universität wien

MAGISTERARBEIT

Titel der Diplomarbeit

„Limiting public information on delinquency.
The perspective of behavioral game theory."

Verfasser

Julian F. Richter Bakk. rer. soc. oec.

angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften
(Mag.rer.soc.oec.)

Wien, 2012

Studienkennzahl lt. Studienblatt:	A 066 913
Studienrichtung lt. Studienblatt:	Magisterstudium Volkswirtschaftslehre
Betreuerin / Betreuer	O. Univ.-Prof. Dr. Jean-Robert Tyran

1. Table of content

2. List of Illustrations

3. Acknowledgements

I would first of all like to thank Michael Eichert for the collaboration especially on the first part of the thesis and I would like to thank my supervisor Jean-Robert Tyran for the guidance, stimulating discussions and his helpful suggestions. Furthermore, I would like to thank my parents (Andreas and Adelheid Richter), Emanuel Serentschy and all my friends for their help, patience and encouragement.

4. Introduction

Cooperation in economics, as in most other human interaction, requires trust and trustworthiness to develop in the absence of complete and enforceable contracts (Cook and Cooper 2003; La Porta et al. 1997). Cooperation problems arise if there are short term incentives to exploit trust instead of cooperating (acting trustworthy). When interaction takes place more or less anonymously cooperation is especially built on mutual trust. In economics and game theory reputation formation is seen as a valid instrument for enhancing trust, trustworthiness and hence mutually beneficial cooperation (see, e.g., Milinski and Rockenbach 2006; Bohnet and Huck 2004; Resnick et al. 2006). Reputation formation is defined as collecting information about a player's past behavior / action to form a reputation profile of this person. On this basis, punishment solutions become feasible even if interaction was anonymous despite the reputation[1]. If interaction between two parties was not continuous, personal reputation formation can be a complicated issue. Therefore, reputation is often provided through an institution who can establish a history line.

Criminal records, credit registries as well as rating mechanisms in online trading facilities like eBay are prominent examples of existing reputation formation institutions. These institutions allow the parties that initiate cooperation (need to trust) to condition their decisions on this public information: their opponents past behavior. Employers get information about possible employees by consulting the criminal records. Banks and other lending institutions base their interest rate decision and the decision whether to engage with a borrower upon the specific credit register or credit reports the potential borrower holds. Also buyers in online market places like eBay consider ratings of sellers to decide with whom to engage in trade and at which price.

Information sharing in form of credit registries increases borrowers repayment rates (Brown and Zehnder 2007) and decreases informational rents banks extract in its absence (Jappelli and Pagano 2000). eBay's peer to peer rating mechanism is another example of beneficial information sharing. Buyers voluntarily pay a price premium to sellers holding a good reputation (Resnick et al. 2006). In general there is evidence for significant correlation between achieved prices and reputation (Melnik

[1] Punishment here is understood in the sense of indirect reciprocity as described by Milinski and Rockenbach (2006).

and Alm 2002). This incentivizes sellers to conform to the rules and provide the promised quality. Higher quality goods will be traded at higher prices which is beneficial for both sellers and buyers.
Evolving from the research on benefits of information sharing, economists argue that apart from its positive effects, excessive forms of information gathering might be harmful to aggregated participants' welfare. In credit markets excessively long credit histories or reports may decrease incentives to exert high effort (Vercammer 1995; Bos and Nakamura 2010). In particular, while restricting information can be beneficial depending mainly on the incentives and the borrowers' quality (Elul and Gottardi 2011), evidence on credit reports from the US suggests that removal of information leads to a decrease in efficiency (Musto 2004). Similarly, peer to peer reputation mechanisms, like the one used by eBay, potentially benefit from information restriction. For example the first bad rating a seller on eBay receives causes damage to their reputation in a way that the incentives to exert high effort afterwards are very low (Cabral and Hortacso 2004). Restricting information in such a case increases incentives and hence effort exerted.

There is also analytical and systematic literature on the effects of the length of public history on outcomes induced by the reputation mechanisms. Focusing on improvements of existing reputation mechanisms (especially eBay like mechanisms) along the lines of restricting information, interesting suggestions are, among others, a form of exponential smoothening of reputation information as past behavior carries a relatively large weight in determining the current reputation (Fan, Tan and Whinston 2005) or reputations mechanisms updating an agents profile every k periods instead of every period (Dellarocas 2006). In this thesis I want to follow these and systematically analyze effects of public information on cooperation and efficiency in a more general setting. I model the above described trust interactions with a sequential trust game. The game I use was designed by Dirk Engelmann and Jean-Robert Tyran as a part of the Ilee project to experimentally analyze the priors to trust and cooperation on the basis of socio economic data. The game is a sequential two player population game with random matching and reputation mechanism. The presumption, throughout the experiment, was that reputation formation is necessary to achieve cooperation and that different institutional settings on this reputation mechanism would lead to significantly different trust rates. Influenced by this idea on

reputation formation and the experimental and behavioral background, I analyze potential benefits and problems arising from publicly available information.

Research Questions

My primary objective in this thesis is to analyze the effects of reputation formation on cooperation and social wealth in the above mentioned sequential two-player trust game. I focus especially on two questions:

First, is public information on delinquencies sufficient to ensure cooperation? I want to analyze if the introduction of public information on delinquencies makes cooperation in form of a stable Nash Equilibrium feasible. In particular, I want to see if the Nash-Equilibrium prediction changed with respect to the public information available under standard theory assumptions.

Second, are aggregated players' payoffs affected by excessive public information if not all players are completely rational and self-interested, i.e. should public information be limited? Supposing that there is a positive effect of public information on cooperation, I want to show that if delinquencies occurred, i.e. in a bounded rationality case, there is harm from excessive information.

The Game

The underlying trust game is a finitely repeated two-player population game. The stage game is played sequentially and I refer to the ones playing first as first movers (FM) and the ones playing second as second movers (SM). The game consists of five rounds, i.e. the stage game is finitely repeated. At the beginning of each round players are matched randomly. The first movers open each round and decide whether to play *Out* or *Trust*. If a first mover chose *Out,* the round ends. If she chose *Trust,* the matched second mover decides whether to play *Betray* or *Honor* and the round ends. If there was no random matching the cooperation problem could be solved by personal reputation building and a trigger strategy by the first mover in the infinitely (Fudenberg and Tirole 1991) but also the finitely repeated game (Andreoni and Miller 1993; Kreps et al. 1982).

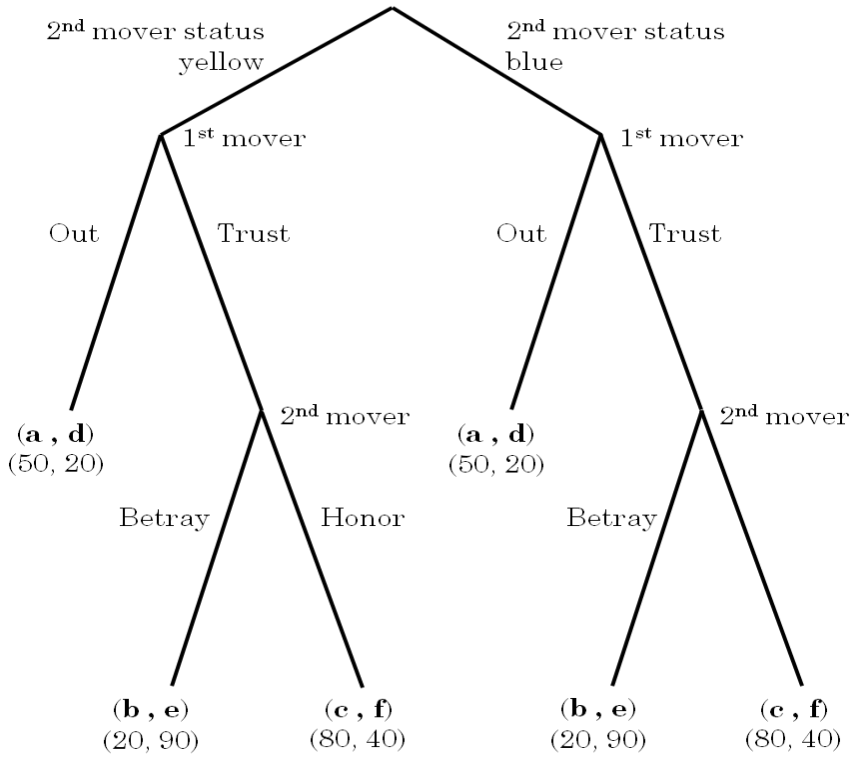

Illustration 1: The stage game with second mover reputation

Payoff structure:

(1) $c > a > b$ *or in a different notation* $(T,H)_1 > (NT,\dots)_1 > (T,B)_1$

(2) $e > f > d$ *or in a different notation* $(T,B)_2 > (T,H)_2 > (NT,\dots)_2$

(3) $(c + f) > (b + e) > (a + d)$ *equivalently* $(T,H)_{1+2} > (T,B)_{1+2} > (NT,\dots)_{1+2}$

Repeating this stage game with random matching for any number of times leads to an accumulation of one shot interactions and the non-cooperative Nash Equilibrium prediction (like in the stage game). Therefore the game is extended with two cooperation enhancing mechanisms:

First, with a binary coded reputation mechanism which provides public information on second movers past behavior despite complete random matching. This enables first movers to condition their strategy on the specific reputation of their opponents. The mechanism distinguishes between the good reputation (no delinquencies committed in the past), signalized by being *Blue* and the bad reputation (delinquencies committed in the past), signalized by being *Yellow*. A second mover holds a bad reputation if and only if they played *Betray* at least once in the last consecutive k rounds. The k, I will from now on refer to as history length limiting public information,

i.e. how far into the history of a second mover the mechanism allows to see. The reputation information is assumed to be correct as it is provided institutionally. Following the argumentation of Bolton and Ockenfels (2008), who refer to a “perfect reputation system”, the correctness assumption enables me to eliminate distortions arising from institutional failures.
Second, with a credible first mover commitment condition which limits the end-game effect. First movers have to commit to a single stage game strategy which they have to play unconditionally on the round they are in. The obligation to commit to one single strategy is crucial for the existence of a cooperative Nash-Equilibrium as it helps to overcome the backwards induction.

I specifically chose a trust game to conduct this analysis because the game represents a situation in which cooperation between first and second movers is not achievable without further enhancement. Furthermore, considering this game as a starting point, my results can be applied to several economic real world interactions such as commercial and retail banking, trade in the absence of enforceable contracts and labor markets.

Bounded Rationality

A fundamental assumption in standard economics is that all players are fully rational and self-interested and that this is common knowledge. I use an approach influenced by behavioral economics which goes beyond standard economic assumptions. I relax the assumption on rationality and selfishness by considering a share of second movers to be boundedly rational to model a more plausible and therefore imperfect world. Selten (1998) among others concluded that perfectly rational behavior is unlikely. Bounded rationality is introduced as trembles that occur with a certain, exogenous probability, along the lines of Selten (1975) and Stahl (1990). I define these trembles as mistakes (delinquencies) committed by second movers (second movers play both actions with an exogenous and positive probability without a clear underlying strategy and independent from the round they are in).

