is exactly the same as in game 2, see also table 3. However, note the following information difference:

At the end of each round, each player will know how many points he/she has earned. If person A did not check on the behaviour of person B, he/she will not know if person B has chosen Left or Right, because his/her pay-off does not reveal this. If person A decides to check on the behaviour of person B, he/she will also know what decision person B has made. This piece of information will appear in the lower-right corner of your screen. If person A does not check, he/she will only know his/her own pay-off. The dotted lines in Figure 3 depict the situations in game 3 when person A does not know what person B has chosen.

A final remark:
After 10 rounds, the game stops. Next, we will play the game again for another 10 rounds. However, before we play the game again, the following two changes will be made:

First, you will switch roles. For example, when you were person A the first time, you will be person B the second time and vice versa.

Second, you will be rematched with another participant. This means that you will play the game with someone other than you did the first time. The computer is programmed in such a way that after rematching you never meet a previous partner again.

Figure 3. The game tree for game 3.

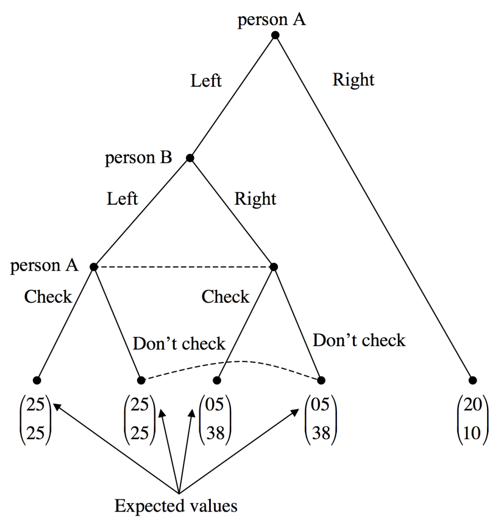


Table 3: The pay-off structure for game 3.

Pay-offs Game 3	Person A	Person B
Left, Left, Check or Don't Check	$\frac{1}{4} \times -5 + \frac{3}{4} \times 35 = 25$ points	25 points
Left, Right, Check or Don't Check	$\frac{3}{4} \times -5 + \frac{1}{4} \times 35 = 5$ points	38 points
Right	20 points	10 points

Appendix E

Questionnaire

First some questions.

What is your age?

[ ]

Are you male or female?
☐ Male ☐ Female

What is your nationality?

[ ]

How many of the participants in this room do you know by first name?

[ ]

What do you study?

[ ]

Would you consider yourself to be a religious person?
☐ Yes ☐ No

If yes, which religion:
☐ Buddhism
☐ Christianity
☐ Confucianism
☐ Hinduism
☐ Judaism
☐ Islam
☐ other

Are you a blood donor?
☐ Yes ☐ No

Have you filled in an organ donor card?
☐ Yes ☐ No

Do you do voluntary work for one or more of the following types of organisations:
☐ Student
☐ Sport
☐ Political
☐ Social
☐ Other
☐ I dont do voluntary work

Do you have a (part-time) job?
☐ Yes ☐ No

If you have a (part-time) job, how many euro do you earn an hour?
| € |

What is your net monthly income?
☐€0-€250
☐€250-€500
☐€500-€750
☐€750-€1000
☐More than €1000
☐No answer

The following two random sequences are the result of tossing a fair coin 8 times, where Heads and Tails are represented by H and T respectively.
Sequence I: HTHTTHTH
Sequence II: HHTTTTTT
Which of the two sequences is more likely to occur:
☐ Sequence I
☐ Sequence II
☐ Both sequences are equally likely to occur

You will now read a few statements. Please state to what degree you agree with them.[1]

No matter what they say, most people inwardly dislike putting themselves out to help other people.

If I know my boss will check on my work output, I would put in a higher amount of effort to make sure I would live up to his expectations.

Telling a lie can be justified, depending on the circumstances.

I don't want to act dishonestly under any circumstances.

[1] On the computer the participants could give their opinion on a 7-point scale, where 0 = completely disagree and 6 = completely agree. For dispositional reasons, the scale is not included in this appendix.

When someone has the option to check on what I am doing, I will change my behaviour, even when he might not actually check.

In this society, one does not need to be constantly afraid of being cheated.

I don't want to miss out on good opportunities while trying to be fair to others.

You will now read some more statements. Please state to what degree you agree with them.

It is easier to trust someone when you know that your relationship with this person does not end tomorrow.

Even when checking is the only way to find out if the trust you placed in someone is justified, you shouldn't check on that person because he will only behave in a less trustworthy manner after being checked.

I am trustful.

I am mindful not to forget the spirit of fair play under any circumstances.

Most people refrain from dishonest conduct to avoid getting a bad reputation.

In this society, one has to be alert, otherwise, someone is likely to take advantage of you.

When in a restaurant a collective bill has to be paid, I will make sure I don't pay more than anybody else.

Most people will respond kindly when they are trusted by others.

Please state to what degree you agree with these final statements.

When you know a relationship is about to end, there is no point in maintaining a good reputation.

In a group I am prepared to make a personal sacrifice when this is in the interest of the entire group.

People are always interested only in their own welfare.

