Essays on
Cooperation

A cumulative dissertation submitted to the

Department of Economics
Faculty of Business, Economics and Social Sciences
Universität Hamburg

for the degree of

Doctor Rerum Politicarum
(Dr. rer. pol., PromO vom 18. Januar 2017)

Submitted by

Arne Pieters

(Born in Nieuwegein, The Netherlands)

Hamburg, 15. April 2019

Datum der Disputation: 28. August 2019

Vorsitzende: Prof. Dr. Anke Gerber

Erstgutachter: Prof. Dr. Andreas Lange

Zweitgutachter: Prof. Dr. Gerd Mühlheußer

Acknowledgements

Writing this dissertation would not have been possible without the help and support of a number of people. First, I would like to thank my supervisor Andreas Lange for giving me the chance to work at his chair. I have enjoyed the freedom he gave me to work on my research, and am grateful for his academic guidance and support. The research presented in this dissertation has benefited greatly from the many rounds of detailed feedback Andreas provided. I am also indebted to Gerd Mühlheußer for his helpful comments and suggestions.

Many colleagues at Universität Hamburg have contributed to the nice time I had there. I am especially grateful for having had Claudia Schwirplies as my office neighbour, colleague and sounding board over the past years. Her econometric skepticism is unmatched, her laugh can drown out any construction-work related noise Hamburg will throw at you, and I have yet to witness her in a bad mood. I would also like to thank Angelika Wilke for her administrative support and for pretending that my German was good enough to converse in. Another colleague I should mention is Arno Apffelstaedt, my co-author for the final chapter of this dissertation, whom I can recommend to anyone for a joint endeavor. I am thankful to Arno for the many hours of productive phone conversations, and more generally for a fruitful cooperation.

My work has benefited from the financial support provided by the WiSo Graduate School at UHH, which facilitated most of my conference travels, but also funded the economic experiments I conducted. In regards to the latter, I am also grateful to the assistance provided by the WiSo research lab team.

One of the challenges in writing a dissertation is to, occasionally, *not* think about research. My parents, grandmother, sister, brother, nephews and nieces have all provided very welcome distractions, for which I am thankful. The same is true for my friends back home, who do not pretend to care about my research any more than I do about their jobs, but are nevertheless a constant source of encouragement and severe entertainment, even from a distance.

My final thanks go to Luise Görges. I am indebted to her for reading my work, for thinking about my work, for her uncondontional belief and support, for being a reason to move to this great city, and in general for showing the world how *it* is done. You're so cool.

Summary

This dissertation comprises three essays that study cooperation. Based on experimental and theoretical analysis, it investigates different aspects of how we cooperate to fund public goods.

Chapter 1 contributes to the understanding of how we negotiate long-term agreements that are not legally binding. A game theoretic model is presented, that studies two parties who repeatedly co-fund a public good. They incur bargaining costs everytime they negotiate contributions, and they cannot contractually commit themselves to future contribution levels. To reduce the frequency of costly negotiations, they can pre-negotiate a non-binding burden sharing agreement. The individual public good benefits randomly vary from period to period, affecting the parties' ex post bargaining positions. The self-enforcing range for a given agreement is derived, within which neither party stands to reduce its contribution so much that it outweighs the costs of renegotiation. In the analysis of the ex ante negotiations, it is shown that symmetric bargaining will lead to a long-term agreement that is equivalent to the expected ex post negotiation outcome. This agreement minimizes the probability of renegotiation, equally divides the expected payoff improvement, and under risk neutrality will be achieved independent of individual benefit variance. If bargaining costs are asymmetric, the agreement will still minimize the renegotiation probability, but this implies that the burden division will favor the party with the lower costs relative to the symmetric case. Chapter 1 goes on to show that asymmetry in bargaining strength is amplified in the pre-negotiated agreement. The resulting burden division will favor the stronger bargainer and will be renegotiated inefficiently often. This may have the adverse consequence that the weaker party will refuse to start pre-negotiations in the first place. Any agreement that is honored with positive probability reduces the scope for risk sharing. This means that if the parties are risk averse, asymmetry in individual benefit variance can alter ex ante bargaining positions, resulting in a long-term agreement that deviates from that which minimizes the renegotiation probability, to compensate the party with the greater individual risk.

Chapter 2 presents the results from a lab experiment that implements a simplified version of the model considered in Chapter 1, specifically designed to test theoretical predictions relating to risk asymmetry. Participants are matched in pairs and, in a repeated setting, have to decide how to divide the costs for a joint project. The first

part of the experiment studies ex post behavior. A non-binding rule for the cost division is exogenously imposed, and as project revenues change from period to period, the participants have to decide whether to stick to the rule or renegotiate contributions. Renegotiation is costly to both participants, but yields an equal division of profits by adjusting the cost division to that period's revenues. The main contribution of this study is the finding that subjects who face less individual revenue risk (ex ante) are more likely to forego the opportunity to force renegotiations (ex post) that would improve their individual payoff, but would reduce the joint earnings. Such cooperative actions effectively extend the self-enforcing range of the imposed rule. The findings presented in Chapter 2 show no indication for a direct effect of risk asymmetry, *per se*, on the degree of cooperation between partners. The second part of the experiment tests the theoretical prediction that, ex ante, the rule is of less value to the party with the greater individual revenue risk. The data provide no evidence in support of this prediction.

Chapter 3 presents the results of an online experiment that contributes to the understanding of pro-social decisions. For many real world social dilemma's, there is a time delay between decisions and their consequences. Existing experimental research on this issue, which is relatively scarce, indicates that such delay can affect social decisions, both positively and negatively. The experiment is designed to study the impact of delay on the impure motives for contributing to a public good, by combining payoff delay with two framing treatments that are known to generate contribution differences. In the positive frame, buying into the public good generates a positive externality by increasing the payoffs of others. In the negative frame, buying into the private good generates a negative externality by reducing others' payoffs. The earnings from these games are either paid out immediately, or thirty days after the experiment takes place. The results show that, without delay, participants contribute significantly more in the negative externality frame than in the positive frame. As the payoff structure of the game is identical under both frames, this difference must be related to "impure" utility elements, that are directly affected by the contribution decision and are not purely based on the resulting payoffs. The framing effect disappears when payoffs are delayed by thirty days. Significantly reduced contributions in the negative frame suggest that the delay enables individuals to not feel bad about reducing the earnings of others. The same change is not observed in the positive frame, where payoff delay results in a non-significant increase in contributions. The observed interaction between framing and payoff delay suggests that, in predicting the effects of different frames on real-world social decisions, it is important to take into consideration the relative timing of decisions and their consequences.

Zusammenfassung

Diese Dissertation umfasst drei Aufsätze, die verschiedene Aspekte menschlichen Kooperationsverhaltens bei der Finanzierung öffentlicher Güter anhand experimenteller und theoretischer Analysen untersuchen.

Kapitel 1 geht der Frage nach, wie langfristige, nicht rechtsverbindliche Vereinbarungen ausgehandelt werden. . Im Zentrum steht ein spieltheoretisches Modell, das das Verhalten zweier Parteien abbildet, die wiederholt gemeinsam ein öffentliches Gut finanzieren. Den Parteien entstehen für jede neue Verhandlung Kosten und sie können künftige Beitragsniveau nicht vertraglich bindend festlegen. Um die Häufigkeit kostspieliger Verhandlungen zu reduzieren, können sie jedoch eine unverbindliche Lastenteilungsvereinbarung vorverhandeln. Die individuellen Erträge aus dem öffentlichen Gut variieren zufällig von Periode zu Periode und beeinflussen die Ex-post-Verhandlungspositionen der Parteien. Es wird der sich selbst durchsetzende Bereich für eine gegebene Vereinbarung abgeleitet, innerhalb dessen keine Partei ihren Beitrag so stark reduzieren kann, dass diese Ersparnis die Kosten für Neuverhandlungen überwiegt. In der Analyse der Ex-ante-Verhandlungen wird gezeigt, dass symmetrische Verhandlungen zu einer langfristigen Vereinbarung führen werden, die dem erwarteten Ergebnis der Ex-post-Verhandlungen entspricht. Diese Vereinbarung minimiert die Wahrscheinlichkeit einer Neuverhandlung, teilt die erwartete Auszahlungsverbesserung gleichmäßig auf und wird unter Risikoneutralität unabhängig von den individuellen Ertragsabweichungen erzielt. Wenn die Verhandlungskosten asymmetrisch sind, wird die Neuaushandlungswahrscheinlichkeit durch die Vereinbarung immer noch minimiert. Allerdings begünstigt die Lastenteilung die Partei mit den niedrigeren Kosten gegenüber dem symmetrischen Fall. Kapitel 1 zeigt weiter, dass die Asymmetrie der Verhandlungsstärke in der vorverhandelten Vereinbarung verstärkt wird. Die sich daraus ergebende Lastenteilung wird die verhandlungsstärkere Partei begünstigen und wird ineffizient häufig neu verhandelt. Dies kann die nachteilige Folge haben, dass die schwächere Partei Vorverhandlungen verweigert. Jede Vereinbarung, die mit einer positiven Wahrscheinlichkeit erfüllt wird, verringert den Spielraum für Risikoteilung. Dies bedeutet, im Falle von risikoscheuen Parteien, dass eine Asymmetrie in die individuellen Varianz die Ex-ante-Verhandlungspositionen so verändern kann, dass die langfristige Vereinbarung, nicht länger die Wahrscheinlichkeit einer Neuverhandlung minimiert, um die Partei mit dem höheren individuellen Risiko zu entschädigen.