I do not oppose arguments brought in for other behavioral measures[2] influencing the outcome of the game by assuming only bounded rationality. In the context of this thesis, however, I purposely focus only on boundedly rational players as I am not trying to explain the priors to trust and cooperation. Focusing solely on boundedly rational players may lead to an underestimation of the ability for cooperation (Bolton and Ockenfels 2004). I, nevertheless, show that cooperation is still feasible and stable even if there is a share of boundedly rational second movers included.

Outline and Results

The first part of the thesis is dedicated to analyze if providing public information through a reputation mechanism, in the presence of the first mover commitment condition, is sufficient to sustain mutually beneficial cooperation as a Nash-Equilibrium. The model I (*see Model 1*) set up is based on standard economic assumptions, i.e. players are completely rational and self-interested, and this is common knowledge. I use an approach that is an adjusted game tree analysis. In a game tree all possible combinations of actions are laid out and then compared. My approach uses the rules set out in the game to conduct an analysis comparing all feasible but different second movers' combinations of *Betray*, *Honor* and *Out*, allocated over five rounds. Second movers, assuming that they know the first mover strategy and the history length, can compare their feasible combinations. Thus in particular need to payoff maximize over these possibilities. In a second step first movers, knowing the second movers' best responses (payoff maximizing choice), payoff maximize over their four different strategies[3]. The combination of the best responses defines the Nash-Equilibriums played. In the underlying trust game, I determine two different Nash-Equilibriums which are played uniquely with respect to the history length. The cooperative Nash-Equilibrium in which first movers play a trigger strategy (conditioned on the reputation) incentivizing second movers to play *Honor* as long as the threat of a bad reputation persists, i.e. up to the last round of the game. The non-cooperative Nash-Equilibrium is similar to the one round outcome where first movers have no incentive to play anything else than *Out* and the second

[2] Possible behavioral aspects influencing trust and cooperation are, among others, reciprocity (Berg, Dickhaut and McCabe 1995 ; Dufwenberg and Kirchsteiger 2004) and fairness and other trusting "characteristics" of players (Gächter, Herrmann and Thöni 2004).

[3] Each first mover commits to one of the four stage game strategies (Trust and Betray respective to the reputation).

movers never get the chance to play. The specific history length k at which equilibrium predictions switch is denoted by k*.

$$Nash - EQ^*(k, \pi_i^1, \pi_i^2)$$
$$= \begin{cases} S^1(\text{Trigger})\ and\ S^2(Honor) \quad if\ k \geq 4\ \wedge\ \ b + 4c \geq 5a \\ S^1(\text{Trigger})\ and\ S^2(Honor) \quad if\ k \leq 3 \wedge b + 4c \geq 5a \wedge (k+1)f \geq e + kd \\ S^1(\text{Out})\ and\ S^2(Betray'Invest) \quad otherwise \end{cases}$$

Therefore I conclude that in this framework public information in combination with the first mover commitment condition is a valid instrument to sustain mutually beneficial cooperation. There is a direct effect of the history length on the Equilibrium played. I derive from this that the robustness of the cooperative Nash-Equilibrium increases with history length k as incentives increase. The subsequent intuition is: the longer negative reputation prolongs, the smaller is the incentive to play *Betray* and the "easier" the cooperative Nash-Equilibrium is reached.
The payoffs achieved in the Equilibriums differ significantly. From a social perspective, sum over payoffs over all players, the non-cooperative outcome is the least best. The cooperative outcome is the best achievable outcome and pareto better than the non-cooperative one. My results from this part support the ideas of economists working on the effects of reputation in the credit market (Brown and Zehnder 2007; Jappelli and Pagano 2000) and on interactions in online marketplaces like eBay (Melnik and Alm 2002; Resnick et al. 2006). Public information, in fact, has the power to sustain cooperation.

In the second part of the thesis I focus on the question whether limiting public information has positive effects on social payoffs. For the analysis I use the same model apart from one significant difference (*see* ***Model 2***). I assume that a fraction of second movers is boundedly rational. This is modeled as probability for second movers to make mistakes (act not money maximizing) due to cognitive limitations. Delinquencies, in form of mistakes to cooperate, are essential for considerations that excessive punishment could be harmful. Without delinquencies punishment never occurs, stays a threat and therefore has no direct effect on payoffs other than incentivizing second movers.

The intuition for social payoffs to be affected negatively by excessive information is the following. If delinquencies were committed, not only the incentives, but the costs of punishment increase as punishment takes place. Hence, the aggregated players' welfare is decreasing. The question is whether it decreases solely because of delinquency or also because of the severity of the punishment. I use an approach in which I compare any two history lengths, bigger than the Nash-Equilibrium sustaining k* (from the completely rational part) to show that the respectively shorter one is generating more overall wealth. In particular, I derive conditions for which increasing history length from k'>k* to k''>k' is costly and by doing so prove that the respectively smaller history length is socially preferable. Hence, k* is the respectively shortest possible history length sustaining cooperation and therefore is the socially optimal one.

This result is only valid in the context of a stable equilibrium. I show that there exists a cooperative Nash-Equilibrium that survives bounded rationality. In particular I show that the Nash-Equilibrium derived in Model 1 is robust against some bounded rationality second movers. Under the conditions needed to sustain these equilibriums, social payoffs are still decreasing in history length, but on average first movers' would be better off from an increase in history length. Second movers' payoffs on the other hand decrease by more than the first movers' payoffs increased and overall there is a net loss from a longer history length. Again the shortest possible history length sustaining cooperation is socially optimal, even if it is not pareto better. If this Nash-Equilibrium is not leading to stable cooperation, I show that Coordination-Equilibriums sustaining cooperation, could be stable, i.e. equilibriums where first movers agree to play different strategies (*trigger* and the *always trust* strategy). If the Coordination-Equilibrium was played, limiting public information increases social payoffs and leads to a pareto better outcome. Nevertheless, the assumption that the Coordination-Equilibriums are stable is critical. One might argue its existence with experimental evidence for the willingness to punish even though it incurs substantial costs (Milinski and Rockenbach 2006).

From a theoretical point of view limiting public information is beneficial for society in cases where incentives for cooperative behavior persist for both players. Considering the controversial findings described before (Musto 2004; Bos and Nakamura 2010), I would see the differences in their results and conclusions coming from different frameworks and institutions, as well as from differently implemented restrictions on

information (removal of bad remarks after differing number of years). Also, the payoff structure used in my model might be different to the one present in the real world. In particular, $(T, B) > (NT, \dots)$ could be ranked in the opposite way which changes the conditions that need to be fulfilled for social payoffs to decrease with history length. Specifically, these conditions are not fulfilled for all populations and do not lead to the before described result of a falling payoff function. Underlying populations in the empirical studies could be different, where one fulfills the specific conditions and the other one does not. This could explain different effects of such a regulation on different populations.

Nevertheless, regarding my results I would advise a limitation / restriction of public information. This restriction on public information should be implemented by a regulatory entity as it is unclear whether self implementation would result in a social optimum. The optimal history length needs extensive assessment and might differ from population to population.

My main contribution to the literature is that I show the cooperation enhancing features of reputation formation and the positive effects of restrictions on reputation formation in one and the same trust game, assuming bounded rationality. Thereby, I give a benchmark for further experiments testing the ability to reach the socially optimal cooperation outcomes.

Concluding Remarks

The concluding implication is intuitive as most people act upon the principle described for finding the aggregated welfare maximizing solution anyway. People make mistakes and we know from personal relationships with friends, family and most notably with children that forgiveness is essential to sustain these relationships. Translated into economic interaction, in the form of this trust game, this means restricting information is essential as rationality sometimes prevents us from forgiving. But there should be a clear incentive for cooperation through the threat of potential punishment. This punishment should incentivize people to be good but should not punish them excessively. So, information restriction can have positive effects on aggregated players' welfare.

5. Equilibrium prediction under full rationality

In this part of the thesis I show that the game has a Nash-Equilibrium that supports cooperation. If there were no rational cooperation Nash-Equilibriums, the model I use to show the effects of limiting public information loses its predictive power. In particular, I suppose that there exists a Nash-Equilibrium in which first movers play the trigger strategy inducing second movers to cooperate. The idea is that first movers incentivize second movers to play *Honor* by playing a trigger strategy (punishing bad and rewarding good reputation). Punishment is understood as an incentivizing mechanism as rational and self-interested players incorporate potential losses due to punishment into their maximization problem. If avoiding punishment was less costly than being punished, there is, in fact, no punishment as second movers play accordingly. This punishment threat needs to be credible for this to be a Nash-Equilibrium, i.e. first movers need to have an incentive to punish if a second mover had a bad reputation.

5.1. Model 1

In a Nash-Equilibrium first and second movers play best response on the other players' best response respectively and therefore have no incentive to deviate from these strategies. Thusly, the equilibrium is stable. In the sequential trust game discussed here first movers play first and hence, have the power to lead the game towards their favored outcome. In my approach to find the Nash-Equilibrium I use this characteristic of the game. I consider all possible first mover strategies which are limited to the four stage game strategies due to the first mover commitment condition. In fact, if all first movers were completely rational and self-interested all strategies in which first movers do not play *Out* facing a *Blue* second mover lead to the same result as second movers start *Blue (recall the reputation mechanism: yellow only if played Betray at least once in the last rounds)*. I use a combinatorial approach to find second movers' best responses with respect to history length k and the three different first mover strategies. Second movers face the problem of finding the payoff maximizing way to allocate playing *Betray* and *Honor* over the five rounds given first movers' strategy and the history length. Then, I use the second movers' best response to find the payoff maximizing strategy for first movers with respect to history length. Throughout the entire thesis I consider an average pair of players and not

specific players, to determine the dynamics of the game, to avoid that the calculations become unnecessarily complicated.

Assumptions in Model 1:

I. Both player types, first and second movers, are rational and self-interested.
II. Players are maximizing expected payoffs.
III. All assumptions on rationality are common knowledge.
IV. The underlying game is a population game with role asymmetry, i.e. a share of the entire population is considered to be first and the rest is considered to be second movers. Furthermore the population is big enough for interaction between randomly matched players to be anonymous (and facing the same opponent twice is unlikely).