Most people are trustworthy.

I would mind if my boss checked to see that I was doing my work well.

I am trustworthy.

When in a business relationship it is unclear to me what my business partner is doing, it is only natural for me to find this out, even when I trust him.

Samenvatting

Dit proefschrift gaat over lange termijn vertrouwensrelaties waarbij de vertrouwensgever (trustor) niet direct kan observeren wat de reactie is van de vertrouwensontvanger (trustee). Deze informatieachterstand wordt in dit proefschrift aangeduid met informatie asymmetrie. De vraag wordt opgeworpen welke invloed informatie asymmetrie heeft op het niveau van samenwerking tussen vertrouwensgever en vertrouwensontvanger. Hiernaast wordt bekeken in hoeverre een check option, waarbij de vertrouwensgever alsnog kan achterhalen wat de reactie van de vertrouwensontvanger is geweest, een positieve invloed heeft op de mate van samenwerking.

In het eerste hoofdstuk van dit proefschrift wordt de positie van dit proefschrift in de literatuur duidelijk gemaakt. Hierna wordt het vertrouwensprobleem beschreven en wordt dit probleem vertaald in een vertrouwensspel. Vervolgens wordt ingegaan op het effect van temporele inbedding, wordt het asymmetrische vertrouwensspel gepresenteerd en wordt de check-optie toegelicht. Tot slot worden de onderzoeksvragen gepresenteerd.

De vertrouwensliteratuur is zeer uitgebreid. Dit proefschrift moet worden geplaatst in de tak van de literatuur die vertrouwen op het microniveau bestudeert, waarbij interpersoonlijke vertrouwensrelaties centraal staan. Binnen deze tak van de literatuur kan een onderscheid gemaakt worden tussen studies die zich richten op de determinanten van vertrouwen en studies waar de focus ligt op het vinden van oplossingen voor het vertrouwensprobleem. Dit proefschrift vormt een toevoeging aan de laatste groep van studies.

Het vertrouwensprobleem kent een drietal elementen:

(1) De vertrouwensgever is de persoon die kan kiezen of hij met de vertrouwensontvanger gaat samenwerken of niet.

(2) De vertrouwensontvanger is de persoon die wanneer hij vertrouwen ontvangt van de vertrouwensgever moet beslissen of hij dit gaat honoreren of misbruiken.

(3) Wanneer de vertrouwensgever de vertrouwensontvanger zijn vertrouwen geeft, komt hij in een toestand van onzekerheid terecht. Vergeleken met de situatie waar hij zijn vertrouwen niet geeft, is hij slechter af wanneer zijn vertrouwen wordt misbruikt, maar is hij beter af wanneer zijn vertrouwen wordt gehonoreerd.

Het derde element is cruciaal om van vertrouwen te kunnen spreken. Het kan niet zo zijn dat de vertrouwensgever zeker weet dat zijn vertrouwen wordt gehonoreerd. Gegeven de beschikbare informatie zal de vertrouwensgever een inschatting maken over hoe waarschijnlijk het is dat de vertrouwensontvanger zijn vertrouwen zal honoreren.

De elementen van het vertrouwensprobleem worden het best weergegeven door het binaire vertrouwensspel met incomplete informatie. In dit spel kan de vertrouwensgever een eerlijke en een oneerlijke vertrouwensontvanger ontmoeten. De eerlijke vertrouwensontvanger zal gegeven vertrouwen altijd honoreren. Indien het vertrouwensspel eenmalig wordt gespeeld zal de oneerlijke vertrouwensontvanger van het geplaatste vertrouwen misbruik maken. Indien de vertrouwensgever dit anticipeert en hij de kans te klein acht dat hij een eerlijke vertrouwensontvanger ontmoet, is het voor hem beter om niet voor samenwerking te kiezen. Het vertrouwensprobleem ontstaat aangezien de vertrouwensgever en de oneerlijke vertrouwensontvanger verschillende belangen hebben. Door middel van sociale inbedding kunnen deze belangen op één lijn gebracht worden. Sociale inbedding kan worden onderverdeeld in drie subcategorieën, te weten institutionele, netwerk en temporele inbedding. De analyse die in dit proefschrift wordt gepresenteerd beperkt zich tot temporele inbedding.

Het binaire vertrouwensspel kan op twee manieren worden herhaald: Eindig en onbepaald. Indien het eindig wordt herhaald dan staat het einde van te voren vast. Wanneer het onbepaald wordt herhaald, dan weten de spelers dat het spel een einde heeft, ze weten alleen niet precies wanneer het eindigt. In dit proefschrift wordt dit verschil aangeduid met het verschil in horizon perspectief.

Door het binaire vertrouwensspel te herhalen kunnen de preferenties van de oneerlijke vertrouwensontvanger door temporele inbedding worden beïnvloed. Afhankelijk van de effectiviteit van temporele inbedding kan de vertrouwensontvanger er de voorkeur aangeven om zich coöperatief op te stellen en gegeven vertrouwen te honoreren.