Kapitel 2 präsentiert die Ergebnisse eines Laborexperiments, in dem eine vereinfachte Version des in Kapitel 1 betrachteten Modells implementiert wird, die speziell zum Testen theoretischer Vorhersagen in Bezug auf die Risikoasymmetrie entwickelt wurde. Die Teilnehmer*innen werden paarweise zusammengeführt und müssen in wiederholt entscheiden, wie die Kosten für ein gemeinsames Projekt aufgeteilt werden. Der erste Teil des Experiments untersucht das Ex-post-Verhalten. Eine unverbindliche Regel für die Kostenteilung ist exogen gegeben, und da sich die Projekterlöse von Periode zu Periode ändern, müssen die Teilnehmer*innen entscheiden, ob sie an der Regel festhalten oder Beiträge neu verhandeln möchten. Die Neuverhandlung ist für beide Teilnehmer*innen kostspielig, führt jedoch zu einer symmetrischen Gewinnverteilung, da die Kostenteilung an die Erträge dieser Periode angepasst wird. Der Hauptbeitrag der Studie liegt in der Erkenntnis, dass Personen, die einem geringeren individuellen Einkommensrisiko (ex-ante) ausgesetzt sind, eher auf die Gelegenheit verzichten, Neuverhandlungen (ex-post) zu erzwingen, die ihre individuelle Auszahlung verbessern, jedoch die gemeinsamen Einkünfte verringern würden. Durch solche kooperativen Maßnahmen wird der sich selbst durchsetzende Bereich der Regelung wirksam erweitert. Die in Kapitel 2 dargestellten Ergebnisse zeigen keinen Hinweis auf eine direkte Auswirkung der Risikoasymmetrie (*per se*) auf den Grad der Zusammenarbeit zwischen den Partnern. Der zweite Teil des Experiments testet die theoretische Vorhersage, dass die Regel vorab für die Partei mit dem höheren individuellen Einkommensrisiko von geringerem Wert ist. Die Daten liefern keine Belege für diese Vorhersage.

Kapitel 3 präsentiert die Ergebnisse eines Online-Experiments, das zum Verständnis prosozialer Entscheidungen beiträgt. Viele soziale Dilemmata der realen Welt zeichnen sich durch eine zeitliche Verzögerung zwischen Entscheidungen und ihren Folgen aus. Die wenigen verfügbaren Ergebnisse aus der experimentellen Forschungen zu diesem Thema zeigen, dass eine solche Verzögerung soziale Entscheidungen sowohl positiv als auch negativ beeinflussen kann. Das Experiment soll die Auswirkungen zeitlicher Entkoppelung auf die "unreinen" Motivationen für individuelle Beiträge zu einem öffentlichen Gut untersuchen. Dazu wird die Auszahlungsverzögerung mit zwei Framing-Verfahren kombiniert, von denen bekannt ist, dass sie unterschiedliche Entscheidungen hervorrufen. Im positiven Framing erzeugt die Investierung in öffentlichen Gütern eine positive Externalität, indem die Auszahlungen anderer erhöht werden. Im negativen Framing erzeugt die Investierung im privatem Gut eine negative Externalität, indem er die Auszahlungen anderer reduziert. Die Einnahmen aus diesen Spielen werden entweder sofort oder dreißig Tage nach dem Experiment ausgezahlt. Die Ergebnisse zeigen, dass die Teilnehmer ohne Verzögerung wesentlich mehr im negativen als im positiven Framing beitragen. Da die Auszahlungsstruktur des Spiels in beiden Frames identisch ist, muss sich dieser Unterschied aus den "unreinen" Nutzenkomponenten speisen, die direkt von der Beitragsentscheidung betroffen sind und nicht ausschließlich auf den daraus resultierenden Auszahlungen basieren. Der Framingeffekt verschwindet, wenn die Auszahlungen um dreißig Tage verzögert werden. Deutlich reduzierte Beiträge im negativen Rahmen deuten darauf hin, dass die Verspätung es den Betroffenen ermöglicht, sich weniger schlecht zu fühlen, wenn es darum geht, das Einkommen anderer zu reduzieren. Dieselbe Änderung wird nicht im positiven Rahmen beobachtet, wo die Auszahlungsverzögerung zu einer nicht signifikanten Erhöhung der Beiträge führt. Die beobachtete Interaktion zwischen Framing und Auszahlungsverzögerung legt nahe, dass es bei der Vorhersage der Auswirkungen verschiedener Rahmen auf soziale Entscheidungen in der realen Welt wichtig ist, den relativen Zeitpunkt der Entscheidungen und ihre Konsequenzen zu berücksichtigen.

List of Included Essays

Chapter 1: Negotiating Non-Binding Burden Sharing Agreements when Renegotiations are Costly
Authors: Arne Pieters

Chapter 2: Asymmetric Risk and Non-Binding Rules: An Experiment
Author: Arne Pieters

Chapter 3: Cooperating Tomorrow: Warm Glow vs. Cold Prickle Revisited
Authors: Arno Apffelstaedt and Arne Pieters

Contents

Introduction

> *"Ain't no rules, ain't no vow, we can do it anyhow: I'n'I will see you through. 'Cos everyday we pay the price with a little sacrifice, Jammin' till the jam is through."*
>
> — Bob N. Marley, 1977

This dissertation is about cooperation. The human ability to cooperate is important for any economic situation that involves individuals or entities whose interests are, as John Nash (1953) put it, "neither completely opposed nor completely coincident". Most of our decisions and actions do not only affect ourselves, but also have external effects on the welfare of others. If everyone ignores such *externalities*, and makes isolated decisions that are purely based on self-interest, we usually leave room for mutual improvement. Whenever interests are not fully aligned, some degree of cooperation is required to realize the best joint outcome.

In the three essays presented in this dissertation, I use theoretical and empirical analysis to study different aspects of how we fund public goods. Without cooperation, public goods tend to be underproduced, typically because their costs do not outweigh their benefits at the individual level. To a certain extent, governments are useful instruments to facilitate public good provision and cooperative behavior in general. They can legally enforce the agreement between citizens not to hurt each other, steal from each other, or do anything else they agree is mutually undesirable. Governments have the legal power to collect taxes, allowing citizens to jointly fund public goods such as infrastructure or education. They do not solve every problem relating to public goods or externalities, however. First, even when technically and legally feasible, society must agree on the extent to which government intervention is desirable, which is not always straightforward. Second, not every situation involving externalities can be governed, in detail, by law. Third, when it comes to international public goods, there simply is no supranational government.

In situations that do not allow for legally binding contracts, we can only write agreements that are not binding. Chapters 1 and 2 consider such non-binding agreements. For public goods that require repeated investments, it is not uncommon to have in place a long term agreement on how to divide the financial burden, even when that agreement is not a binding contract. Over time, changing circumstances can affect what we see as a fair burden division and, in the case of a non-binding agreement, every party involved can unilaterally decide to demand renegotiations. Why might such agreements nevertheless exist? Why do we stick to an agreement, even though we could improve by renegotiating? One reason may be that we fear not sticking to the deal will create hostility among the other parties (Hart and Moore, 2008). It could also be explained by an intrinsic desire to be trustworthy, and not to renege on a made promise (Dufwenberg et al., 2017). However, even in absence of such "behavioral" motivations, people might stick to agreements because they want to avoid renegotiations, which cost time and effort that could be spent elsewhere. In Chapter 1, I use a game-theoretic setting to study agreements that are enforced by this desire to avoid costly negotiations. Chapter 2 employs an empirical approach to the same topic, using a lab experiment to test several theoretical predictions pertaining to a non-binding cost division.