Notation:

k…history length, where $\{k \in \mathbb{N} | \ 1 \leq k \leq 5\}$

Blue…indicating a good reputation

Yellow… indicating a bad reputation

Payoffs:

– single payoffs also indicated as $\pi_{1,2}^{i}$ respectively

a…FM (Out/…) *b*…FM (Trust/Betray) *c*…FM (Trust/Honor)

d…SM (Out/…) *e*…SM (Trust/Betray) *f*…SM (Trust/Honor)

where:

(1) $c > a > b \ or \ in \ a \ different \ notation \ (T,H)_1 > (Out,\dots)_1 > (T,B)_1$

(2) $e > f > d \ or \ in \ a \ different \ notation \ (T,B)_2 > (T,H)_2 > (Out,\dots)_2$

(3) $(c + f) > (b + e) > (a + d) \ \ equivalently \ \ (T,H)_{1+2} > (T,B)_{1+2} > (Out,\dots)_{1+2}$

Strategies:

S_1 (…) ,S_2(…)…first and second mover strategy respectively

BR_1(k,S_2, $\pi_{1,2}^{i}$) and BR_2(k,S_1, $\pi_{1,2}^{i}$) …best response of the first and second mover respectively

Three different second mover (SM) actions:

B…play *Betray* if trusted

H…play *Honor* if trusted

Out…one cannot actually choose to play *Out* as it is not an action but as mentioned here the second mover could "choose" to play *Out* by playing *Betray* in (one of) the round(s) before

Four different first mover (FM) strategies:

S_1(Trust if second mover (SM) is Blue and Out if SM is Yellow) = $S_1(1,0)$
S_1(Out if second mover (SM) is Blue and Out if SM is Yellow) = $S_1(0,0)$
S_1(Trust if second mover (SM) is Blue and Trust if SM is Yellow) = $S_1(1,1)$
S_1(Out if second mover (SM) is Blue and Trust if SM is Yellow) = $S_1(0,1)$

5.1.1. Second Movers' best response

a) Suppose all first movers play S_1(Trust if SM is Blue and Out if SM is Yellow) = $S_1(1,0)$ and this is common knowledge.

For k=1 there are 6 second mover strategies which differ in their outcome. The second movers have to decide whether and how often they want to play *Betray* or *Honor* given the round they are in and the fact that playing *Betray* is followed by a bad reputation and exclusion for the next rounds as FMs will play *Out*.
In particular, SMs can decide whether to play *Betray* never, once, twice or three times and *Honor respectively*:

Second movers' payoffs from playing a specific strategy (where e, d, and f are second movers' payoffs from a specific action)

$S_2(B,Out,B,Out,B) = 3e + 2d$
$S_2(B,Out,B,Out,H) = 2e + 2d + f$
$S_2(B,Out,H,H,B) = 2e + d + 2f$
$S_2(B,Out,H,H,H) = e + d + 3f$
$S_2(H,H,H,H,B) = e + 4f$
$S_2(H,H,H,H,H) = 5f$

Combining these and the payoff structure the best responses are:

$$BR_2\left(k=1, S_1(1,0), \pi_2^i\right) = \begin{cases} S_2(B,Out,B,Out,B)\ if\ 2f < e+d \\ S_2(H,H,H,H,B)\ if\ 2f > e+d \end{cases}$$

For k=2 SMs can play *Betray* at most twice throughout the five rounds.

Second movers' payoffs from playing a specific strategy (where e, d, and f are second movers' payoffs from a specific action)

S_2(*B,Out,Out,B,Out*) = 2e +3d
S_2(*B,Out,Out,H,B*) = 2e + 2d + f
S_2(*B,Out,Out,H,H*) = e + 2d + 2f
S_2(*H,H,H,B,Out*) = e +d + 3f
S_2(*H,H,H,H,B*) = e + 4f
S_2(*H,H,H,H,H*) = 5f

Combining these and the payoff structure the best responses are:

$$BR_2\left(k=2, S_1(1,0), \pi_2^i\right) = \begin{cases} S_2(B,Out,Out,H,B)\ if\ 3f < e+2d \\ S_2(H,H,H,H,B)\ if\ 3f > e+2d \end{cases}$$

For k=3 there are again different strategies to consider where SMs can play *Betray* at most twice throughout the five rounds.

Second movers' payoffs from playing a specific strategy (where e, d, and f are second movers' payoffs from a specific action)

S_2(*B,Out,Out,Out,B*) = 2e +3d
S_2(*B,Out,Out,Out,H*) = e + 3d + f
S_2(*H,H,B,Out,Out*) = e + 2d + 2f
S_2(*H,H,H,B,Out*) = e +d + 3f
S_2(*H,H,H,H,B*) = e + 4f
S_2(*H,H,H,H,H*) = 5f

Combining these and the payoff structure the best responses are:

$$BR_2\left(k=3, S_1(1,0), \pi_2^i\right) = \begin{cases} S_2(B,Out,Out,Out,B)\ if\ 4f < e+3d \\ S_2(H,H,H,H,B)\ if\ 4f > e+3d \end{cases}$$

For k=4 there are again different strategies to consider when SMs can play *Betray* at most once throughout the five rounds. The **k=5** case is equivalent because after playing Betray once a second mover is never trusted again.

Second movers' payoffs from playing a specific strategy (where e, d, and f are second movers' payoffs from a specific action)

S_2(B,Out,Out,Out,Out) = e +4d
S_2(*H,B,Out,Out,Out*) = e + 3d + f
S_2(*H,H,B,Out,Out*) = e + 2d + 2f
S_2(*H,H,H,B,Out*) = e +d + 3f
S_2(*H,H,H,H,B*) = e + 4f
S_2(*H,H,H,H,H*) = 5f

Combining these and the payoff structure the best responses are:

$$BR_2\left(k = 4, S_1(1,0), \pi_2^i\right) = S_2(H, H, H, H, B)$$

Best response on $S_1(1,0)$

Considering the results from above, I distinguish two different strategies that are relevant for second movers supposing that $S_1(1,0)$ is played. First, the *always Honor* strategy in which a second mover plays *Honor* in each round but plays *Betray* in the last round as it does not affect their reputation anymore (no punishment threat in the last round). Second, the *Betray'Invest* strategy in which a second mover plays *Betray* whenever they are trusted but always invest in a good reputation for the last round, i.e. plays *Honor* in a way that they have a good reputation in the last round as playing *Betray* in the last round is least costly.

$$BR_2\left(S_1(1,0), k, \pi_2^i\right) = \begin{cases} S_2(Betray'Invest)\ if\ (k+1)f < e + kd \wedge\ k \leq 3 \\ S_2(Honor)\ if\ (k+1)f \geq e + kd \wedge\ k \leq 3 \\ S_2(Honor)\ if\ k \leq 4 \end{cases}$$

Using the payoffs from the experiment, this translates in the following best response function:

$$BR_2(S_1(1,0),k) = \begin{cases} S^2(\text{Betray}'\text{Invest}) \; if \; k \leq 3 \\ S^2(Honor) \; if \; k > 3 \end{cases}$$

b) Suppose all first movers play S_1(Out if SM is Blue and Out if SM is Yellow) = S_1(0,0) / S_1(Out if SM is Blue and Trust if SM is Yellow) = S_1(0,1) and this is common knowledge.

For all $k \in [1,5]$ the SMs never get the chance to play as FMs play *Out* every period of the game because SMs start *Blue*.

Best response on S_1(0,0) and S_1(0,1)

$$BR_2\left(S_1(0,0) \; or \; S_1(0,1), k, \pi_2^i\right) = Betray \; every \; round$$

c) Suppose all first movers play S_1(Trust if SM is Blue and Trust if SM is Yellow) = S_1(1,1) and this is common knowledge.

For all $k \in [1,5]$ the SMs will play the single round payoff maximizing action, i.e. SMs play *Betray* as FMs play *Trust* in every period of the game and do not condition their strategy on the reputation information.

Best response on S_1(1,1)

$$BR_2\left(S_1(1,1), k, \pi_2^i\right) = Betray \; every \; round$$

d) Best response on S_1 and k

Second movers best response function with respect to first mover strategy, history length and payoffs:

$$BR_2\left(S_1, k, \pi_2^i\right) = \begin{cases} S_2(Honor) \; if \; (k+1)f \geq e + kd \wedge \; k \leq 3 \wedge S_1(1,0) \\ S_2(Honor) \; if \; k \leq 4 \wedge S_1(1,0) \\ S_2(Betray'Invest) \; otherwise \end{cases}$$

Using the payoffs from the experiment, this translates in the following best response function:

$$BR_2\left(S_1, k, \pi_2^i\right) = \begin{cases} S_2(Honor) \; if \; k \geq 3 \wedge \; S_1(1,0) \\ S_2(Betray'Invest) \; otherwise \end{cases}$$

5.1.2. First movers' best response

To find the first movers best response for the five round game I insert the second movers' best response into the first movers' payoff function.

$$\pi_1(S_1(1,1), BR_2, k, \pi_i^1, \pi_i^2) = 5b$$

$$\pi_1\big(S_1(0,0)\ and\ S_1(0,1), BR_2, k, \pi_1^i, \pi_2^i\big) = 5a$$

$$\pi_1\big(S_1(1,0), BR_2, k, \pi_1^i, \pi_2^i\big) = \begin{cases} b + 4c\ \ldots if\ (k+1)f \geq e + kd \wedge\ k \leq 3 \\ b + 4c\ \ldots if\ k \leq 4 \\ (4-k)b + (k+1)a\ \ldots\ otherwise \end{cases}$$

Best response on BR_2 and k:

First movers best response function with respect to second mover best response, history length and payoffs:

$$BR_1\big(k, BR_2, \pi_1^i, \pi_2^i\big) = \begin{cases} S_1(1,0) \ldots if\ b + 4c \geq 5a\ \wedge (k+1)f \geq e + kd \wedge\ k \leq 3 \\ S_1(1,0) \ldots if\ b + 4c \geq 5a\ \wedge k \geq 4 \\ S_1(0,0) or\ S_1(0,1)\ otherwise \end{cases}$$

Using the payoffs from the experiment, this translates in the following best response function:

$$BR_1(BR_{\ 2}, k) = \begin{cases} S_1(1,0)\ if\ k \geq 3 \\ S_1(0,0)\ if\ k < 3 \end{cases}$$

5.1.3. The Nash-Equilibrium

Combining the two best responses gives the Nash-Equilibrium prediction with respect to single round payoffs and history length:

$$Nash - EQ^*(k, \pi_1^i, \pi_2^i)$$

$$= \begin{cases} S_1(1,0)\ and\ S_2(Honor) \quad if\ k \geq 4\ \wedge\ \ b + 4c \geq 5a \\ S_1(1,0)\ and\ S_2(Honor)\ \ if\ k \leq 3 \wedge b + 4c \geq 5a \wedge (k+1)f \geq e + kd \\ S_1(0,0)\ and\ S_2(Betray)\ \ otherwise \end{cases}$$

Using the payoffs from the experiment, this translates in the following Nash-Equilibrium:

$$Nash - EQ^*(k) = \begin{cases} S_1(1,0)\ and\ S_2(Honor) & if\ k \geq 3 \\ S_1(0,0)\ and\ S_2(Betray) & if\ k < 3 \end{cases}$$

5.1.4. Cutoff level for history length k

The non-cooperative Nash-Equilibrium is played if the history length is not sufficiently long with respect to the payoffs. The cooperative Nash-Equilibrium is played only if the history length is longer than the cutoff value k*. The cutoff level can be understood as the history length for which punishment becomes more costly than avoiding the punishment. It represents therefore the smallest history length which is cooperation sustaining, depending on the payoff structure.