Hierbij is het belangrijk om een onderscheid te maken tussen twee mechanismen: Leren en controleren. Controleren geeft het idee weer dat de vertrouwensgever de vertrouwensontvanger kan straffen door niet langer met hem samen te werken en dat de vertrouwensontvanger beseft dat zijn beslissing in het heden van invloed kan zijn op de toekomstige beslissing van de vertrouwensgever. Leren refereert aan het feit dat de vertrouwensgever zijn verwachting kan bijstellen ten aanzien van het type vertrouwensontvanger dat hij tegenover zich heeft op basis van zijn observaties van het gedrag van de vertrouwensontvanger uit het verleden en dat de vertrouwensontvanger zich hier bewust van is.

In onbepaald herhaalde vertrouwensspellen van complete informatie is controleren het dominante mechanisme dat samenwerking theoretisch kan verklaren. In eindig herhaalde vertrouwensspellen met incomplete informatie kan samenwerking theoretisch worden verklaard door het gecombineerde effect van controleren en leren. Het positieve effect van controleren en/of leren op het samenwerkingsniveau in herhaalde vertrouwensspellen wordt ook door experimentele studies ondersteund.

In dit proefschrift worden lange termijn vertrouwensrelaties bestudeerd onder asymmetrische informatie. Hiertoe wordt in hoofdstuk 1 het asymmetrische vertrouwensspel geïntroduceerd. In dit spel kan de vertrouwensgever een hoge of een lage opbrengst verdienen. De kansverdeling die gebruikt wordt om te bepalen welke opbrengst de vertrouwensgever verdient, hangt af van de reactie van de vertrouwensontvanger. Als de vertrouwensontvanger er voor kiest om het vertrouwen van de vertrouwensgever te honoreren is het meer waarschijnlijk dat de vertrouwensgever een hoge opbrengst krijgt en minder waarschijnlijk dat hij een lage opbrengst ontvangt. In het geval dat de vertrouwensontvanger er daarentegen voor kiest het vertrouwen te misbruiken is het meer waarschijnlijk dat de vertrouwensgever een lage opbrengst krijgt en minder waarschijnlijk dat hij een hoge opbrengst krijgt. De verwachte waarde van de opbrengst van de vertrouwensontvanger is gelijk aan de waarde van de opbrengst die hij in het binaire vertrouwensspel ontvangt, hiermee is het eenmalig gespeelde asymmetrische vertrouwensspel strategisch equivalent aan het eenmalig gespeelde binaire vertrouwensspel.

Het aantal studies dat asymmetrische informatie in herhaalde spellen bestudeert is zeer beperkt. Deze studies suggereren dat samenwerking wordt bemoeilijkt door de aanwezigheid van asymmetrische informatie. Onder de assumptie dat vrijwillige samenwerking tot een Pareto-verbetering leidt, is het belangrijk te achterhalen of door middel van een check-optie samenwerking kan worden gestimuleerd. De check-optie kan door de vertrouwensgever worden gebruikt nadat zowel hij als de vertrouwensontvanger hun keuze in het asymmetrische vertrouwensspel hebben gemaakt, maar voordat de opbrengsten worden bekend gemaakt. Indien de vertrouwensgever van de check-optie gebruik maakt, ontvangt hij naast zijn opbrengst ook informatie over de reactie van de vertrouwensontvanger. De vertrouwensontvanger wordt op de hoogte gesteld of de vertrouwensgever van de check-optie gebruik heeft gemaakt.

In dit proefschrift wordt er vanuit gegaan dat de check-optie gratis is. Dit betekent dat de vertrouwensontvanger het asymmetrische vertrouwensspel kosteloos kan veranderen in een symmetrisch vertrouwensspel waar hij over dezelfde informatie beschikt als in het binaire vertrouwensspel, maar waar zijn opbrengst echter nog steeds een verwachte waarde is.

Gegeven het feit dat veel economische situaties kunnen worden beschreven door herhaalde asymmetrische vertrouwensspellen met incomplete informatie, denk bijvoorbeeld aan de financiële crisis van 2008, is het belangrijk om beter inzicht te krijgen in dit type spellen. Hiertoe is er voor gekozen om in dit proefschrift onbepaald en eindig herhaalde asymmetrische vertrouwensspellen zowel theoretisch te analyseren, als experimenteel te testen met als treatment variabele de check-optie. Hiernaast begon elk experiment met een aangepaste versie van het eenmalig gespeelde binaire vertrouwensspel. Op basis van dit spel werd naast de eerlijke en oneerlijke vertrouwensontvanger ook een onderscheid gemaakt tussen de vertrouwensvolle en niet-vertrouwensvolle vertrouwensgever. Op basis van de theoretische modellen zijn de experimentele opbrengsten zo gekozen, dat indien evenwichtsstrategieën zouden worden gespeeld, de verschillende types vertrouwensgever en vertrouwensontvanger zich in hun gedrag zouden moeten onderscheiden tussen de treatments.

In dit proefschrift wordt getracht op de volgende drie onderzoeksvragen een antwoord te geven:

1. Wat is het theoretische en empirische effect van de check-optie in herhaalde asymmetrische vertrouwensspellen met incomplete informatie?

2. Kan het effect van het horizon-perspectief empirisch worden geïdentificeerd in herhaalde asymmetrische vertrouwensspellen met incomplete informatie?