Even without *any* agreement with others, people tend to voluntarily contribute to public goods. There is plenty of evidence that we give more than standard theory would predict. In public good game experiments, people often contribute positive amounts voluntarily.[1] In the real economy, there is clearly a willingness to contribute to privately provided public goods, such as charity.[2] People can be motivated to do so by a "purely" altruistic desire to improve the lives of others. In addition, it may serve the objective of gaining prestige or respect, of avoiding a feeling of guilt, or of experiencing a "warm glow". Such factors, that are more directly related to the own contribution, to doing good, rather than to the achieved result (making someone else better off), are referred to as "impure" motives (Andreoni, 1989, 1990). Although this term has a somewhat negative connotation, the world would be a great place if everyone was impurely motivated to generate love, peace and clean air. Especially in situations where the marginal impact of an individual good deed is negligible, it helps when the motivation can be found within oneself and is not too dependent on the achieved effect.

Impure motives are the focus of Chapter 3. It reports the results of an online public good game experiment that studies how impure motives drive intertemporal choice. Many cooperative decisions we face are intertemporal in nature. How much we choose to pollute the environment *today* will affect the state of the planet in the *future*. In other cases, we decide today about cooperating in the future, for instance when we register as an organ donor, or commit to a future donation to charity. Pure motives relate to outcomes. It is reasonable to assume that most people care less about outcomes that take a long time to

[1] See Chaudhuri (2011) for an overview.
[2] See, e.g., Andreoni and Payne (2013).

occur than they do about those that arrive immediately. Impure motives, however, may be affected by time differently, as they are not directly linked to outcomes. For example, if we sign up for future charitable donations because we want to show our friends that we are good people, does it really matter when the donation actually occurs? Or is the timing of the decision more important? These type of questions motivate the research presented in Chapter 3.

Below, I shortly outline the contribution of each chapter in more detail.

Chapter 1 contributes to the understanding of how we negotiate long-term agreements that are not legally binding. I present a game theoretic model that studies two parties who repeatedly co-fund a public good. They incur bargaining costs everytime they negotiate contributions, and they cannot contractually commit themselves to future contribution levels. To reduce the frequency of costly negotiations, they can pre-negotiate a non-binding burden sharing agreement. The individual public good benefits randomly vary from period to period, affecting the parties' ex post bargaining positions. The non-binding nature of the agreement implies that each party has the power to unilaterally opt out ex post, for instance to exploit a particularly strong bargaining position by renegotiating towards a lower contribution level. I derive the self-enforcing range for a given agreement, within which neither party stands to reduce its contribution so much that it outweighs the costs of renegotiation.

To analyze the ex ante negotiations on the long-term agreement, I first consider a basic version of the model in which the parties are risk neutral, face the same bargaining cost, and have symmetric bargaining power. I show that the parties will reach a long-term agreement that is equivalent to the expected ex post negotiation outcome. This agreement minimizes the probability of renegotiation, equally divides the expected payoff improvement, and is achieved independent of individual benefit variance. I then consider several variations to this basic model by separately relaxing some of its assumptions. First, if bargaining costs are asymmetric, the negotiated agreement will still minimize the renegotiation probability, but will not reflect the expected outcome of future negotiations. Relative to that, it favors the party with the lower bargaining costs. Second, I show that asymmetry in bargaining power is amplified by a pre-negotiated agreement. The stronger bargainer negotiates a burden division that is so favorable that, ex post, it is renegotiated inefficiently often. This may have the adverse consequence that, foreseeing this, the weaker party will refuse to start pre-negotiations in the first place. Third, I show that under risk averse preferences, the agreement may not be independent of individual benefit variance. The reason is that any agreement that reduces the probability of ex post renegotiation, also reduces the scope for risk sharing. Therefore, when parties face asymmetric individual benefit variance, the agreement will asymmetrically impact the allocation of risk. If the parties are risk averse, such asymmetry can result in an agreement that deviates from that which minimizes the renegotiation probability, to compensate the party with the greater individual risk.

Chapter 2 presents the results from a lab experiment that implements a simplified version of the model considered in Chapter 1, specifically designed to test theoretical predictions relating to risk asymmetry. Participants are matched in pairs and, in a repeated setting, have to decide how to divide the costs for a joint project. The first part of the experiment studies ex post behavior. A non-binding rule for the cost division is exogenously imposed, and as project revenues change from period to period, the participants have to decide whether to stick to the rule or renegotiate contributions. Renegotiation is costly to both participants, but yields an equal division of profits by adjusting the cost division to that period's revenues. The main contribution of this study is the finding that subjects who face less individual revenue risk (ex ante) are more likely to forego the opportunity to force renegotiations (ex post) that would improve their individual payoff, but would reduce the joint earnings. Such cooperative actions effectively extend the self-enforcing range of the imposed rule. Previous studies[3] have found payoff asymmetry to have a negative effect on cooperation levels. The restults presented in Chapter 2 show no indication that this adverse effect of asymmetry extends to the risk domain: I find no evidence in support of a direct effect of risk asymmetry, *per se*, on cooperation levels. The second part of the experiment tests the theoretical prediction that, ex ante, the rule is of less value to the party with the greater individual revenue risk. I find no evidence to support this prediction.

Chapter 3 is joint work with Arno Apffelstaedt, and presents the results of an online experiment that contributes to the understanding of pro-social decisions. For many real world social dilemma's, there is a time delay between decisions and their consequences. The existing experimental research on this issue, which is relatively scarce, indicates that such delay can affect social decisions.[4] We study its impact on the impure motives for contributing to a public good, by combining payoff delay with two well-known framing treatments proposed by Andreoni (1995). In the positive frame, buying into the public good generates a positive externality by increasing the payoffs of others. In the negative frame, buying into the private good generates a negative externality by reducing others' payoffs. The earnings from these games are either paid out immediately, or thirty days after the experiment takes place. We find that, without delay, participants contribute significantly more in the negative externality frame than in the positive frame. As the payoff structure of the game is identical under both frames, this difference must be related to "impure"

[3] Examples include Ahn et al. (2007); Beckenkamp et al. (2007).

[4] Breman (2011) and Andreoni and Serra-Garcia (2017) find that people are more generous when deciding about future donations (compared to immediate donations). Contrastingly, Kovarik (2009) and Dreber et al. (2016) find that giving in dictator games *decreases* when the implementation is delayed.

utility elements, that are directly affected by the contribution decision and are not purely based on the resulting payoffs. The framing effect disappears when payoffs are delayed by thirty days. Significantly reduced contributions in the negative frame suggest that the delay enables individuals to not feel bad about reducing the earnings of others. We do not observe this in the positive frame, where payoff delay results in a non-significant increase in contributions. The observed interaction between framing and payoff delay suggests that, in predicting the effects of different frames on real-world social decisions, it is important to take into consideration the relative timing of decisions and their consequences.

References

Ahn, T.K., Myungsuk Lee, Lore Ruttan, and James Walker, "Asymmetric Payoffs in Simultaneous and Sequential Prisoner's Dilemma Games," *Public Choice*, 2007, *132* (3/4), 353–366.

Andreoni, James, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, 1989, *97* (6), 1447–1458.

_ , "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal*, 1990, *100* (401), 464–477.

_ , "Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments*," *The Quarterly Journal of Economics*, 1995, *110* (1), 1–21.

_ **and A. Abigail Payne**, "Chapter 1 - Charitable Giving," in Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, eds., *handbook of public economics, vol. 5*, Vol. 5 of *Handbook of Public Economics*, Elsevier, 2013, pp. 1 – 50.

_ **and Marta Serra-Garcia**, "Time-Inconsistent Charitable Giving," 2017. NBER Working Paper No. 22824.

Beckenkamp, Martin, Heike Hennig-Schmidt, and Frank Maier-Rigaud, "Cooperation in Symmetric and Asymmetric Prisoner's Dilemma Games," *Working Paper Series of the Max Planck Institute for Research on Collective Goods*, 03 2007.

Breman, Anna, "Give more tomorrow: Two field experiments on altruism and intertemporal choice," *Journal of Public Economics*, 2011, *95* (11-12), 1349–1357.