The k*-function: $k^*\left(\pi_2^i, b + 4c \geq 5a\right)^4 = \min\{k | k \in \mathbb{N}\} > \frac{e+f}{f+d}$

For the payoffs from the experiment by Tyran and Engelmann the cut-off value is $k^*=3$.

Looking at the k* - function it becomes clear that the specific Nash-Equilibrium realization and the history length k are strongly interdependent. A longer history length increases the cooperative Nash-Equilibrium's robustness with respect to payoff structure. If the difference in single payoffs for second and first movers was small, the k* increases and vice versa.

5.2. Results

Given a certain history length there exists a unique Nash-Equilibrium for the five round game. In particular, there are two different Equilibriums to consider. First, the cooperative Nash-Equilibrium where FMs play "Trust if SM is Blue and Out if SM is Yellow" (the trigger strategy) and SMs play the *Honor* strategy. Second, the non-

[4] The $b + 4c \geq 5a$ condition is necessary to induce first movers to play the trigger strategy, as in this case being betrayed once (in the last round) and honored the rest of the time is better than the save Out-Option.

cooperative Nash-Equilibrium where FMs play “Out if SM is Blue and Out if SM is Yellow” and SMs play the *Betray* strategy. Which of these is played depends on the history length and the payoff structure. The two Nash-Equilibriums differ significantly in the payoffs achieved. The cooperative Nash-Equilibrium is pareto better than the non-cooperative one. This manifests the importance to choose a history length $k \geq k^*$. Cooperation is in general feasible if $4c+b \geq 5a$. The cooperative Nash-Equilibrium leaves first movers with $4c+b$ and second movers with $4f+e$ which are by assumption larger than the corresponding non-cooperative Equilibrium outcomes of $5a$ (for first movers) and $5d$ (for second movers) as can be seen from the payoff structure.
→ The cooperative Nash-Equilibriums give both players a higher payoff than the non-cooperative Nash Equilibriums.

Using the payoffs from the experiment, this translates in the following Nash-Equilibrium:
Cooperative Eq.: FMs get 340 and SMs get 250
Non-cooperative Eq.: FMs get 250 and SMs get 100

6. Effects of limiting public information on delinquencies

In this part of the thesis I provide a proof for the existence of a socially optimal level of publicly available information on delinquencies, i.e. a socially optimal history length for this game.

Following standard theory, i.e. assuming completely rational and self-interested players, the specific history length is only decisive for the realization of the Nash-Equilibriums. I defined the cut-off value, k* as the smallest history length that enables punishment. If the history length is smaller than k*, the non-cooperative Nash-Equilibrium is played. Therefore, limiting public information in the sense of reducing history length from any level k'>k* to another level k''>k' has no effect on the game's outcome. In the rational and self-interested player case a change in history length influences the outcome if and only if it crosses the level k*.

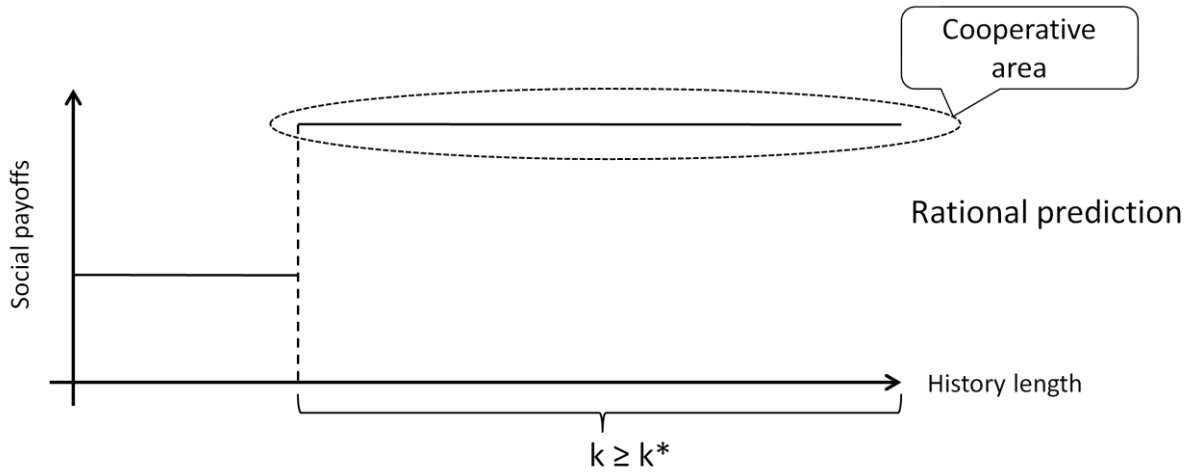

Illustration 2: Rational prediction for the social payoff function

If the two Nash-Equilibriums described in the previous part were ranked by social payoffs, the cooperation outcome is pareto better than the non-cooperation outcome. Therefore, without knowing what k* is, the longest possible history length is always at least as good as every other k. Unlimited public information, i.e. remembering delinquencies for the complete lifespan of a second mover, is the most obvious in guaranteeing the cooperative Nash-Equilibrium.

This picture changes dramatically if considering a world with relaxed assumptions on rationality. Including a chance of delinquency, i.e. allowing for mistakes, the game's outcome will change with every change in history length. This is the case because,

other than in the rational case, the punishment threat transforms into a real punishment as delinquencies occur.

My hypothesis is that punishment in the game comes at costs for both players. First movers, who carry out the punishment by playing *Out* when faced with a second mover who previously committed a delinquency, are giving up potential payoffs. Second movers who committed delinquencies get punished by receiving the smallest possible payoff through withholding. In this sense, enabling a fresh start through the limitation of information on delinquencies bears potential benefits on social payoffs.

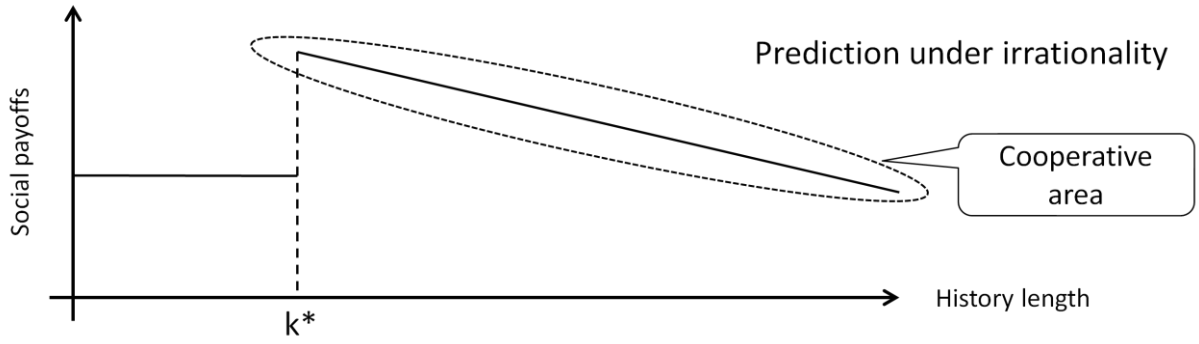

Illustration 3: Boundedly rational prediction for the social payoff function

Following the intuition, the social payoff function is falling in *history length*, if a fraction of second movers committed delinquencies (are boundedly rational). In the non-cooperative area social payoffs are independent from the *history length*. At the level k* there is a single jump in history length and cooperation becomes feasible. At k* the highest possible payoff is reached and from there it decreases in *history length*.

Therefore, I claim that the shortest possible history length enabling punishment is payoff maximizing if not all players were completely rational and self-interested, i.e. eventually forgetting delinquencies can be optimal from a social point of view.

6.1. Model 2

Here I prove a falling social payoff function. I compare any two history lengths, bigger than k*, to show that the respectively shorter one is generating more overall wealth. I particularly derive conditions for which increasing history length is costly if k>k* and by doing so prove that the k* is the respectively best history length.

Assumptions in Model 2:

I. First movers are completely rational and self-interested.
II. There are two types of second movers. First, there is a fraction of rational and self-interested players. Second, there is the fraction of boundedly rational players.
III. Bounded rationality is modeled as a probability to commit delinquencies, i.e. boundedly rational second movers play *Betray* with a certain probability and do not have a clear underlying strategy.
IV. This probability to play *Betray* is independent of stakes[5] and the specific round a player is in. It is considered as completely exogenous.
V. All assumptions on rationality and bounded rationality as well as the specific probability to play *Betray* are common knowledge.
VI. The underlying game is a population game with role asymmetry, i.e. a share of the entire population is considered to be first and the rest is considered to be second movers. Furthermore the population is big enough for interaction between randomly matched players to be anonymous (and facing the same opponent twice is unlikely).
VII. The cooperative Nash-Equilibrium exists in the completely rational case, i.e. the specific payoff structure makes cooperation in principle feasible.

Notation:

k…history length, where $\{k \in \mathbb{N} | \ 1 \leq k \leq 5\}$

k*…cutoff value for the history length (decisive for the Nash-Equilibrium realization)

s…fraction of rational and self interested SMs, where $s \in [0,1]$

x…probability to commit a delinquency, where $x \in [0,1]$

$B(S_1, S_2, k, x) = \sum_{i=1}^{5} b_i$ …sum over five rounds of the probability to be *Blue* for SMs where: $1 \geq b_i \geq 0$ and 1 is being *Blue* and 0 is being *Yellow* and $B(S_1, S_2, k, x)$ is not treated as a probability anymore, where $\sum_{i=1}^{5} b_i \in [1,5]$

Blue…indicating a good reputation

Yellow… indicating a bad reputation

[5]If the probability of mistakes was not exogenous, the calculations would become significantly more complex and need quantal response equilibrium (QRE) considerations along the lines of McKelvey and Palfrey (1998).

Payoffs:

– five round payoffs:

- $\Pi^{\text{rational}}(k)$ … sum of FM and SM average payoffs over 5 rounds if SM is rational and self-interested at history length k
- $\Pi^{\text{bounded}}(k)$ … sum of FM and SM average payoffs over 5 rounds if SM is boundedly rational at history length k
- $\Pi^{\text{social}}(k)$ … weighted average of $\Pi_{\text{k}}^{\text{rational}}$ and $\Pi_{\text{k}}^{\text{bounded}}$ at history length k

– single payoffs also indicated as π_1^i, π_2^i respectively

a…FM (Out/…) *b*…FM (Trust/Betray) *c*…FM (Trust/Honor)
d…SM (Out/…) *e*…SM (Trust/Betray) *f*…SM (Trust/Honor)

where:

(1) $c > a > b \; or \; in \; a \; different \; notation \; (T,H)_1 > (Out,\dots)_1 > (T,B)_1$

(2) $e > f > d \; or \; in \; a \; different \; notation \; (T,B)_2 > (T,H)_2 > (Out,\dots)_2$

(3) $(c + f) > (b + e) > (a + d) \; equivalently \; (T,H)_{1+2} > (T,B)_{1+2} > (Out,\dots)_{1+2}$

6.1.1. Falling payoff function

Proposition 1:

The social payoff function is decreasing in k if the critical value of k* was surpassed.