3. Kunnen we theoretische voorspellingen ten aanzien van het gedrag van vertrouwensgever en vertrouwensontvanger types bevestigen in onbepaald en eindig herhaalde asymmetrische en symmetrische vertrouwensspellen met incomplete informatie?

In hoofdstuk 2 wordt de grim-trigger strategie beschreven voor zowel het onbepaald herhaalde asymmetrische als symmetrische vertrouwensspel. Een vergelijking van de modellen maakt duidelijk dat in geval van asymmetrische informatie, ten aanzien van de reactie van de vertrouwensontvanger, samenwerking moeilijker ontstaat. De voornaamste reden hiervoor ligt in het feit dat de vertrouwensgever een strengere grim-trigger strategie moet spelen in geval van asymmetrische informatie. De strengere grim-trigger strategie straft de vertrouwensontvanger niet wanneer hij er voor kiest het vertrouwen te misbruiken, maar wanneer de vertrouwensontvanger een lage opbrengst observeert. De strengere grim-trigger strategie zorgt ervoor dat minder snel aan de evenwichtscondities is voldaan. Gegeven het feit dat de opbrengsten van de vertrouwensgever geen volledig beeld geven van de reactie van de vertrouwensontvanger kan het voorkomen dat goed gedrag wordt gestraft, terwijl slecht gedrag ongestraft kan blijven. Temporele inbedding in de vorm van controleren wordt hierdoor minder effectief met als gevolg dat samenwerking minder makkelijk totstandkomt.

In hoofdstuk 3 worden de resultaten van het onbepaald herhaalde asymmetrische vertrouwensspel experiment gepresenteerd. In de treatment met de check-optie (Checking Treatment (CT)) maakten de vertrouwensgevers 86% van de tijd gebruik van deze check-optie. In deze treatment is de kans groter dat vertrouwensgevers voor samenwerking en vertrouwensontvangers voor honoreren kiezen in vergelijking tot de treatment zonder check-optie (No-Checking Treatment (NCT)). De experimentele resultaten ondersteunen bijna alle theoretische voorspellingen ten aanzien van de type indicatoren. Zoals theoretisch voorspeld kiezen vertrouwensvolle vertrouwensgevers in de NCT vaker voor samenwerking, terwijl in de CT er geen onderscheid gemaakt kan worden in het gedrag van de verschillende vertrouwensgever en vertrouwensontvanger types. Eerlijke vertrouwensontvangers kiezen daarentegen niet vaker voor honoreren in de NCT in vergelijking tot oneerlijke vertrouwensontvangers, terwijl ze dit volgens de theorie wel zouden moeten doen.

We vonden dat in de CT de opbrengstrealisaties van invloed waren op de gemaakte keuzes. Wanneer hun vertrouwen was gehonoreerd was het waarschijnlijker dat vertrouwensgevers de volgende ronde wederom voor samenwerking zouden kiezen, dit effect was echter sterker wanneer ze een hoge opbrengst hadden verdiend. Hiernaast bestaat er een kleine kans dat vertrouwensvolle vertrouwensontvangers nog steeds bereid zijn om de volgende periode voor samenwerking te kiezen wanneer ze observeren dat van hun vertrouwen misbruik is gemaakt, zolang ze maar een hoge opbrengst hebben gekregen.

In hoofdstuk 4 wordt een model voor het eindig herhaalde asymmetrische vertrouwensspel gepresenteerd. Het sequentiële evenwicht is gebaseerd op het principe dat de vertrouwensgever kan leren over het type van de vertrouwensontvanger waar hij mee speelt, op basis van de informatie die hij gedurende het verloop van het spel ontvangt. In vergelijking tot het eindig herhaalde symmetrische vertrouwensspel, is de kwaliteit van de informatie lager in geval van asymmetrische informatie. In het evenwicht moeten oneerlijke vertrouwensontvangers eerder willekeurige keuzes maken op basis van de berekende kansen, aangezien de drempel op basis waarvan de vertrouwensgever besluit of hij voor samenwerking kiest eerder stijgt. De reden hiervoor ligt in het feit dat in geval van asymmetrische informatie de vertrouwensgever een hogere garantie wil hebben dat hij met een eerlijke vertrouwensontvanger speelt. Hiernaast is het in het evenwicht vereist dat de vertrouwensgever de vertrouwensontvanger indifferent maakt. Wanneer de oneerlijke vertrouwensontvanger begint om willekeurig tussen honoreren en misbruik maken te kiezen, moet hij een hoger kans aan misbuik maken toekennen in vergelijking tot het symmetrische vertrouwensspel. Gegeven de strengere evenwichtscondities van het eindig herhaalde asymmetrische vertrouwensspel in vergelijking tot het eindig herhaalde symmetrische vertrouwensspel, is het moeilijker om samenwerking vol te houden.