Chaudhuri, Ananish, "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature," *Experimental Economics*, Mar 2011, *14* (1), 47–83.

Dreber, Anna, Drew Fudenberg, David K. Levine, and David G. Rand, "Self-Control, Social Preferences and the Effect of Delayed Payments," 2016. mimeo.

Dufwenberg, Martin, Maros Servatka, and Radovan Vadovic, "Honesty and informal agreements," *Games and Economic Behavior*, 2017, *102*, 269 – 285.

Hart, Oliver and John Moore, "Contracts as Reference Points," *Quarterly Journal of Economics*, 2008, *CXXIII*, 1–48.

Kovarik, Jaromir, "Giving it now or later: Altruism and discounting," *Economics Letters*, 2009, *102* (3), 152–154.

Marley, Bob N. and The Wailers, *Jamming*, Tuff Gong / Island Records, 1977.

Nash, John F., "Two-Person Cooperative Games," *Econometrica*, 1953, *21* (1), 128–140.

Chapter 1

Negotiating Non-Binding Burden Sharing Agreements when Renegotiations are Costly

Author: Arne Pieters

Abstract: I study two parties in a long-term relationship, repeatedly co-funding a public good. To avoid the costs of repetitive bargaining over their contributions, they pre-negotiate a non-binding burden sharing agreement. The individual public good benefits randomly vary from period to period, affecting the parties' ex post bargaining positions. I show that symmetric bargaining will lead to a long-term agreement that is equivalent to the expected ex post negotiation outcome. This minimizes the probability of renegotiation, equally divides the expected payoff improvement, and under risk neutrality will be achieved independent of individual benefit variance. If bargaining costs are not equal for both parties, the agreement will still minimize the renegotiation probability, but this implies that the burden division will favor the party with the lower costs. I show that asymmetry in bargaining strength is amplified in the pre-negotiated agreement. The resulting burden division will favor the stronger bargainer and will be renegotiated inefficiently often. This may have the adverse consequence that the weaker party will refuse to start pre-negotiations in the first place. Any agreement that is honored with positive probability reduces the scope for risk sharing. This means that if the parties are risk averse, its value is negatively (positively) related to one's (partner's) individual benefit variance. Greater individual benefit risk could thereby improve one's bargaining position in the pre-negotiations.

JEL: C70, C73, D86, F51, F55

Keywords: Public Goods, Burden Sharing, Repeated Bargaining, Costly Negotiation, Non-binding Agreements

1.1 Introduction

Many economic relations are governed by agreements that to some extent cannot be enforced by a court. It may be altogether infeasible to write binding contracts when courts are unreliable or simply not available. For instance, when two or more countries sign an agreement on how to share the future financial burden for some international institution, they cannot rely on a supranational court to enforce this agreement. Accordingly, the slightest change in bargaining positions can be exploited by a unilateral push for renegotiations. However, the implied instability of international burden sharing rules is generally not observed in reality.

As an example, consider the EU budget, which is determined annually in terms of expenditures, constrained by ceilings that are determined in the Multi-annual Financial Framework. The lion's share of the required revenue comes from direct member state transfers that are set as a percentage of each member's GNI, thereby ignoring change in all other relevant economic, financial and political variables. Prior to the recent decision by the United Kingdom to withdraw from the EU entirely, hardly any deviations from this divison rule occurred, with the most notable exception the "UK Rebate" negotiated by Margaret Thatcher in 1984.[1]

One reason for infrequent renegotiations is that they are costly. In situations such as the EU example, these costs may stem from the time and effort involved for all the negotiating parties, but could also reflect that renegotiations cause uncertainty about, or delay of, the funding of the public good. This provides a common incentive to pre-negotiate a standard division of financial contributions, and to stick to this *burden sharing rule (BSR)* as much as possible.

This paper studies the use of non-binding, long-term agreements that are motivated by a desire to avoid repetitive negotiations. I consider a simple model where two parties repeatedly fund a public good. In the model, the parties incur an exogenous bargaining cost every time they negotiate how to divide the burden. I assume they cannot make a long-term contractual commitment to future contribution levels. However, they can agree on a simple rule that can be implemented repeatedly without the need to negotiate. From period to period, there is random variation in the individual public good benefits. While the optimal total public good investment is fixed, the individual benefits do affect the parties' bargaining positions with respect to who contributes what share. A party with an improved bargaining position may decide to unilaterally opt out of the non-binding agreement to negotiate a better burden division, but will only do so if the anticipated improvement outweighs the incurred bargaining costs. These costs thus have a dual function in the model: they are the reason to establish a long term rule, and at the same time provide

[1] Effective in 1985 - See: 85/257/EEC, Euratom: Council Decision of 7 May 1985 on the Communities' system of own resources.

that rule with a self-enforcing quality.

The main focus of the paper is on what shapes the long term agreement. Under the assumption that the parties are risk neutral and that negotiation outcomes correspond to the symmetric Nash bargaining solution,[2] I establish a baseline result where the parties agree on the optimal rule. It reflects the expected outcome of future negotiations, which minimizes the probability of either party opting out of the agreement (it is the *least-renegotiated rule*), and equally divides the value it generates. The paper then discusses the implications of relaxing several assumptions in the basic model. I show that with unequal bargaining costs, the parties will still establish the least-renegotiated rule, but that this rule is no longer based on the expected outcome of future negotiations. Relative to that, it favors the party with the lower bargaining costs.

While the baseline result is independent of the variance in individual public good benefits, this is not the case when the parties are risk averse. Repetitive bargaining, while costly, allows the parties to share risk by adjusting to the variability of their public good benefits. Therefore, any pre-agreed rule that is implemented with positive probability increases the overall risk exposure, and may impact how risk is allocated between the parties. When the individual benefit variance is greater for one party than for the other, that party has a better bargaining position regarding the BSR. This can result in an agreement that deviates from the least-renegotiated rule, compensating one party for taking on more risk.

When allowing for asymmetric bargaining power, the model reveals that a pre-negotiated agreement amplifies this asymmetry between the parties relative to per-period negotiations. The least-renegotiated rule will reflect the anticipated asymmetry in future renegotiations. However, the stronger bargainer will negotiate a rule that reduces its contribution level beyond this. This creates the possible adverse consequence that the weaker bargainer will reject pre-negotiations altogether, and shows why a strong bargainer may be better off committing to a fair outcome before starting the negotiations, if possible.

It is useful to compare the current paper on non-binding agreements to the existing literature on incomplete contracts. This literature generally considers a contract incomplete if it leaves certain decisions or transactions open to be determined later (Bolton and Dewatripont, 2005), for instance because actions and events are not verifiable ex post, or impossible to describe ex ante. This lack of commitment can create the opportunity for one party to hold up the other. Various studies propose solutions to this problem, e.g., by shifting property rights (Grossman and Hart, 1986) or by combining incomplete contracts with a specified process for future revisions(Hart and Moore, 1988; Chung, 1991). These solutions exploit the ability to contract on certain elements to affect the parties' options

[2] See Nash (1950, 1953). All results qualitatively hold for any bargaining solution that gives weight to both efficiency and equity (e.g., Kalai and Smorodinsky (1975)), but I will use the Nash bargaining solution throughout.

and bargaining power in the future, thereby achieving the desired impact on other elements that are not contractible. Such solutions are not viable in the setting considered by the current paper, as they require some degree of enforcement by court. In achieving efficiency with incomplete contracts, it can be problematic that the parties cannot commit ex ante to refrain from renegotiations ex post (Che and Hausch, 1999). For the non-binding agreements considered here, this problem is worsened by the ability of either party to unilaterally opt out ex post, which pulls the threat point for renegotiations down to the noncooperative outcome.

More recent literature attempts to explain the prevalence of incomplete contracts using behavioral ideas, considering the effects of contracts and agreements beyond their legally binding provisions. Hart and Moore (2008) develop a theory based on the idea that contracts set reference points, shaping parties' expectations and entitlements. Hart (2009) applies this theory to a buyer and a seller in a long-term relationship, facing uncertainty ex ante about the buyer's value and the seller's cost. If either deviates too much from what is expected, one party will have an incentive to force renegotiation of the price, which leads to a welfare loss by turning the relation hostile. The model studied by Hart is similar to that presented in the current paper, in that the main challenge in writing an ex ante agreement is payoff uncertainty rather than noncontractible investments. In both models, the desire to avoid renegotiations gives the initial agreement a self-enforcing range. A difference is that the current paper considers an exogenous cost to bargaining ex post, that does not depend on what the parties agreed on ex ante (but rather provides a reason for them to agree on a rule). In Hart's model, the threat of hostility is endogenous to the initial contract and the entitlements it creates. My approach differs in a similar manner from recent experimental and theoretical work on agreements that are entirely non-binding, which is focused on the psychological or emotional costs of reneging on an agreement (Miettinen, 2013; Dufwenberg et al., 2017).