Proof:

I prove this by showing that increasing history length from any level k'>k* to a new level k'+1>k* under specific conditions leads to a decrease in social payoffs. The payoff function is of the following structure:

$\Pi^{\text{social}}(k') = s * [4(c+f) + (b+e)] + (1-s) * \big[B * \big(x(b+e) + (1-x)(c+f)\big) + 5 - B*(a+d)$

This payoff function is decoupled from the exact number of players[6] as it represents the payoff of an average first and an average second mover who play for five rounds. I rewrite the social payoff function as a weighted average of $\Pi^{\text{rational}}(k')$ and $\Pi^{\text{bounded}}(k')$:

$$\Pi^{\text{social}}(k') = s * \Pi^{\text{rational}}(k') + (1 - s) * \Pi^{\text{bounded}}(k')$$

- $\Pi^{\text{rational}}(k') = [4(c + f) + (b + e)]$.
- $\Pi^{\text{bounded}}(k') = \left[B * \left(x(b + e) + (1 - x)(c + f)\right) + (5 - B) * (a + d)\right]$

I start at a point in which cooperation equilibriums can be realized (see assumption 7) and the specific *history length* is k'. Increasing the *history length* by one round to k'+1 does not change $\Pi^{\text{rational}}(k')$. This is the case because rational and self-interested first and second movers already play best response as history length is increased from a k' already sustaining cooperation even though the part of the social payoff function that is indicated by $\Pi^{\text{bounded}}(k')$ is changing. The only factor that changes is the probability of being *Blue* for boundedly rational second movers. The $B(S_1, S_2, k, x)$ is decreasing in k (see Appendix B).

$B(S_1, S_2, k, x)$ changes by $-\Delta B \rightarrow\ 5 - B(S_1, S_2, k, x)$ changes by ΔB, where $\Delta B \geq 0$ (note that $\Delta B = 0$ only holds if the $B(S_1, S_2, k, x)$ was already at its upper boundary of 5) therefore $\Pi^{\text{bounded}}(k')$ is changing if $\Delta b \neq 0$:

rewriting $\Pi^{\text{bounded}}(k')$ in terms of changes:

$if\ k' \rightarrow k' + 1$

$$\Longrightarrow \Delta\Pi^{\text{bounded}}(k' + 1) = (-\Delta B) * \overbrace{[x * (b + e) + (1 - x) * (c + f)]}^{u} + \Delta B * \overbrace{[a + d]}^{v} =$$

$$(-\Delta B) * u + \Delta B * v$$

This means that if **u** is bigger than **v** than $\Pi^{\text{bounded}}(k')$ is decreasing if $\boldsymbol{u \geq v}$

[6] The number of players needs to be sufficiently large to guarantee anonymity through the random matching.

$\boldsymbol{u} \geq \boldsymbol{v}$ if $x * (b + e) + (1 - x) * (c + f) \geq a + d$

Due to the assumptions on the payoffs this is true. The non-cooperative single round outcome (a+d) is socially the least best and the weighted average of two better outcomes. Therefore the left side is always bigger than the right side (recall: $(c + f > b + e > a + d)$.

$\Longrightarrow \Pi_{k'+1}^{bounded} \leq \Pi_{k'}^{bounded}$.

$\Pi_k^{rational}$ stays constant if history length increased.

Therefore Proposition 1 is correct:
$\Pi^{\text{social}}(k'+1) = s * \Pi^{\text{rational}}(k'+1) + (1-s) * \Pi^{\text{bounded}}(k'+1) \leq \Pi^{\text{social}}(k') = s * \Pi^{\text{rational}}(k') + (1-s) * \Pi^{\text{bounded}}(k') \; if \; k' \geq k^* \; (Assumption \text{ VII})$ ■

6.1.2. Result

For the specific payoff structure implemented on this game the social payoff function is in fact decreasing in history length, if the presumed equilibrium with k* existed. For all payoffs that make the punishment solution feasible, and that follow the structure presumed in this game, the social payoff function is falling in k if the k was bigger or equal to k*. It is important to now focus on the existence of stable cooperation in the bounded rationality case.

6.2. Equilibrium considerations in the bounded rationality case

For the results of the payoff function to be conclusive, I have to show that there exists a state of stable cooperation if a share of second movers is boundedly rational. In Model 2, assumption 7, I claimed that punishment through the trigger strategy is feasible, and presumed that there is still a cut off value for the history length which marks the point at which the Nash-Equilibrium outcomes switch. This level I denoted as k*. In the following part I consider two options that lead to beneficial and stable cooperation sustained through an equilibrium concept.

First, I show that the Nash-Equilibrium developed in Model 1 is to some extent robust against bounded rationality of second movers. The Nash-Equilibrium prediction is still licit if bounded rationality second movers play *Betray* with a high probability but their number is small, i.e. x needs to be large and s needs to be small.

Second, I show that there are coordination equilibriums which could sustain cooperation if the cooperative Nash-Equilibrium described became impossible to reach, i.e. there are incentives (for first movers) to play $S_1(1,0)$ if x is relatively small. These coordination equilibriums are based on the assumption that first movers can coordinate themselves to play the trigger strategy, even though they have an incentive to deviate from the trigger strategy.

6.2.1. Robustness of the cooperative Nash-Equilibrium

In this extension to Model 2, I derive conditions on s, the size of the fraction of boundedly rational second movers, and x, the exogenous probability to play *Betray.*

$$Condition\ 1\text{:}\quad \frac{a-c}{b-c} < x \leq 1$$

$$Condition\ 2\text{:}\quad s(k,x) \geq \frac{5a - y(x,k)}{4c + b - y(x,k)}$$

These conditions secure the Nash-Equilibriums described in Model 1 for the set of Assumptions of Model 2. Particularly, I derive conditions guaranteeing that $S_1(1,0)$ can still be first movers' best response by inducing the rational second movers to play the *Honor* strategy. There are two possible deviation strategies for first movers. First, there is the *Out*-option($S_1(0,0)$). Second, there is the always *Trust* strategy ($S_1(1,1)$). I therefore split the analysis into two separate cases, guaranteeing that $S_1(1,0)$ is still the best response respective to the deviation strategies. The intersection of the two

set of solutions ($S_1(1,1)$ and $S_1(0,0)$ as deviation strategies) gives the set of solutions for which $S_1(1,0)$ stays best response.

a) check whether $S_1(0,0)$ becomes BR

$\Pi_1\left(S_1(1,0), k, s, x, \pi_1^i\right) = s * \Pi_1^{rational}(k) + (1-s) * \Pi_1^{bounded}(k) \geq$
$\Pi_1(S_1(0,0), k, s, x, \pi_1^i)$

$\rightarrow Equation\ 1: s * [4c + b] + (1-s) * \left[y(x, k, \pi_1^i)\right] \geq [5a]$

Where, $y\left(x, k, \pi_1^i\right) = B(k) * (xb + (1-x)c) + (5 - B(k))a$ is the expected payoff for a first mover from playing a boundedly rational second mover.

Case 1:

Equation 1 is true if both factors on the left side were bigger than the factor on the right side.

$[4c + b] \geq [5a]\ and\ y(x, k) \geq 5a$ needs to hold.

From the calculations in Model 1 I recall that $[4c + b] \geq [5a]$ is a necessary condition for the Nash-Equilibrium to be feasible and can be considered to be true (otherwise cooperation was in general not feasible).

$y(x, k) \geq 5a$ is the case if $xb + (1-x)c \geq a \rightarrow$ hence $x \leq \frac{a-c}{b-c}$

$\rightarrow$ hence $\text{Eq. } 1: s * [4c + b] + (1-s) * \left[y(x, k, \pi_1^i)\right] \geq [5a]$ is fulfilled if:

$x \leq \frac{a-c}{b-c}$ and 2.

$[4c + b] \geq [5a]$.

Case 2:

The case where $\frac{a-c}{b-c} < x \leq 1$ and $[4c + b] \geq [5a]$. From above I know that in this case $y(x, k) \leq 5a$.

$\rightarrow$ Hence, there needs to be a condition on s(k,x) that makes sure that depending on π_1^i and the history length fulfills Equation 1:

$$s(k, x) \geq \frac{5a - y(x, k)}{4c + b - y(x, k)}\ \forall\ \frac{a-c}{b-c} \leq x \leq 1$$

Combining cases 1 and 2 leads to the result that $S_1(1,0)$ is better than $S_1(0,0)$ if the following condition was fulfilled:

$$\Rightarrow \begin{cases} 0 \leq x \leq \frac{a-c}{b-c} \wedge [4c+b] \geq [5a] \\ or \\ \frac{a-c}{b-c} < x \leq 1 \wedge s(k,x) \geq \frac{5a-y(x,k)}{4c+b-y(x,k)} \wedge [4c+b] \geq [5a] \end{cases}$$

For the payoffs used in the experiment the problem reduces to $s * [340] + (1-s) * \left[y(x,k,\pi_1^i)\right] \geq [250]$ and $S_1(1,0)$ is better than $S_1(0,0)$ if the following condition was fulfilled:

$$\Rightarrow \begin{cases} 0 \leq x \leq 0{,}5 \\ or \\ 0{,}5 \leq x \leq 1 \wedge s(k,x) \geq \frac{250-y(x,k)}{340-y(x,k)} \end{cases}$$

b) check whether $S_1(1,1)$ becomes BR

$S_1(1,1)$ becomes BR for the first movers if the expected payoff that a first mover received from a boundedly rational second mover is better than the payoff they receive from playing *Out.*

As $+(1-x)c \geq 5a \ \forall\, 0 \leq x \leq \frac{a-c}{b-c}$, the probability x to play *Betray* has to be $\frac{a-c}{b-c} < x \leq 1$ for $S_1(1,0)$ to be better than $S_1(1,1)$.

Using the payoffs from the experiment, this translates in the following condition for $S_1(1,1)$ becomes BR:

$x20 + (1-x)80 \leq 50 \rightarrow hence\ if\ 0{,}5 < x \leq 1$

6.2.2. Nash-Equilibrium

Combining a) and b) the cooperative Nash-Equilibrium still exists if and only if:

$$Condition\ 1: \ \frac{a-c}{b-c} < x \leq 1$$

$$Condition\ 2: \ s(k,x) \geq \frac{5a-y(x,k)}{4c+b-y(x,k)}$$

$$\rightarrow Nash - EQ^*(\pi_1^i, \pi_2^i, b + 4c \geq 5a, k \geq k^*, x, s)$$

$$= \begin{cases} S_1(1,0)\ and\ S_2(Honor)\ if\ \dfrac{a-c}{b-c} < x \leq 1 \wedge s(k,x) \geq \dfrac{5a - y(x,k)}{4c + b - y(x,k)} \\ S_1(0,0)\ and\ S_2(Betray'Invest) \qquad otherwise \end{cases}$$

Using the payoffs from the experiment, the cooperative Nash-Equilibrium can be sustained if and only if:

$$Condition\ 1\text{:}\ \ 0{,}5 \leq x \leq 1$$

$$Condition\ 2\text{:}\ \ s(k,x) \geq \frac{250 - B(k) * (80 - 60x) + \big(5 - B(k)\big) * 50}{340 - B(k) * (80 - 60x) + \big(5 - B(k)\big) * 50}$$

$$\rightarrow Nash - EQ^*(\ k \geq k^*, x, s)$$

$$= \begin{cases} S_1(1,0)\ and\ S_2(Honor)\ if\ 0{,}5 < x \leq 1 \wedge s(k,x) \geq \dfrac{250 - B(k) * (80 - 60x) + \big(5 - B(k)\big) * 50}{340 - B(k) * (80 - 60x) + \big(5 - B(k)\big) * 50} \\ S_1(0,0)\ and\ S_2(Betray'Invest) \qquad otherwise \end{cases}$$

6.2.3. Coordination Equilibriums

Here I look at the case where first movers have incentives to play $S_1(1,1)$ which eventually drives rational and self-interested second movers to play *Betray* in each round. By this presumption on coordination, second movers' probability to play *Betray* has to be: $0 \leq x \leq \frac{a-c}{b-c}$. This in return induces rational and self-interested (by Assumption I all first movers) to play $S_1(0,0)$ and cooperation breaks down.