In hoofdstuk 5 worden de resultaten van het eindig herhaalde asymmetrische vertrouwensspel experiment gepresenteerd. In de CT maakten vertrouwensgevers 77.54% van de tijd gebruik van de check-optie. Zowel vertrouwensgevers als vertrouwensontvangers kiezen met een grotere kans voor samenwerking in de CT in vergelijking tot de NCT. In overeenstemming met de theorie kiezen vertrouwensvolle vertrouwensgevers met een grote kans voor samenwerking in vergelijking tot niet-vertrouwensvolle vertrouwensgevers en kiezen eerlijke vertrouwensontvangers met een grotere kans voor honoreren in vergelijking tot oneerlijke vertrouwensontvangers. De theoretische aanname dat eerlijke vertrouwensontvangers gedurende het hele spel eerlijk blijven wordt door de experimentele data ondersteund. De data levert ook bewijs voor leren. Volgens de theorie moeten vertrouwensgevers hun verwachting, ten aanzien van het type vertrouwensontvanger waar ze mee spelen, op basis van Bayes' Rule aanpassen. Vertrouwensvolle vertrouwensgevers hebben een relatief hoge beginverwachting over het aandeel van eerlijke vertrouwensontvangers in de populatie ten opzichte van niet-vertrouwensvolle vertrouwensgevers. Dit betekent dat vertrouwensvolle vertrouwensgevers de kans waarmee zij voor samenwerking kiezen sterker zullen moeten verlagen, indien zij met bewijs worden geconfronteerd dat zij met een oneerlijke vertrouwensontvanger aan het spelen zijn, in vergelijking tot niet-vertrouwensvolle vertrouwensgevers. We vonden dat vertrouwensvolle vertrouwensgevers de kans waarmee zij voor samenwerking kiezen sterker verlagen in vergelijking tot niet-vertrouwensvolle vertrouwensgevers nadat zij misbruik hadden geobserveerd in de CT. Dit resultaat wordt ondersteund door de waarneming dat beide type vertrouwensgevers met een kleinere kans voor samenwerking kiezen in de latere periodes van het herhaalde spel, waarbij dit effect sterker is voor vertrouwensvolle vertrouwensgevers.

Gegeven de opzet van het asymmetrische vertrouwensspel kunnen de opbrengsten van de vertrouwensgevers variëren. De opbrengstrealisaties zijn van invloed op het gedrag van de vertrouwensgevers, zelfs wanneer de checkoptie gebruikt wordt en de reactie van de vertrouwensontvanger bekend is. Vertrouwensgevers, onafhankelijk van hun type, kiezen samenwerking met een grotere kans, wanneer ze een hoge opbrengst hebben verdiend.

In hoofdstuk 6 wordt een vergelijking gemaakt tussen eindig en onbepaald herhaalde asymmetrische vertrouwensspellen. In overeenstemming met eerdere studies, vinden we in de data geen significant verschil tussen de kansen waarmee vertrouwensgevers en vertrouwensontvangers voor samenwerking en respectievelijk honoreren kiezen in eindig en onbepaald herhaalde vertrouwensspellen, wanneer zij geen ervaring hebben met het spelen hiervan.

In beide treatments kiezen vertrouwensgevers die een lage opbrengst observeren met een kleinere kans voor samenwerking. Echter, dit effect doet zich significant sterker voor in het onbepaald herhaalde asymmetrische vertrouwensspel. Dit resultaat ondersteunt het theoretische verschil tussen controleren dat geïnitieerd wordt door een grim-trigger stategie en controleren dat geïnitieerd wordt door leren. In onbepaald herhaalde vertrouwensspellen komt controleren tot uiting in het feit dat er gestraft wordt zodra misbruik wordt geobserveerd. Leren en controleren in eindig herhaalde vertrouwensspellen stelt dat de samenwerking stopt zodra de verwachting van de vertrouwensgever dat hij een eerlijke vertrouwensontvanger tegenover zich heeft beneden de evenwichtsdrempel zakt. Dit verschil wordt ook door de volgende waarneming zichtbaar gemaakt: In het eindig herhaalde vertrouwensspel kiezen vertrouwensvolle vertrouwensgevers in eerste instantie samenwerking met een hoger kans in vergelijking tot niet-vertrouwensvolle vertrouwensgevers. Over de tijd, neemt de kans dat vertrouwensvolle vertrouwensgevers voor samenwerking kiezen sterker af, zie ook hoofdstuk 5. In het onbepaald herhaalde vertrouwensspel daarentegen bestaat er geen verschil in de kans waarmee vertrouwensvolle en niet-vertrouwensvolle vertrouwensgevers voor samenwerking kiezen.

In hoofdstuk 7 wordt een antwoord gegeven op de drie hierboven vermelde onderzoeksvragen. Zowel vanuit theoretisch als empirisch oogpunt heeft de check-optie een positief effect op het samenwerkingsniveau in eindig en onbepaald herhaalde asymmetrische vertrouwensspellen met incomplete informatie. Zoals te verwachten was hebben we geen verschil gevonden in samenwerkingsniveau tussen eindig en onbepaald herhaalde asymmetrische vertrouwensspellen. De data leverde daarentegen wel bewijs dat vertrouwensgevers verschillende strategieën spelen onder de twee horizon perspectieven. Het theoretische onderscheid tussen vertrouwensvolle en niet-vertrouwensvolle vertrouwensgevers en eerlijke en oneerlijke vertrouwensontvangers wordt grotendeels door de data ondersteund en heeft dus niet alleen theoretische relevantie.