The exogenous bargaining costs I consider are more similar to transaction costs, which have been studied in relation to the design of contracts. Masten and Crocker (1985) argue that the optimal length of a contract reflects a tradeoff between the cost of period-by-period negotiations and the inflexibility of being bound to an agreement. Masten (1988) considers a buyer-seller model where, similar to my setting, transaction costs drive a wedge between actual and potential outcomes. He shows that an important aspect of contractual design is 'Hazard Equilibration', equating the expected costs of opportunistic behavior on both side of the relation. This corresponds to my baseline result where both parties are equally likely to opt out of the agreement ex post, but, as my model shows, it is not always achieved when there is asymmetry in risk or in bargaining power.

My analysis of how the burden sharing rule affects risk sharing relates to work by Perloff (1981), who demonstrates that allowing breaches of forward contracts under extreme circumstances can improve welfare by reducing the income risk of farmers. His focus is on the optimal breach policy for a given contract, not on how the choice of forward contract is impacted. Polinsky (1987) does consider how risk aversion and the degrees of uncertainty on both sides of the relation determine what is preferred between a spot price or a fixed priced contract, while also considering a hybrid form that combines a spot price with a floor price. A burden sharing rule with a limited range of self-enforcement is similar to a different type of hybrid contract: one that implements a fixed price unless it deviates too much from the spot price.

The remainder of this paper is organized as follows. The next section describes the burden sharing stage game, the negotiation of contributions and establishes the bargaining solution for the pre-negotiated rule. Section 1.3 explores several separate extentions to the model, showing how the agreement can be impacted by risk averse preferences, asymmetric bargaining costs, asymmetry in bargaining power, and correlation between the individual benefits. In Sections 1.2 and 1.3, I focus on one-shot Nash equilibrium strategies regarding the parties ex post decision to honor the agreement. Section 1.4 considers the alternative trigger strategy equilibriua in which the rule has an expanded self-enforcing range. I show that while such equilibria can be subgame perfect, the implied strategy profiles themselves are vulnerable to renegotiation. Section 1.5 concludes.

1.2 The Model

Consider a long-term relationship between two parties $i \in \{A, B\}$, who both benefit from the provision of a public good. Time is modeled in discrete periods $t \in \{1, 2, ...\}$ and the relationship is assumed to have an infinite time horizon, with both parties discounting the future at rate δ. In each period t, the parties have to determine who contributes how much to the public good. I refer to the situation they face in each period as the stage game, and will describe its strategic structure in more detail in subsection 1.2.1. In case they have no prior agreement, the parties can meet to negotiate contributions $\boldsymbol{q_t} = \{q_t^A, q_t^B\}$ on a period-to-period basis. In the process, they both incur bargaining cost $\alpha > 0$, regardless of the outcome. To avoid repetitive bargaining costs, the parties can pre-negotiate a simple burden sharing rule (BSR), $\boldsymbol{q_r} = \{q_r^A, q_r^B\}$, that can be implemented upon mutual consent, without the need to negotiate.

There is uncertainty about how much each party will benefit from the public good in future periods. The state of the world in period t is captured by $\{h_t^A, h_t^B\}$, the set of public good benefits. Each h_t^i is drawn every period from an ex ante commonly known distribution, that is assumed to be independent of previous draws, implying that expectations about the future state of the world are not affected by the current state or by previous states. This assumption allows the analysis to focus on how a BSR divides future surplus and how likely it is to be honored ex-post, and avoids the need to consider future renegotation of the agreement itself. There is no private information: at the start of each period t, the state of the world is fully revealed to both parties. The public good benefits follow mutually independent, unimodal, symmetric distributions with support $h_t^i \in [h_{\min}^i, h_{\max}^i]$. Their cumulative distribution functions are denoted $F_i(\cdot)$, with density function $f_i(\cdot)$, mean μ_i and standard deviation σ_i. I make the following assumptions concerning their support:

$$h_{\min}^A + h_{\min}^B \geq 1 + 2\alpha \tag{1.1}$$

$$h_{\max}^i < 1 \quad \forall\ i. \tag{1.2}$$

As will become clear in the next subsection, this limits attention to ranges of public good benefits that are (i) jointly always greater than the costs, (ii) large enough to justify costly negotiations ex post, and (iii) individually small enough to necessitate joint investment.

Two types of negotiating occur within the long-term relationship. There are *ex ante* negotiations on the non-binding rule, and in each period t there may be *ex post* negotiations on contributions in that period.[3] The model does not specifically describe the bargaining process. Instead, both ex ante and ex post, agreement formation is captured by implementing the Nash bargaining solution, and the parties are assumed to perfectly foresee this outcome. I do not allow direct payoff transfers, or side payments, between the parties.

Within the model, parties can commit to contributions they negotiate in the current period, but cannot bind themselves to future contributions. If one party, ex post, decides to *opt out* of the burden sharing agreement, because it would rather not contribute or renegotiate towards a smaller share of the burden, there is no court that can prevent this. As an example of such limited ability to commit, think of national governments negotiating with each other about annual financial contributions to an international project. From one year to the next, a country may change governments, and the current administration generally does not have the legislative power to bind its successor to a foreign policy. Even when there is no change in government, the circumstances one year from now may warrant a reconsideration of the contribution level. These problems are avoided with contributions that are made immediately after reaching agreement.

After observing the state of the world, the parties know what to expect from renegotiations, which informs their decision between sticking to the pre-agreed BSR and opting out. Regarding that decision, the analysis in the current and in the next section will be focused exclusively on strategies that correspond to the one-shot Nash equilibrium. In section 1.4 I consider alternative equilibria, in which the parties employ trigger strategies to establish a wider range of circumstances under which they stick to the rule. These equilibria, while subgame perfect, are vulnerable to renegotiation when the continuation payoffs after a

[3] I do not consider a bargaining cost that is specific to ex ante negotiations on the rule, say α_r. This is not entirely without loss of generality, but incorporating an extra cost α_r would not affect any of the model's qualitative results.

deviation are not Pareto efficient ex ante. I show that constructions involving asymmetric punishment do not provide a way around this in the current setting.

1.2.1 The Stage Game

Contribution effects are assumed to be symmetric, additively separable and linear. Total capacity $Q = \sum_i q^i$ is constrained at a maximum of one per period. The stage game payoff function that party i seeks to maximize is given by

$$\pi_t^i(q) = h^i Q - q^i \quad \text{with} \quad Q = \min\{\sum_i q^i\ ,\ 1\}. \tag{1.3}$$

All uncertainty affecting t is resolved at the start of the period. After observing $\{h_t^A, h_t^B\}$, the parties' decisions determine their contribution levels for that period, producing payoffs $\{\pi_t^A, \pi_t^B\}$. If both parties consent to honoring the pre-agreed rule, contributions are made and the resulting payoff is given by

$$\pi_t^i(q_r) = h_t^i Q_r - q_r^i. \tag{1.4}$$

Should one of the parties decide to opt out of the agreement, (costly) renegotiations can take place, but only if both parties agree to this. If either party does not want to renegotiate, the parties will choose their contribution levels in a non-cooperative manner, separately setting q^i to maximize (1.3). It is straightforward to see that as long as $h^i < 1$, this means the parties will contribute zero, and payoffs will be zero:

$$\pi_t^i(q^*) = 0. \tag{1.5}$$

If the parties do meet and negotiate new contribution levels, they incur bargaining cost α. Upon agreeing to a new burden division $\{q_t^A, q_t^B\}$, the parties contribute accordingly, resulting in the following payoff for party i:

$$\pi_t^i(q_t) = h_t^i Q_t - q_t^i - \alpha. \tag{1.6}$$

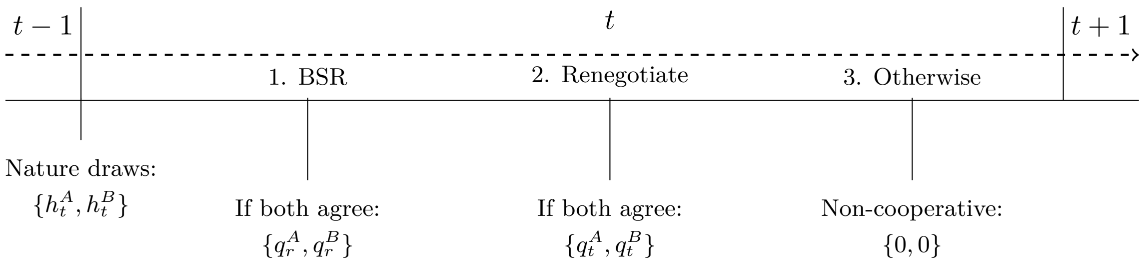
The sequence of events and decisions is summarized by Figure 1.1. First, the state of the world is observed by both parties. They then decide how to contribute, where each party has the unilateral power to decline the stated option and move on to the next.