The coordination Equilibrium I introduce here, under strong assumptions on the ability of first movers to coordinate themselves, has the power to sustain stable cooperation. The intuition is that, knowing of the dynamic that evolves, the population of first movers could agree to play a combination of $S_1(1,0)$ and $S_1(1,1)$. First movers could agree on a fraction that plays $S_1(1,1)$ and a fraction that plays $S_1(1,0)$. Depending on the history length first movers have to choose a level of deviation towards $S_1(1,1)$ that still induces rational and self-interested second movers to play cooperatively. This approach needs coordination as the fraction playing $S_1(1,1)$ is better off than the fraction playing $S_1(1,0)$. This persists as long as a sufficient

amount of first movers continues to play $S_1(1,0)$ but every first mover has incentives to deviate to $S_1(1,1)$.

Nevertheless, I suppose in this context that such a form of coordination is possible and derive the cut-off values for the fraction sizes that define the coordination Equilibriums possible. There is, in fact, experimental evidence supporting this assumption. Punishment, even though costly and irrational takes place if there are also social incentives involved (Milinski and Rockenbach 2006 Fehr, Fischbacher and Gächter 2002). Costly punishment solutions are frequently chosen in the public good game (see, e.g., Fehr and Gächter 2002). This could also be the case in this game, especially if being rational might not be that rational after all, as rationality leads to the non cooperative, hence, the least best outcome. It is important to note that the payoffs have to enable punishment solution, i.e. there needs to be a k* for which the cost of punishment exceed the costs of avoiding punishment for second movers (Assumption VII).

I compare relevant second movers' strategies to the cooperative Nash-Equilibrium strategy to determine at which level of first mover deviation (towards $S_1(1,1)$) the second movers' best response switches. In particular, I show at which level of deviation second movers switch from Honor to the next best strategy depending on the history length k.

Notation:

t…share of FM playing $S_1(1,1)$ instead of $S_1(1,0)$

(1-t)…share of FM playing $S_1(1,0)$

$t^*(S_1,S_2,\ldots)$…cut-off level

Given the FMs' strategies, there are two relevant deviation strategies for SMs to be considered:

S_2(Betray)… play *Betray* whenever they are trusted

S_2(Betray'invest)… play *Betray* whenever they are trusted but always invest in a good reputation for the last round (play honor to guarantee being blue in the last round)

(it is important to note that S_2(Invest) is in the case where first movers are completely rational and self-interested and playing $S_1(1,0)$ equivalent to S_2(Betray) therefore there are only two realistic strategies to consider)

S_2(Invest)… play *Betray* whenever they are trusted and immediately afterwards invest in a good reputation again

Deviation Strategies

For k≥4 *Betray* is the relevant SM strategy (incentives to play *Betray* increase with t) while *Betray'invest* due to the specific k is equal to the *Honor* strategy

Find t for which $\Pi_2(\text{Betray}) \leq \Pi_2(\text{Honor})$

$$e + 4(te + (1-t)d) \leq 4f + e$$

$$\rightarrow t \leq \frac{f-d}{e-d}$$

For k=3 and depending on the specific parameters in the game (single round payoffs) one of the two strategies is the "closest" to *Honor*:

$$BR_2\left((1-t)(1,0), t(1,1), k=3, \pi_2^i, t=t^*\right)$$

$$= \begin{cases} Betray'invest & if\ t^{Betray} \geq t^{Betray'invest} \\ Betray & if\ t^{Betray} \leq t^{Betray'invest} \end{cases}$$

or

$$\mathrm{t}^* = \min\{\mathrm{t}^{\mathrm{Betray}}, \mathrm{t}^{\mathrm{Betray'Invest}}\}$$

Find t for which $\Pi_2(\text{Betray}) \leq \Pi_2(\text{Honor})$

$$e * (1 + (1-t)^3) + (5 - (1 + (1-t)^3) * (te + (1-t)d) \leq 4f + e$$

$$\rightarrow 6t^2 - 4t^3 + t^4 \leq \frac{4f - e - 3d}{e - d}$$

Find t for which $\Pi^{\mathrm{SM}}(\text{Betray}'\text{invest}) \leq \Pi^{\mathrm{SM}}(\text{Honor})$

$$2e + 3(te + (1-t)d) \leq 4f + e$$

$$\rightarrow t \leq \frac{4f - e - 3d}{3(f-d)}$$

For k=2 and depending on the specific parameters in the game (single round payoffs) one of the two strategies is the "closest" to *Honor*:

$$BR_2\left((1-t)(1,0), t(1,1), k=2, \pi_2^i, t=t^*\right)$$
$$= \begin{cases} Betray'invest & if\ t^{Betray} \geq t^{Betray'invest} \\ Betray & if\ t^{Betray} \leq t^{Betray'invest} \end{cases}$$

or

$$t^* = \min\{t^{\text{Betray}}, t^{\text{Betray'Invest}}\}$$

Find t for which $\Pi_2(\text{Betray'invest}) \leq \Pi_2(\text{Honor})$

$$e + 2(te + (1-t)d) + ((1-t)^2 + t(1-t)^2) * e + (2 - ((1-t)^2 + t(1-t)^2)) * (te + (1-t)d) \leq 4f + e$$
$$\rightarrow 2t^2 + 2t^3 - t^4 \leq \frac{4f - e - 3d}{(e-d)}$$

Find t for which $\Pi_2(\text{Betray}) \leq \Pi_2(\text{Honor})$

$$2e + (te + (1-t)d) + (tf + (1-t)d) + (1-t)f + t(tf + (1-t)d) \leq 4f + e$$
$$\rightarrow t(e-d) + t^2(f-d) \leq 3f - e - 2d$$

6.2.4. <u>Coordination Equilibrium is characterized as following:</u>

A fraction t of first movers plays $S_1(1,1)$ and the fraction (1-t) plays $S_1(1,0)$. Second movers play the *Honor* strategy if they are rational and self-interested (a fraction s of all second movers). They play *Betray* with probability x and *Honor* with probability (1-x) if they are boundedly rational (a fraction (1-s) of all second movers). The exact number of such feasible coordination equilibriums is not particularly important but in principle every t ≤ t* describes a possible coordination equilibrium. The t* describes the smallest t that makes a deviation strategy for second movers better than the *Honor* strategy given the history length.

$$Coordination\ EQ^*\left(t, k, s, x, S_1, S_2, \pi_1^i\right)$$
$$= \begin{matrix} \boldsymbol{FM}\ \{\boldsymbol{t} * \boldsymbol{S_1}(\mathbf{1},\mathbf{1}) + (\mathbf{1}-\boldsymbol{t})\boldsymbol{S_1}(\mathbf{1},\mathbf{0}) \\ \boldsymbol{SM}\{\boldsymbol{s} * \boldsymbol{S_2}(\boldsymbol{Honor}) + (\mathbf{1}-\boldsymbol{s}) * \boldsymbol{S_2}(\boldsymbol{xB} + (\mathbf{1}-\boldsymbol{x})\boldsymbol{H}) \end{matrix} \quad if\ \boldsymbol{s} > 0\ \wedge\ \boldsymbol{t} \leq \boldsymbol{t}^*$$

7. Discussion

What was the objective?

My primary objective in this thesis was to analyze the effects of reputation formation, i.e. public information on delinquencies, on cooperation and social wealth in a sequential two-player trust game. I specifically focused on two questions. First, is reputation formation a valid institution to ensure cooperation? Second, are aggregated players' payoffs affected by excessive public information in a bounded rationality case, i.e. should public information be limited?

Results

In Model 1 I showed that there exists cooperation as a Nash-Equilibrium if the history length was long enough and all players were assumed to be rational and self-interested. Social payoffs in this cooperative Nash-Equilibrium are higher than in the non cooperative Nash-Equilibrium that is played if the history length was not long enough. Therefore, the reputation mechanism in the completely rational case is a valid institution to sustain cooperation.

In Model 2 I showed that there are two possible ways in which limiting public information on delinquencies, i.e. restricting history length, leads to a stable and socially better outcome than unrestricted information. First, a cooperative "Nash-Equilibrium" exists, even though a share of second movers is boundedly rational. And while second movers' and overall wealth increases, first movers' wealth on average decreases compared to the unrestricted case. For this Nash-Equilibrium first movers have to have incentives to play *Out* facing a boundedly rational second mover. Second, if first movers were assumed to be able to find a coordination equilibrium agreement, there exist several coordination equilibriums in which a fraction of first movers still plays the trigger ($S_1(1,0)$) and the rest a forgiving strategy ($S_1(1,1)$). Given that one of these coordination equilibriums is played, first and second movers' wealth increases compared to the unrestricted case.

Therefore, I conclude that players' payoffs are affected by excessive public information in a bounded rationality case, i.e. public information should be restricted if one of the above mentioned cases occurs.

7.1. Critiques that might limit the generalizability of my results

The models I use in this thesis are based on a sequential trust game that stylizes interactions between two parties where, due to anonymity, cooperation problems arise. Here, and throughout the thesis, I presumed that cooperation is beneficial for society and the parties involved. In the following I describe potential problems and critiques that might arise if a general concept on the ability and the setup of reputation formation was derived from the game's particulars.

First, the reputation mechanism is of the simplest form. It creates a black and white world in which reputation can be either good or bad. The information available only indicates that a bad reputation player committed at least one delinquency in one of the previous k periods. The reputation mechanism therefore omits a great deal of information, such as the time of occurrence of the delinquency, how often a player committed delinquencies, and all actions that happened before the previous k periods.

However, I show that even this simple form of reputation is sufficient to give first movers the tool to induce second movers by punishment in a binary choice setup if reputation was implemented accurately. Whether more detailed reputation information would be more efficient needs further consideration. Especially, if the situation was not limited to a binary choice problem but had a more diversified choice set. In such a case first movers could condition the terms of their engagement not only on the reputation information, but also on the time the delinquency was committed and its frequency. This could lead to inefficiencies. First movers could overcharge (if granting loans) or squeeze prices under their actual value (in an online market) depending on the situation, instead of granting second chances. This again might lead to inefficiencies and might make the restriction of information (forgetting) even more beneficial following the same argumentation used in this thesis.