Hiernaast wordt in Hoofdstuk 7 een aantal verbeterpunten ten aanzien van het experimentele ontwerp besproken en worden richtingen voor toekomstig onderzoek voorgesteld. Tot slot wordt kort beschreven hoe een check-optie in de praktijk zou kunnen worden vormgegeven.

Gegeven het feit dat veel dagelijkse situatie beschreven kunnen worden door herhaalde asymmetrische vertrouwenspellen met een check-optie, is het belangrijk om ons fundamentele begrip van deze spellen verder te ontwikkelen.

References

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica 56*(2), 383–396.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Alchian, A. A. and H. Demsetz (1972). Production, information costs, and economic organization. *The American Economic Review 62*(5), 777–795.

Anderhub, V., D. Engelmann, and W. Güth (2002). An experimental study of the repeated trust game with incomplete information. *Journal of Economic Behavior and Organization 48*(2), 197–216.

Aoyagi, M. and G. R. Fréchette (2008). Collusion as public monitoring becomes noisy: Experimental evidence. mimeo.

Arrow, K. J. (1974). *The limits of organization.* New York: W.W. Norton.

Barrera, D. (2005). *Trust in embedded settings.* Ph. D. thesis, Utrecht University.

Benoît, J.-P. and V. Krishna (1985). Finitely repeated games. *Econometrica 53*(4), 905–922.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. *Games and Economic Behavior 10*, 122–142.

Binmore, K. (1999). Why experiment in economics? *The Economic Journal 109*(453), F16–F24.

Bohnet, I., B. S. Frey, and S. Huck (2001). More order with less law: On contract enforcement, trust and crowding. *The American Political Science Review 95*(1), 131–144.

Bohnet, I. and S. Huck (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. *The American Economic Review 94*(2), 362–366.

Bohnet, I. and R. Zeckhauser (2004). Trust, risk and betrayal. *Journal of Economic Behavior and Organization 55*(4), 467–484.

Bolton, G. E. and A. Ockenfels (2000). ERC: A theory of equity, reciprocity, and competition. *The American Economic Review 90*(1), 166–193.

Bornhorst, F., A. Ichino, O. Kirchkamp, K. Schlag, and E. Winter (2004). How do people play a repeated trust game? experimental evidence. SFB 504 Discussion Paper 04-43.

Bornhorst, F., A. Ichino, K. Schlag, and E. Winter (2005). Trust and trustworthiness among europeans: South-north comparison. CEPR Discussion Paper No. 4378.

Bower, A. G., S. Garber, and J. C. Watson (1996). Learning about a population of agents and the evolution of trust and cooperation. *International Journal of Industrial Organization 15*(2), 165–190.

Brandts, J. and N. Figueras (2003). An exploration of reputation formation in experimental games. *Journal of Economic Behaviour and Organisation 50*(1), 89–115.

Bush, R. and F. Mosteller (1955). *Stochastic Models for Learning.* New York: John Wiley and Sons.

Buskens, V. (1999). *Social Networks and Trust.* Ph. D. thesis, Utrecht University.

Buskens, V. (2003). Trust in triads: Effects of exit, control, and learning. *Games and Economic Behavior 42*, 235–252.

Buskens, V. and W. Raub (2002). Embedded trust: Control and learning. In S. R. Thye and E. J. Lawler (Eds.), *Group Cohesion, Trust and Solidarity*, Volume 19 of *Advances in Group Processes*, pp. 167–202. JAI Press.

Buskens, V. and W. Raub (2008). Rational choice research on social dilemmas: Embeddedness effects on trust. In R. Witteck, T. A. B. Snijders, and V. Nee (Eds.), *Handbook of Rational Choice Social Research.* New York: Russell Sage. In preparation.

Camerer, C. F. (2003). *Behavioural Game Theory - Experiments in Strategic Interaction.* Princeton, New Jersey: Princeton University Press.

Camerer, C. F. and R. H. Thaler (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives 9*(2), 209–219.

Camerer, C. F. and K. Weigelt (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica 56*(1), 1–36.

Chong, J., C. Camerer, and T. Ho (2006). A learning-based model of repeated games with incomplete information. *Games and Economic Behaviour 55*(2), 340–371.

Coleman, J. S. (1988). Social capital in the creation of human capital. *The American Journal of Sociology 94*, S95–S120.

Coleman, J. S. (1990). *Foundations of Social Theory*, Chapter Relations of Trust, pp. 91–116. Cambridge, London: The Belknap Press of Harvard University Press.

Colman, A. M. (1982). *Game Theory and Experimental Games.* Oxford: Pergamon Press.

Cook, K. S., R. Hardin, and M. Levi (2005). *Cooperation Without Trust?* Russell Sage Foundation.

Coricelli, G., L. Gonzalez Morales, and A. Mahlstedt (2006). The investment game with asymmetric information. *Metroeconomica 57*(1), 13–30.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior 46*, 260–281.

Dal Bó, P. (2005). Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *The American Economic Review 95*(5), 1591–1604.

Dal Bó, P. and G. R. Fréchette (2007). The evolution of cooperation in infinitely repeated games. mimeo.