1.2.1.1 Ex Post Negotiations

The zero contribution levels and payoffs, $\{\pi_t^A(q^*), \pi_t^B(q^*)\}$, serve as the threat point for ex post negotiations on contributions $\boldsymbol{q_t}$. The resulting contribution levels maximize the Nash product[4] of their respective payoff gains relative to that threat point:

[4]Bargaining strength is assumed to be symmetric, i.e., equal on both sides. This assumption is relaxed in section 1.3.3.

Figure 1.1: Sequence of events in t



$$\boldsymbol{q_t} = \underset{\boldsymbol{q_t}}{\arg\max}\ (h_t^A Q_t - q_t^A) \times (h_t^B Q_t - q_t^B) \tag{1.7}$$

Note that neither the negotiation costs nor the future contributions enter the Nash product, as these elements are independent of the negotiation outcome $\boldsymbol{q_t}$. The cost α is incurred regardless of whether or not the negotiations are successful, while the set of feasible agreements on future contributions is not impacted by any agreement, or disagreement, on contributions in t. The resulting agreement $\boldsymbol{q_t}$ is described by the following Lemma.

Lemma 1.1. *When the parties negotiate contributions ex post, they jointly invest up to the capacity constraint. Individual contributions depend on the relative public good benefits, and the resulting payoffs are equal for both parties:*

$$q_t^A + q_t^B = 1, \tag{1.8}$$

$$q_t^i = \frac{1}{2} + \frac{h_t^i - h_t^j}{2} \qquad \forall\ i, \quad j \neq i, \tag{1.9}$$

$$\pi_t^A(q_t) = \frac{h_t^A + h_t^B - 1}{2} - \alpha \ = \pi_t^B(q_t). \tag{1.10}$$

The proof is provided in Appendix 1.6.1. We can see from (1.9) that the difference between individual contributions matches that in public good benefits. This reflects that a smaller benefit h_t^i gives party i a better bargaining position, as does a greater benefit h_t^j for the other party. Note that the payoffs given by (1.10) are never below zero given the assumption on the smallest possible benefits. This means that both parties will always be willing to commence negotiations.[5]

[5]Section 1.3.2 explores a setting where this may not be the case in all periods.

1.2.1.2 Self-enforcing Range

In period t, the BSR will by honored only if neither party opts out. The parties make this decision having observed $\{h_t^A, h_t^B\}$. Anticipating that party j will always consent to renegotiations, party i knows it can achieve payoff $\pi_t^i(q_t) \geq 0$ by opting out. Therefore, in order for the agreement to be honored, we need $\pi_t^i(q_t) \leq \pi_t^i(q_r)$ for both i. By describing the range of public good benefits for which this holds, we can identify the self-enforcing range (SER) of a given BSR. For notational simplicity, I will use Lemma 1.1 to express the renegotiated contributions as a single variable q_t that denotes the contribution by party A: $\boldsymbol{q_t} = \{q_t^A, q_t^B\} = \{q_t, 1 - q_t\}$. The same type of notation can be applied to the pre-agreed contributions, as any pareto optimal BSR must specify contributions such that $q_r^A + q_r^B = 1$ (see Appendix 1.6.2 for the proof). I will therefore use q_r to denote the BSR contribution for party A: $\boldsymbol{q_r} = \{q_r^A, q_r^B\} = \{q_r, 1 - q_r\}$.

In deciding whether or not to opt out, the parties face a choice between the renegotiation payoff given by (1.10) and the following 'BSR payoffs', respectively:

$$\pi_t^A(q_r) = h_t^A - q_r, \tag{1.11}$$

$$\pi_t^B(q_r) = h_t^B - (1 - q_r). \tag{1.12}$$

The payoff-maximizing choice is determined by the difference in public good benefits, i.e., by the parties' bargaining positions, as is described by Lemma 1.2.

Lemma 1.2. *The difference in public good benefits determines whether the parties stick to the rule. Neither will opt out to trigger renegotiations if this difference falls in the following range, defined by* $\boldsymbol{q_r}$*:*

$$2q_r - 1 - 2\alpha \leq h_t^A - h_t^B \leq 2q_r - 1 + 2\alpha \tag{1.13}$$

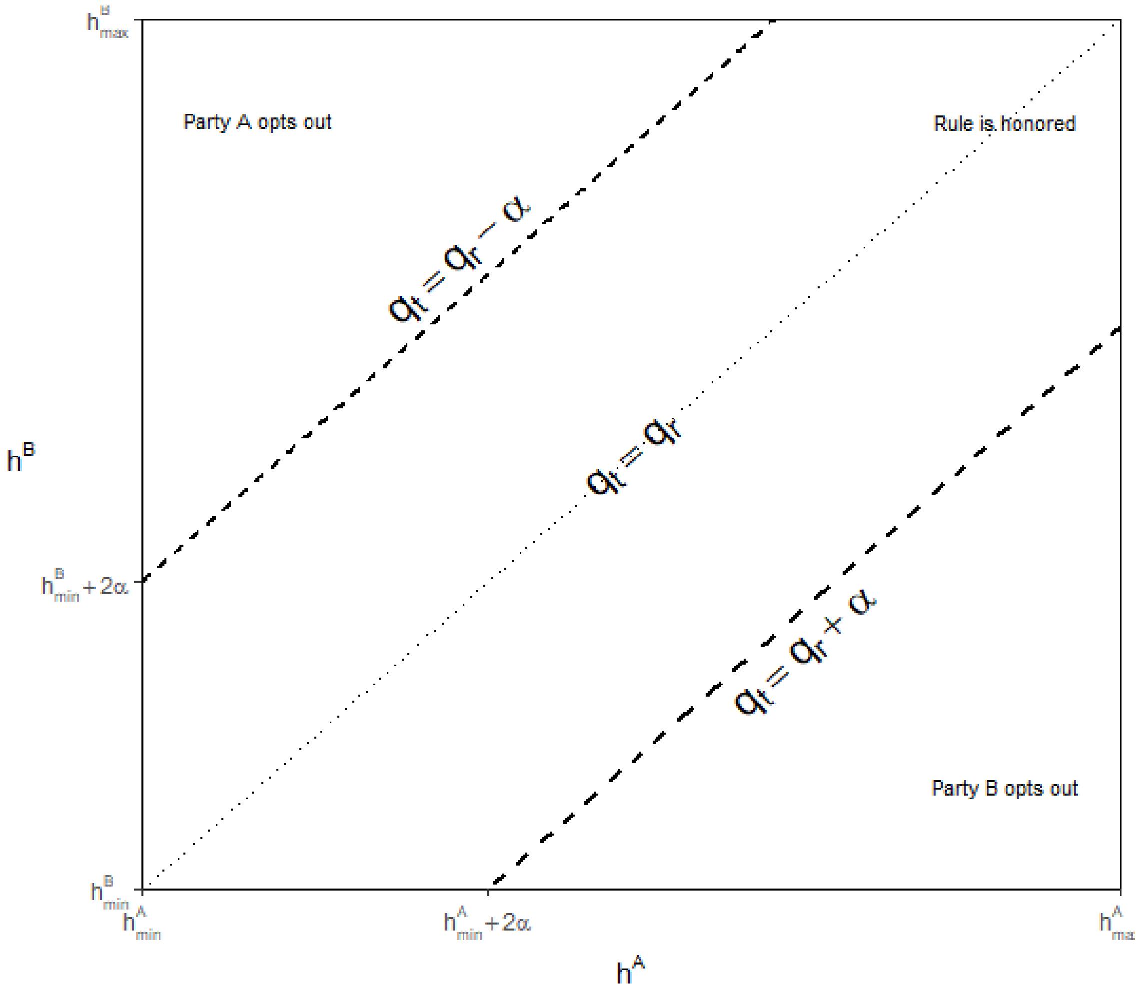
This condition, that determines the SER, is equivalent to $\pi_t^i(q_r) \geq \pi_t^i(q_t)$ for both i. The difference in benefits $[h_t^A - h_t^B]$ captures the bargaining positions in period t: the smaller (larger) this difference is, the stronger the bargaining position for party A (B). In periods where there is an unusually strong bargaining position for one of the parties, that party will opt out of the BSR to exploit this position. We can rewrite 1.13 in terms of BSR contributions and how they compare to the q_t that would result from renegotiations:

$$q_r^L = q_r - \alpha \leq q_t \leq q_r + \alpha = q_r^H. \tag{1.14}$$

Here, q_r^L and q_r^H are determined by q_r, and they represent the boundary levels of q_t for which the agreement is honored. These boundaries are further apart the higher negotiation cost α.