Second, the commitment condition on first movers' strategies which was revealed as essential for rational cooperation to occur, is critical. If it was self-induced by players there is a credibility issue as it prevents them from playing best response in each period.

If the game was not finite or had no pre-specified end (the exact time of end was unknown) which in the real world is likely to be the case, this commitment condition is

not necessary to sustain cooperation. The end game effect is cancelled out and therefore cooperation would be even easier to achieve. Furthermore, the condition could be seen as a company policy ensured by a contractual agreement and therefore not completely random.

Third, the assumption that deviation from best response and the actual deviation behavior in the bounded rational case are common knowledge is highly unlikely.
I use it to be able to show the dynamics of the game if the assumption held. Nevertheless, I draw conclusions from this as it indicates what happens if collective beliefs were of a certain structure.

Fourth, to assume that bounded rationality is modeled as trembles is critical but there is literature supporting my choice to model delinquencies that way (Selten 1975; Stahl 1990).
If enough of the first movers play the cooperation inducing strategy ($S_1(1,0)$), it is clearly a mistake for second movers to play *Betray, as* they ultimately will be worse off if the history length is longer than k*. Therefore I assume that not all second movers have the ability to fully understand the game and play best response in every period. This has been an easy way of modeling bounded rationality. Different ways of modeling bounded rationality could be interesting for further research and lead to significantly different results.

Fifth, the assumption that first movers have the chance to coordinate themselves into a, for them favorable, coordination equilibrium if they had incentives to be forgiving, i.e. to play Trust unconditionally, can be backed by behavioral economics.
Coordination of such a kind requires that, depending on the likeliness of delinquency, a share of first movers plays the trigger strategy knowing this is not best response. This is without doubt hard to achieve because a free riding problem occurs. Every first mover has an incentive to act forgiving no matter what the agreement was if the likeliness of delinquencies was low and first movers got on average more from forgiving than from carrying out the punishment. Still, even if it is unlikely, such coordination could be a matter of a contract between first movers and therefore feasible. Furthermore, evidence suggests that strong reciprocity (voluntarily cooperate if treated fairly but punish non-cooperators) exists even if it was costly

(Fehr and Gächter 2002; Fehr, Fischbacher and Gächter 2002; Milinski and Rockenbach 2006). This could be a possible behavioral economics explanation for coordination to exist, as some first movers might choose to play the trigger strategy even though it is costly.

Sixth, the payoff structure that I impose on the game ranks social payoffs in the stage game as following:

$(c + f) > (b + e) > (a + d)\ equivalently\ (T,H)_{1+2} > (T,B)_{1+2} > (NT,\dots)_{1+2}$

This might lead to the critique that reducing history length was only beneficial because of the fact that from a social perspective the Out-Option (NT,...) is the least best. Therefore even if all second movers played Betray in each round that they got to play earlier due to a restriction on information, social payoffs improve. The critique could be that my results only work because of that specific structure.

I can show that this is not the case. Changing the structure accordingly, I can reproduce my results.

$$(c + f) > (a + d) > (b + e)\ equivalently\ (T,H)_{1+2} > (NT,\dots)_{1+2} > (T,B)_{1+2}$$

The conditions on the parameters of the game for which a restriction of public information is beneficial change, but the general conclusion stays the same:

$$Condition\ 1a\text{:}\ \frac{a+d-c-f}{b+e-c-f} \geq x \geq \frac{a-c}{b-c}$$

$$Condition\ 2a\text{:}\ s(k,x) \geq \frac{5a-y(x,k)}{4c+b-y(x,k)}$$

$$\rightarrow Nash-EQ^{*}(\pi_1^i, \pi_2^i, b+4c \geq 5a, k \geq k^{*}, x, s)$$

$$= \begin{cases} S_1(1{,}0)\ and\ S_2(Honor)\ if\ \frac{a+d-c-f}{b+e-c-f} \geq x \geq \frac{a-c}{b-c} \wedge\ s(k,x) \geq \frac{5a-y(x,k)}{4c+b-y(x,k)} \\ S_1(0{,}0)\ and\ S_2(Betray'Invest) \qquad otherwise \end{cases}$$

The payoff function is falling in history length and the k* is still socially favorable in the respective area. For a detailed analysis see Appendix A.

7.2. Concluding Remarks

Based on my results and also incorporating potential drawbacks of the models I would conclude that public information in the form of reputation has positive effects on real world societies. If information about delinquencies was available long enough to punish the delinquents in a way that their costs from punishment (exclusion) are higher than the gain from betraying, cooperation is beneficial for both parties and the society. In particular, if there was no information available there is space for the market to set up such an institution “privately”. Assuming that an institution providing public information is costly there are different possibilities to finance it. I consider that confiders, the agents who base their trust on information about their counterparts, as the most likely case financiers of such an institution. The aggregate of all confiders has all the information necessary to create complete profiles about their counterparts past behavior. Therefore they could set up an institution that collects all information and then distributes the right profile to the right confider. An institution financed by possible delinquents is hardly plausible as after committing a delinquency there exist strong incentives not to report this truthfully. Thus such an institution would have a credibility problem.

Taking this into consideration the confiders could influence or decide on the amount of information available. Considering my results from Model 2, this is not necessarily the optimal amount of information from a social perspective. Depending on the expected likeliness of delinquencies, confiders would either decide for the longest possible history length (marking delinquents for the rest of their lives) or the shortest possible history length that is still inducing cooperative behavior. This is the case if the information was coded binary (as in the reputation mechanism used in the game) and no further information was available. With a binary coded reputation confiders cannot distinguish between the history length (how long a delinquent holds the bad reputation) and the punishment. If the information was more detailed and the entire history of an opponent was known, confiders would always want the entire information available. Confiders could decouple punishment from reputation and decide independently for how long delinquents should be excluded from playing after having committed a delinquency. These punishments match the specific history lengths chosen in the binary reputation case.

Nevertheless in both cases confiders punish delinquents too hard (choose a history length that is too long) from a social perspective if the expected payoffs confiders

receive from trusting are less than from excluding a boundedly rational opponent. Here, limitation of information would have positive effects on the society. Specifically, such a policy would increase delinquents' wealth at costs of confiders. This, even though not pareto better, is socially optimal as the gains of delinquents would be higher than the losses of confiders.

But such a policy is clearly distortive. Confiders who already carry the costs of the institution providing information are forced to forego payoffs on behalf of delinquents. Therefore, if implementing such a policy on the backs of confiders, it has to be secured in such a way that confiders are still better off than they would be in the non-cooperative case. Otherwise there are clearly incentives to change their strategy to $S_1(0,0)$. Confiders would never engage in interaction and cooperation would break down. Thus I conclude that the creation of an institution that is collecting and reallocating information, in the manner of the reputation mechanism discussed, has the potential to increase social payoffs. There is space for "private" financing of this institution through confiders. Whether such an institution provides the optimal amount of information (a delinquent is marked long enough or too long) depends on the likeliness of delinquencies. If few delinquencies were committed, the market gives the right incentives and confiders choose the optimal amount of information. If the likeliness of delinquencies is above a certain level, confiders have personal incentives to "sort out" delinquents (mark and punish them forever after they have committed a delinquency). This is not optimal from a social perspective and would ask for government intervention. Limiting public information increases social payoffs but needs to be assessed carefully as otherwise the confiders would be driven out of cooperative equilibriums.

Thus I give an argument for the common saying: *Everybody deserves a second chance*.

8. ***Appendix A: Change in payoff structure***

Reproduction of the findings on the Nash-Equilibrium prediction and the falling payoff function for a changed payoff structure.
The payoff structure I imposed on the game and used throughout this paper ranks social payoffs in the stage game as following*:*

$$(c + f) > (b + e) > (a + d) \ equivalently \ (T,H)_{1+2} > (T,B)_{1+2} > (NT,\dots)_{1+2}$$

In the following section I reproduce these findings with a payoff structure where (Trust/Betray) and the (Out/…) payoffs switch according to the social payoff rank:

$$(c + f) > (a + d) > (b + e) \ equivalently \ (T,H)_{1+2} > (NT,\dots)_{1+2} > (T,B)_{1+2}$$

and the rest of the payoff structure stays the same:

$$(1) \quad c > a > b \ or \ in \ a \ different \ notation \ (T,H)_1 > (NT,\dots)_1 > (T,B)_1$$
$$(2) \quad e > f > d \ or \ in \ a \ different \ notation \ (T,B)_2 > (T,H)_2 > (NT,\dots)_2$$

Even if the specific conditions change, all general results can be reproduced and the general conclusion stays the same.

Proposition:
The payoff function is falling in history length even if the payoff structure is changed in the above discussed manner.

Proof:
I use the same approach as in section 5.1.1.

The payoff function is of the following structure:
$\Pi^{\text{social}}(k') = s * [4(c + f) + (b + e)] + (1 - s) * \big[B * \big(x(b + e) + (1 - x)(c + f)\big) + (5 - B) * (a + d)\big]$

This payoff function is decoupled from the exact number of players[7] as it represents the payoff of an average first and an average second mover who play for five rounds. I rewrite the social payoff function as a weighted average of $\Pi^{\text{rational}}(k')$ and $\Pi^{\text{bounded}}(k')$:

$$\Pi^{\text{social}}(k') = s * \Pi^{\text{rational}}(k') + (1 - s) * \Pi^{\text{bounded}}(k')$$

$$\Pi^{\text{rational}}(k') = [4(c + f) + (b + e)].$$
$$\Pi^{\text{bounded}}(k') = \left[B * \left(x(b + e) + (1 - x)(c + f)\right) + (5 - B) * (a + d)\right]$$

I start at a point in which cooperation equilibriums can be realized (see assumption 7) and the specific *history length* is k'. Increasing the *history length* by one round to k'+1 does not change $\Pi^{\text{rational}}(k')$. This is the case because rational and self-interested first and second mover already play best response as history length is increased from a k' already sustaining cooperation. Though the part of the social payoff function that is indicated by $\Pi^{\text{bounded}}(k')$ is changing. The only factor that changes is the probability of being *Blue* for boundedly rational second movers. The $B(S_1, S_2, k, x)$ is decreasing in k (see Appendix B).

$B(S_1, S_2, k, x)$ changes by $-\Delta B \rightarrow\ 5 - B(S_1, S_2, k, x)$ changes by ΔB, where $\Delta B \geq 0$ (note that $\Delta B = 0$ only holds if the $B(S_1, S_2, k, x)$ was already at its upper boundary of 5) therefore $\Pi^{\text{bounded}}(k')$ is changing if $\Delta b \neq 0$:

rewriting $\Pi^{\text{bounded}}(k')$ in terms of changes:

$if\ k' \rightarrow k' + 1$

$$\Longrightarrow \Delta\Pi^{\text{bounded}}(k' + 1) = (-\Delta B) * \overbrace{[x * (b + e) + (1 - x) * (c + f)]}^{u} + \Delta B * \overbrace{[a + d]}^{v} =$$
$$(-\Delta B) * u + \Delta B * v$$

This means that if **u** is bigger than **v** than $\Pi^{\text{bounded}}(k')$ is decreasing if $\boldsymbol{u \geq v}$

[7] The number of players needs to be sufficiently large to guarantee anonymity through the random matching.