Dasgupta, P. (1988). Trust as a commodity. In D. Gambetta (Ed.), *Trust - Making and Breaking Cooperative Relations*, Oxford, pp. 51–72. Blackwell.

Deci, E. and R. Ryan (1985). *Instrinsic Motivation and Self-Determination in Human Behaviour.* New York: Plenum Press.

Dickinson, D. and M. Villeval (2005). Does monitoring decrease work effort? the complementarity between agency and crowding-out theories. Appalachian State University, Working Paper 05-12.

Duffy, J. and J. Ochs (2008). Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior DOI:10.1016/j.geb.2008.07.003*.

Engle-Warnick, J. (2007). Five indefinitely repeated games in the laboratory. CIRANO Working Papers No. 2007s-11.

Engle-Warnick, J. and R. L. Slonim (2004). The evolution of strategies in a repeated trust game. *Journal of Economic Behavior and Organization 55*(4), 553–573.

Engle-Warnick, J. and R. L. Slonim (2006). Learning to trust in indefinitely repeated games. *Games and Economic Behavior 54*(1), 95–114.

Erev, I. and A. Roth (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed-strategy equilibria. *The American Economic Review 88*(4), 848–881.

Falk, A. and M. Kosfeld (2006). The hidden cost of control. *American Economic Review 96*(5), 1611–1630.

Fehr, E. and U. Fischbacher (2003). The nature of human altruism. *Nature 425*(6960), 785–791.

Fehr, E. and S. Gächter (1998). Reciprocity and economics: The economic implications of homo reciprocans. *European Economic Review 42*, 845–859.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 144*(3), 817–868.

Fischbacher, U. (1999). z-tree – zurich toolbox for readymade economic experiments - experimenter's manual. Working Paper No. 21.

Frey, B. (1993). Does monitoring increase work effort? the rivalry with trust and loyalty. *Economic Inquiry 31*(4), 663–670.

Frey, B. (1997a). *Not Just for the Money.* Cheltenham, UK: Edward Elgar.

Frey, B. S. (1997b). A constitution for knaves crowds out civic virtues. *The Economic Journal 107*(443), 1043–1053.

Friedman, D. and S. Sunder (1994). *Experimental Methods: A Primer for Economist*. Cambridge: Cambridge University Press.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. Cambridge, MA: MIT Press.

Fukuyama, F. (1995a). Social capital and the global economy. *Foreign Affairs 74*(5), 89–103.

Fukuyama, F. (1995b). *Trust: The Social Virtues and the Creation of Prosperity*. New York: Free Press.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. New Jersey: Prentice-Hall Inc.

Giddens, A. (1990). *The consequences of modernity*. Cambridge: Polity Press.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *The American Journal of Sociology 91*(3), 481–510.

Green, E. J. and R. H. Porter (1984). Noncooperative collusion under perfect price information. *Econometrica 52*(1), 87–100.

Gulati, R. (1995). Does familiarity breed trust? the implications of repeated ties for contractual choice in alliances. *The Academy of Management Journal 38*(1), 85–112.

Güth, W., S. Huck, and P. Ockenfels (1996). Two-level ultimatum bargaining with incomplete information: An experimental study. *The Economic Journal 106*(436), 593–604.

Hardin, R. (2002). *Trust and Trustworthiness*. The Russell Sage Foundation Series on Trust. New York: Russell Sage Foundation.

Ivaldi, M., B. Jullien, P. Rey, P. Seabright, and J. Tirole (2003). *The Economics of Tacit Collusion*. Toulouse: IDEI Working Paper.

Knack, S. and P. Keefer (1997). Does social capital have an economic payoff? a cross-country investigation. *The Quarterly Journal of Economics 112*(4), 1251–1288.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Hart, Schaffner, and Marx Prize Essays, no. 31. Boston, MA: Houghton Mifflin.

Korczynski, M. (2000). The political economy of trust. *The Journal of Management Studies 37*(1), 1–21.

Kramer, R. M. (1999). Trust and distrust in organizations: emerging perspectives, enduring questions. *Annual Review of Psychology 50*, 569–598.

Kreps, D. M. (1990). Corporate culture and economic theory. In J. Alt and K. Shepsle (Eds.), *Perspectives on Positive Political Economy*, pp. 90–143. Cambridge, UK: Cambridge University Press.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory 27*, 245–252.

Kreps, D. M. and R. Wilson (1982a). Reputation and imperfect information. *Journal of Economic Theory 27*(2), 253–279.

Kreps, D. M. and R. Wilson (1982b). Sequential equilibria. *Econometrica 50*(4), 863–894.

La Porta, R., F. L. de Silanes, A. Shleifer, and R. W. Vishny (1997). Trust in large organizations. *The American Economic Review, Papers and Proceedings 87*(2), 333–338.

Lane, C. (2000). Introduction: Theories and issues in the study of trust. In C. Lane and B. R. (Eds.), *Trust Within and Between Organizations: Conceptual Issues and Empirical Applications.*, pp. 130. Oxford: Oxford University Press.

Lepper, M. and D. Greene (1978). *The Hidden Costs of Reward: New Perspectives on Psychology of Human Motivation.* Hillsdale, New York: Erlbaum.