Given the BSR, we now know whether or not the agreement will be honored for any

Figure 1.2: BSR Self-Enforcing Range



combination of benefit factors $\{h_t^A, h_t^B\}$. Figure 1.2 graphically illustrates this. The area between the thick dashed lines represents the range in which the rule will be honored. The diagonal distance between these two parallel lines depends on the exogenous α, while their positions are determined by the choice of q_r. The figure reflects that for neither party to opt out, the bargaining positions should not deviate too much from those are consistent with the pre-agreed rule: $q_t = q_r$.

1.2.2 Pre-negotiating the Burden Sharing Agreement

Having determined what will happen in t for a given state $\{h_t^A, h_t^B\}$, and how payoffs will depend on q_r, we can turn to the ex ante negotiations on the burden sharing agreement. Each party enters these negotiations with the objective of maximizing the BSR's expected value to them. Accordingly, the bargaining solution for q_r maximizes the Nash product of the gains in expected payoffs generated by the rule. The threat point for these negotiations is having no BSR. The outcome of the negotiations does not affect the parties' capacity to negotiate or renegotiate a BSR in the next period. For the bargaining solution, I will therefore apply a duration of one period to the difference between agreement and threat point.

In the remainder of this section, I will describe how the expected value depends on the design of the rule, and then derive the bargaining solution. Before doing so, however, it is helpful to note the following: The ex post bargaining solution for q_t is a function of the realizations h_t^A and h_t^B, so that we can define a distribution F_q over q_t that maps h_t^A and h_t^B onto $F_q(q_t)$:

$$q_t = \frac{1}{2} + \frac{h_t^A - h_t^B}{2}. \tag{1.15}$$

As a linear combination of the two independent distributions F_A and F_B, F_q is also unimodal and symmetric, while its variance σ_q^2 is given by:

$$\sigma_q^2 = \frac{\sigma_A^2 + \sigma_B^2}{4}. \tag{1.16}$$

1.2.2.1 Expected Value of the Agreement

The expected value of the BSR depends on what contributions $\boldsymbol{q_r}$ it specifies, for two reasons. First, it affects the individual payoffs when the rule is honored, by prescribing who contributes how much. Second, it determines under what range of circumstances neither party opts out.

For the rule's ex post value in t, denote V_t^i as the difference in payoff between having and not having a BSR. As opting out is free of costs, the agreement will have zero value if not honored. We can therefore limit attention to the SER, where both parties decide stick to the BSR, and its value is given by

$$V_t^i = \pi_t^i(q_r) - \pi_t^i(q_t) = \frac{h_t^i - h_t^j + 1}{2} + \alpha - q_r^i. \tag{1.17}$$

Note that Lemma 1.2 ensures that V_t^i never falls below zero, and that the sum of the individual values is equal to the avoided negotiation costs, $V_t^A + V_t^B = 2\alpha$. We can rewrite (1.17) to obtain the per period BSR values to each party as a function of q_t and q_r:

$$V_t^A = q_t - q_r + \alpha \tag{1.18}$$

$$V_t^B = q_r - q_t + \alpha$$

The parties know with what probability every q_t occurs, and can thereby form expectations about V_t^i as a function of q_r. This gives the expected increase in next period's payoff that the parties achieve with the agreement:

$$EV^A = \int_{q_r-\alpha}^{q_r+\alpha} q_t - q_r + \alpha \ \ dF_q(q_t), \tag{1.19}$$

$$EV^B = \int_{q_r-\alpha}^{q_r+\alpha} q_r - q_t + \alpha \ \ dF_q(q_t),$$

where $EV^i = E_t(V^i_{t+1}(q_r))$.

1.2.2.2 The Bargaining Solution

While the cost division is zero-sum in terms of payoff, the parties have a common interest in minimizing the probability of renegotiation. Both elements are relevant for the bargaining solution, which maximizes the Nash product of its expected value to A and B for the next period:[6]

$$q_r = \arg\max_{q_r} \quad (EV^A) \times (EV^B). \tag{1.20}$$

This gives us the following proposition.

Proposition 1.1. *The BSR reflects the expected negotiation outcome, depending only on the mean benefit factors::*

$$q_r = \mu_q = \frac{\mu_A - \mu_B + 1}{2} = E[q_t]. \tag{1.21}$$

The prescribed contributions $\boldsymbol{q_r}$ *do not depend on the variances in benefit factors. The value of the BSR to both parties depends only on total* $\sigma_A^2 + \sigma_B^2$*, not on individual variance.*

Appendix 1.6.3 contains the complete proof. By maximizing a product of values, the Nash bargaining solution represents a balance between maximizing the sum of these values and minimizing the difference between them. This can be interpreted as a combination of efficiency and equity, compromising between the two if necessary.[7] In this case, the solution $q_r = \mu_q$ generates the maximum total value, which is shared exactly evenly between the parties.

It is clear from (1.18) that whenever honored, the BSR will create the same total extra payoff: the total avoided negotiation costs 2α. This is true irrespective of q_r. The greatest total expected value of the agreement is therefore achieved by the *least-renegotiated rule*, i.e., the agreement that maximizes the probability that neither party opts out ex post. This probability is given by: $F_q(q_r + \alpha) - F_q(q_r - \alpha)$.

[6] To be more precise: the discounted expected value to A and B, (δEV^A) and (δEV^B). Without the possibility for transfers at $t = 0$, however, time preferences have no effect on the bargaining solution as they are factored out of the Nash product.

[7] Efficiency can, in part, be traced back to Nash's Pareto-Optimality axiom. The equity aspect is not clearly represented in the axioms, although the symmetry requirement implies a degree of fairness. Trockel (1996, 2006) relates the underlying fairness of the Nash solution to that of the Walrasian equilibrium, and explores its links to different representations of the solution itself. The tradeoff between efficiency and fairness is studied in more detail by Rachmilevitch (2015).

The unimodality of distribution F_q implies that to minimize the occurence of renegotiations, the rule should equate the probability density at the boundaries of the self enforcing range:

$$f_q(q_r + \alpha) = f_q(q_r - \alpha). \tag{1.22}$$

For a symmetric distribution, this is achieved at $q_r = \mu_q$.

To understand how the rule's value is divided between the parties, note that the distribution of extra payoff in one period varies (in a linear fashion) from one end of the self-enforcing range to the other. At the lower boundary $(q_r - \alpha)$, where party A is indifferent between opting out and sticking to the rule, party B has all the benefit, while at the upper boundary $(q_r + \alpha)$, the full extra 2α payoff goes to party A. How this converts to individual expected values EV^i, depends on how probability is distributed within the self enforcing range. At $q_t = q_r = \mu_q$, the extra payoff is exactly α to both parties. The probability distribution for q_t within the self-enforcing range $[q_r - \alpha,\ \ q_r + \alpha]$, is symmetric around its center $q_r = \mu_q$. This solution therefore results in an equal split of expected value from the agreement:

$$EV^A = EV^B = \alpha \int_{\mu_q - \alpha}^{\mu_q + \alpha} dF_q(q_t). \tag{1.23}$$

While this is the expected value for one period, an important motivation for establishing a rule would be to avoid negotiations for a longer time. Given that future benefits are not affected by past realizations or actions, at any date $t' > t$ there is no new information about future payoffs (i.e., payoffs in periods after t'), compared to the information available at date t, that could alter the outcome of negotiations on $\boldsymbol{q_r}$. The rule, once established, is therefore never renegotiated itself. Similarly, if the parties do not agree on a BSR in t, there is no reason why they would in any future period $t' > t$.