$\boldsymbol{u} \geq \boldsymbol{v}$ if $x * (b + e) + (1 - x) * (c + f) \geq a + d$

This is the case if: $x \leq \frac{(a+d)-(c+f)}{(b+e)-(c+f)}$

$\Longrightarrow \Pi^{\text{bounded}}(k' + 1) \leq \Pi^{\text{bounded}}(k')$.

$\Pi^{\text{rationl}}(k')$ stays constant if *history length* increased.

Therefore:

$$\begin{matrix} \Pi^{\text{social}}(k' + 1) = s * \Pi^{\text{rational}}(k' + 1) + (1 - s) * \Pi^{\text{bounded}}(\text{k}' + 1) \\ \leq \\ \Pi^{\text{social}}(k') = s * \Pi^{\text{rational}}(k') + (1 - s) * \Pi^{\text{bounded}}(k') \end{matrix} \quad if\ x \leq \frac{a + d - c - f}{b + e - c - f}$$

■

Nash Equilibrium (in the bounded rationality case with the changed payoff structure)

The calculations are equivalent to the ones from section 5.2.2. but combined with the new condition for the falling payoff structure, the new Nash-Equilibrium is as following:

The cooperative Nash-Equilibrium exists if and only if:

$$Condition\ 1a\text{:}\ \frac{a - c}{b - c} < x \leq \frac{(a + d) - (c + f)}{(b + e) - (c + f)}$$

$$Condition\ 2a\text{:}\ s(k, x) \geq \frac{5a - y(x, k)}{4c + b - y(x, k)}$$

$$\rightarrow Nash - EQ^{*}(\pi_1^i, \pi_2^i, b + 4c \geq 5a, k \geq k^{*}, x, s)$$

$$= \begin{cases} S_1(1{,}0)\ and\ S_2(Honor)\ if\ \frac{a - c}{b - c} < x \leq \frac{(a + d) - (c + f)}{(b + e) - (c + f)} \wedge s(k, x) \geq \frac{5a - y(x, k)}{4c + b - y(x, k)} \\ S_1(0{,}0)\ and\ S_2(Betray'Invest) \qquad otherwise \end{cases}$$

9. Appendix B: Proof of a decreasing likeliness to be Blue

Proposition:

The $B(S_1, S_2, k, x)$ is decreasing in k.

Proof:

The reasoning behind this decrease of a boundedly rational second movers' likeliness to be *Blue,* if the history length increased, is simple. Given a specific set of strategies for first and second movers and a specific likeliness to play *Betray* for boundedly rational second movers, a ceteris paribus change in history length changes the $B(S_1, S_2, k, x)$in the following way:

Increasing k means that if committing a delinquency, second movers hold a bad reputation for a longer time, while nothing else changes therefore $B(S_1, S_2, k, x)$ decreases as it represents the amount of rounds an average second mover holds a good reputation throughout the five round game ■.

10. Appendix C: Abstract in English and German

Abstract in English

In this thesis I analyze the effects of public information on cooperation and efficiency in a game theory setting. My model is based on a sequential and repeated trust game with reputation formation. I determine cooperative Nash-Equilibriums for the standard theory case and analyze the effects on cooperation and the possible payoffs if standard theory assumptions were relaxed, i.e. in a bounded rationality case. I show that reputation formation in this trust game is, from a standard theory point of view, sufficient to sustain cooperation and that limiting the reputation information is beneficial for society in the bounded rationality case.

Abstract in German

In dieser Arbeit analysiere ich die Effekte von „Public Information“ auf Kooperationsverhalten in einem spieltheoretischen Kontext. Mein Modell basiert auf einem wiederholten Vertrauensspiel mit Reputations-Mechanismus. In diesem Kontext zeige ich, dass Reputation, unter Standardannahmen, ausreichend ist, um Kooperation zu garantieren und eine Limitierung der (Reputations-) Information positive Effekte auf die Gesellschaft haben kann.

11. Appendix D: CV

Nationality: Austrian
Date of Birth: February 7th, 1987
Home Address: Wachaustraße 35/24, 1020 Vienna
Telephone: +43 69917186500
Email: julian_f_richter@yahoo.de

Languages: German (mother tongue), English (fluent), French (basic) and Spanish (basic)

Tertiary Education

Oct 2009-present: **Master Program in Economics – University of Vienna, Austria**

- Special interest areas: Game Theory, Political Economy, Experimental / Behavioral Economics
- Currently waiting on my Master – Thesis (Limiting public information on delinquencies. The perspective of behavioral game theory.) approval in Behavioral Game Theory supervised by Prof. Jean – Robert Tyran

Sept 2008-July 2009: **Erasmus Exchange - Universitat Pompeu Fabra, Barcelona, Spain**

- Taught by some of the world's leading economists,
- Focus on continuous assessment and teamwork projects.

Oct 2006-July 2009: **Bachelor Program in Economics – University of Vienna, Vienna, Austria**

- Thesis 1: "Executive Pay - is it set accurately and does it provide the right incentives? "
- Thesis 2: "Pigouvian Taxes in comparison to Emission Permit Markets - market based policies for efficient pollution reduction"
- Both theses written under the supervision of professors from the Universitat Pompeu Fabra, Barcelona

Secondary Education

Sept 2001- June 2005: **Borg 3 Landstraße (Focus on Science), Vienna**

- .Graduated with honors: "Guter Erfolg"

Sept 1997-June 2001: **RG 3 Radetzkystraße (Focus on Science), Vienna**

Employment

October and November 2011: **Internship with the Commercial Section of the Austrian Embassy in Beijing (WKO: Außenwirtschaftsstelle Peking)**

August 2010: **Internship with LG Nexera**

Mid. July – mid. August 2010: **Internship in the Utilities Team with Bearing Point - Infonova**

March and May 2010: **designing a marketing concept and writing the marketing plan for a creative network project (not launched yet)**

2005 - 2008 and 2009 - present**: Waiter, Restaurant Al-Gebra, Vienna**

Oct 2005 – Sept 2006: **Alternative Civilian Service, Caritas der Erzdiözese Wien, Vienna**

Extra-curricular interests

- Extensive travel throughout Europe and Australia.
- Interested in international politics and relations and enjoy reading widely on the subject of economic, political and social issues
- Playing football for over ten years (some as captain of an ambitious hobby team)

Referees: Available on request

12. References

Andreoni, James, and John H. Miller. 1993. Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence. The *Economic Journal* 103(418) (May):570-85.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10(1) (July):122-142.

Bohnet, Iris, and Steffen Huck. 2004. Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change. *American Economic Review* 94(2) (May):362-366.

Bolton, Gary E., Elena Katok and Axel Ockenfels. 2004. Trust among Internet Traders: A Behavioral Economics Approach. Working Paper Series. No. 5, *University of Cologne*

Bolton, Gary E., and Axel Ockenfels. 2008. The limits of trust in economic transaction investigations of perfect reputation systems. Working Paper No. 2216. CESIFO

Bos, Marieke, and Leonard Nakamura. 2010. Should Credit Remarks be Forgotten? Evidence from Legally Mandated Removal. available at: https://fisher.osu.edu/blogs/efa2011/files/FIE_3_2.pdf

Brown, Martin, and Christian Zehnder. 2007. Credit reporting, relationship banking, and loan repayment. *Journal of Money, Credit and Banking* 39 (December): 1883-1918.

Cabral, Luis, and Ali Hortacso. 2004. The Dynamics of seller reputation: Theory and Evidence from eBay. Working Paper 10363, NBER, Cambridge, Mass.

Cook, Karen S., and Cooper, M. Robin. 2003. Experimental studies of cooperation, trust, and social exchange. In *Trust and Reciprocity* edited by Lin Ostrom and Jimmy Walker. 6th ed. New York: Russell Sage Foundation. pp. 209–244.

Dellarocas, Chrysanthos. 2006. How often should reputation mechanisms update a trader's reputation profile? *Information System Research* 17(3) (September): 271-285.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. A theory of sequential reciprocity. *Journal of Economic Behavior & Organization* 55 (August):505–531.

Elul, Ronel, and Piero Gottardi. 2011. Bankruptcy: is it enough to forgive or must we also forget?. Working Paper No. 11-14, *Federal Reserve Bank of Philadelphia, Philadelphia*

Fan, Ming, Yong Tan, and Whinston, Andrew. B. 2005. Evaluation and Design of Online Cooperative Feedback Mechanisms for Reputation Management. *IEEE Trans. Knowledge and Data Engineering* 17 (3) (March):244-254.

Fehr, Ernst, and Simon Gächter. 2002. Altruistic punishment in humans. *Nature* 415 (January): 137-140.

Fehr, Ernst, Urs Fischbacher and Simon Gächter. 2002. Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms. *Human Nature* 13 (March): 1-25.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge: MIT Press.

Gächter, Simon, Benedikt Herrmann, and Christian Thöni. 2004. Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. *Journal of Economic Behavior & Organization* 55(4) (December):505-531.

Jappelli, Tullio, and Marco Pagano. 2000. Information Sharing in Credit Markets: a Survey. Working Paper No. 36. CSEF. University of Salerno.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson. 1982. Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory* 27 (August): 245-252.

La Porta, R., Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1997. Trust in large organizations. *American Economic Association Papers and Proceedings* 87 (2), 333–338.

McKelvey, Richard D., and Thomas R Palfrey.1998. Quantal Response Equilibria for Extensive Form Games. *Experimental Economics* 1: 9-41.

Melnik, Mikhail I., and James Alm. 2002. Does a Seller's Ecommerce Reputation Matter? Evidence from eBay Auctions. *Journal of Industrial Economics* 50(3) (September): 337-49.

Milinski, Manfred, and Bettina Rockenbach. 2006. The efficient interaction of indirect reciprocity and costly punishment. *Nature* 444 (December): 718-723.

Musto, David K. 2004. What Happens When Information Leaves a Market? Evidence from Post bankruptcy Consumers. *Journal of Business* 77(4): 725-748.

Resnick, Paul, Richard Zeckhauser, John Swanson, and Kate Lockwood. 2006. The value of reputation on eBay: A controlled experiment. *Experimental Economics* 9(2) (June): 79-101.

Selten, Reinhard. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. International Journal of Game Theory 4 (1) (March):25–55.

Selten, Reinhard.1998. Features of experimentally observed bounded rationality. European Economic Review 42 (May): 413-436.

Stahl, D.O. 1990. Entropy control costs and entropic equilibria. International Journal of Game Theory 19 (June):129–138.

Vercammen, James A. 1995. Credit Bureau Policy and Sustainable Reputation Effects in Credit Markets. *Economica* 62 (November): 461-78.