Levi, M. and L. Stoker (2000). Political trust and trustworthiness. *Annual Review of Political Science 1*(3), 475–507.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics 1*, 593–622.

Levitt, S. D. and J. A. List (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives 21*(2), 153–174.

Loewenstein, G. (1999). Experimental economics form the vantage-point of behavioural economics. *The Economic Journal 109*(453), F25–F34.

Luhmann, N. (1979). *Trust and Power*. Chichester, UK: John Wiley & Sons.

Luhmann, N. (1988). Familiarity, confidence, trust: Problems and alternatives. In D. Gambetta (Ed.), *Trust - Making and Breaking Cooperative Relations*, pp. 94–107. Oxford: Blackwell.

Macneil, I. (1980). *The new social contract: an enquiry into modern contractual relations.* London: Yale University Press.

Mailath, G. and L. Samuelson (2006). *Repeated Games and Reputations: Long-Run Relationships.* Oxford: Oxford University Press.

McClintock, C. G. (1972). Social motivation - a set of propositions. *Behavioral Science 17*(50), 121–137.

Mitzkewitz, M. and R. Nagel (1993). Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory 22*, 171–198.

Neral, J. and J. Ochs (1992). The sequential equilibrium theory of reputation building: A further test. *Econometrica 60*(5), 1151–1169.

Nooteboom, B. (2002). *Trust; Forms, Foundations, Functions, Failures and Figures.* Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.

Parsons, T. (1951). *The social system.* London: Routledge and Kegan Paul.

Ploner, M. (2006). Trust and detection: An experimental investigation of motivational crowding out. CEEL Working Paper No. 02/2005.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature 37*(1), 7–63.

Putnam, R., R. Leonardi, and R. Y. Nanetti (1993). *Making Democracy Work.* Princeton, New Jersey: Princeton University Press.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review 83*(5), 1281–1302.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature 36*(1), 11–46.

Rasmusen, E. (2001). *Games and Information: An introduction to game theory.* Oxford: Blackwell Publishing Ltd.

Raub, W. (2004). Hostage posting as a mechanism of trust: Binding, compensation and signaling. *Rationality and Society 16*, 319–366.

Raub, W. and G. Keren (1993). Hostages as a commitment device. a game-theoretic model and an empirical test of some scenarios. *Journal of Economic Behavior and Organization 21*, 43–67.

Raub, W. and J. Weesie (1990). Reputation and efficiency in social interactions: An example of network effects. *The American Journal of Sociology 96*(3), 626–654.

Roth, A. (1994). Let's keep the con out of experimental econ.: A methodological note. *Empirical Economics 19*(2), 279–289.

Roth, A. and I. Erev (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behaviour 8*, 164–212.

Roth, A. and J. Kagel (Eds.) (1995). *Handbook of experimental economics.* Princeton, New Jersey: Princeton University Press.

Rubinstein, A. (2001). A theorist's view of experiments. *European Economic Review 45*, 615–628.

Sabel, C. F. (1993). Studied trust: Building new forms of cooperation in a volatile economy. In R. Swedberg (Ed.), *Explorations in Economic Sociology*, New York, pp. 104–144. Russell Sage Foundation.

Samuelson, L. (2005). Economic theory and experimental economics. *Journal of Economic Literature 43*, 65–107.

Selten, R. (1967). Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eimes oligopolexperiments. In H. Sauerman (Ed.), *Beiträge zur experimentellen Wirtschaftsforschung*, pp. 136–168. J.C.B. Mohr (Paul Siebeck), Tübingen.

Slonim, R., S. Helper, and H. Demaree (2006). *Context in Repeated Trust Games.* Mimeo.

Snijders, C. (1996). Trust and commitment. Amsterdam: Thela Thesis.

Snijders, C. and G. Keren (1999). Determinants of trust. In D. V. Budescu, I. Erev, and R. Zwick (Eds.), *Games and Human Behaviour*, pp. 355–385. Mahwah NJ: Lawrence Erlbaum.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Titmuss, R. (1970). *The Gift Relationship*. London: Allen and Unwin.

Uzzi, B. (1997). Social structure and competition in interfirm networks: the paradox of embeddedness. *Administrative Science Quarterly 42*(1), 35–67.

van Witteloosstuijn, A. and M. van Wegberg (2006). Trust attitudes, network tightness and organizational survival: an integrative framework and simulation model. In R. Bachmann and A. Zaheer (Eds.), *Handbook of Trust Research*, pp. 199–217. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.

Verbeek, M. (2000). *A Guide to Modern Econometrics*. Chichester, UK: John Wiley and Sons Ltd.

Williamson, O. E. (1993). Calculativeness, trust, and economic organization. *Journal of Law and Economics 36*, 453–486.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Yamagishi, T. and M. Yamagishi (1994). Trust and commitment in the united states and japan. *Motivation and Emotion 18*(2), 129–166.

Zak, P. J. and S. Knack (2001). Trust and growth. *The Economic Journal 111*(470), 295–321.

Zucker, L. G. (1986). Production of trust: Institutional sources of economic structure, 1840-1920. *Research in Organizational Behaviour 8*, 53–111.