Compared to never agreeing on a rule, the discounted stream of expected extra payoffs it generates is $\frac{\delta}{1-\delta} EV^i$. Although discount factor δ does not affect the bargaining solution, it does impact this total expected value for the BSR, which one can view as an indicator of how important it is for the parties to establish such agreement. Similarly, while the rule does not depend on the variance of the public good benefits, σ_A and σ_B do determine the variance in F_q, which negatively relates to the probability that renegotiation is avoided. From (1.23), it is clear that this probability and the negotiation cost α determine EV^i. We know the variance in bargaining positions, captured by q_t, is $\sigma_q^2 = \frac{1}{4}(\sigma_A^2 + \sigma_B^2)$. Accordingly, an increase in variance on either party's benefit will symmetrically reduce the ex ante value of a given BSR to both parties. Also note that this value depends only on the sum of variances, not on how this total variance is distributed between the two parties' individual benefits. Both these characteristics relate to the fact that party A's ex post bargaining position depends as much on A's public good benefit as it does on B's benefit, and vice versa.

1.3 Extensions

This section will examine whether and how the bargaining solution $q_r = E[q_t]$ changes when several assumptions are relaxed. In section 1.3.1, I allow for risk averse preferences and discuss how the burden sharing rule affects allocation of payoff risk. Section 1.3.2 discusses the assumptions regarding the distributions of h^A and h^B, analyzing the special case where they are fully correlated and they have no lower limit. The effects of asymmetry in bargaining strength are investigated in section 1.3.3, while 1.3.4 introduces asymmetry in the negotiation cost α_i.

Note that these subsections do not build on each other, but rather relax the assumptions one at the time.

1.3.1 Risk Aversion

For the baseline case, the BSR simply reflects the expected future bargaining outcome, and its only economic effect is that with positive probability, it allows the parties to avoid costly negotiations ex post. There are, however, effects on the total level and potentially on the allocation of payoff risk.

Without a BSR, parties negotiate contributions every period, obtain equal payoffs, and thereby fully share the risk. For each party, the individual payoff is as given by (1.10), and its variance is $\frac{\sigma_A^2+\sigma_B^2}{4}$. However, across all periods in which the parties contribute $\boldsymbol{q_r}$, their individual payoffs only vary with their individual benefits. In the extreme case where parties always pay a fixed contribution and renegotiations never occur, an obvious example being the case of a legally binding cost-sharing contract, the variance in the payoff π^i of party i would be equal to σ_i^2, the variance in the individual benefit factor. Between them, the parties clearly experience more payoff variation when they never renegotiate than when they negotiate every period:

$$\sigma_A^2 + \sigma_B^2 \quad > \quad \frac{\sigma_A^2+\sigma_B^2}{4} + \frac{\sigma_A^2+\sigma_B^2}{4} = \frac{\sigma_A^2+\sigma_B^2}{2}$$

The difference between these opposite cases is informative. The BSR that is honored with some positive probability shifts the situation from the prior, i.e., fully flexible contributions, in the direction of the latter, i.e., fixed contributions, affecting payoff risk in two ways. First, it increases total payoff variance, by preventing risk sharing within the rule's self-enforcing range. Second, it can affect the risk allocation between parties, as party i's own benefit variance σ_i^2 becomes more important in determining i's individual payoff risk.

When parties care not only about expected payoff, but also dislike risk, these effects have to be taken into account. Firstly, when both parties face the same benefit variance, the BSR will increase payoff risk for both, which makes it less attractive. Secondly, if one party's variance in h^i is greater than the other's, the BSR will impact payoff risks asymmetrically. This affects their bargaining positions and thereby the rule itself, as is illustrated by the following special case, in which one party's individual benefit does not vary at all.

Assume that party B has a fixed benefit factor $\hat{h}^B$, while h_t^A varies. To allow for risk aversion, let i's utility in t be a (common) function u of π_i, which we have defined as a linear combination of its contribution, benefit factor and potential negotiation cost.

$$u_t^i = \begin{cases} u(h_t^i - q_t^i - \alpha) & \text{outside the self-enforcing range} \\ u(h_t^i - q_r^i) & \text{within the self-enforcing range,} \end{cases} \tag{1.24}$$

where $u' > 0$, $u'' < 0$ and $u''' \geq 0$.

The per-period negotiation outcome q_t is no different than under risk-neutrality, as all uncertainty about the benefits is resolved before those negotiations commence:

$$q_t = \frac{1}{2} + \frac{h_t^A - \hat{h}^B}{2}, \tag{1.25}$$

The corresponding expected BSR values are based on the difference between the utilities in (1.24):

$$EV^A = \int_{\hat{h}^B+2q_r-1-2\alpha}^{\hat{h}^B+2q_r-1+2\alpha} u(h_t^A - q_r) - u(h_t^A - q_t - \alpha) \;\; dF_A(h_t^A), \tag{1.26}$$

$$EV^B = \int_{\hat{h}^B+2q_r-1-2\alpha}^{\hat{h}^B+2q_r-1+2\alpha} u(\hat{h}^B - 1 + q_r) - u(\hat{h}^B - 1 + q_t) \;\; dF_A(h_t^A).$$

Using these expressions for EV^i, it can be shown that the Nash bargaining solution implies $q_r < \mu_q$ (see Appendix 1.6.4). The BSR will therefore specify a higher contribution by party B than would be the case under risk-neutrality. This compensates for the fact that, in all t where the parties stick to the BSR, B has a certain payoff, while party A absorbs all the payoff risk.

One can argue that this impact of individual benefit risk asymmetry on the BSR will also be there when both parties have variance in h^i. We know that, whenever it is honored and $\boldsymbol{q_r}$ is implemented without costly negotiations, the rule generates a total extra payoff of 2α. In any one period, the party that contributes less under the rule than it would after renegotiations ($q_t^i > q_r^i$), takes the larger share of that extra payoff ($V_t^i > \alpha$), and vice versa (see equation (1.18)). Not only does this party avoid the cost α, it also avoids a larger contribution. With unequal benefit variance, the expected share of the 2α that goes to the party with the greater individual spread is larger in periods where the aggregate benefits $h_t^A + h_t^B$ are relatively large, and smaller in periods where they are relatively small.

To a risk averse party, the utility gain from a given increase in payoff is greater when the initial payoff is small. Asymmetry in benefit risk therefore leads to unequal BSR-valuations when both parties are risk averse. The bargaining solution compensates the party with the greater individual variance. This enhances equity, but can also partially be explained as an efficiency gain relative to $q_r = \mu_q$, in terms of utility. While the BSR that reflects the mean negotiation outcome μ_q is still the least-renegotiated rule, the described deviation can result in a utility gain, by shifting the BSR-induced payoff increase towards situations where the total surplus is smaller. There may be less payoff increases overall, but they will occur at lower initial payoff levels.

1.3.2 Correlated Benefits

1.3.2.1 Implications for the Burden Sharing Rule

For a given rule q_r, correlation ρ between benefits affects the BSR's value, because the variance of q_t depends on it:

$$\sigma_q^2 = \frac{\sigma_A^2 + \sigma_B^2 - 2\rho\sigma_A\sigma_B}{4}. \tag{1.27}$$

When positively correlated, the benefit factors will 'move together', making large differences less likely. As it is the difference in benefits that determines whether or not one of the parties will want to renegotiate contributions, positive correlation will increase the probability of the BSR being honored. A negative correlation ($\rho < 0$) would have the opposite effect: large swings in bargaining positions will increase the frequency of renegotiation.

An interpretation of this is that the stability of a rule, and how likely parties are to establish one, depends on what function a public good or institution has to its contributors. If this is the same across parties, then there may well be common exogenous factors that can create a greater or smaller need for these functions (e.g., the weather or the economic climate). This would result in stable relative bargaining positions and low renegotiation frequency, even for large variance in the absolute benefits.

The bargaining solution $q_r = \mu_q$ relies on q_t being symmetrically distributed around μ_q, which it may not be when h^A and h^B are not independent. Without distributional symmetry, equity and efficiency are not achieved by the same q_r. Instead, the bargaining solution will be a compromise between the efficient rule and that which equates individual expected payoff gains. The least-renegotiated rule is efficient and will still be characterized by (1.22), equating the marginal probability density at the boundaries of the SER. That rule will, however, not be of equal expected value to both parties.[8]

The assumptions on $h^i_{\min}$ imply that the bargaining costs are never large enough to deter renegotiations, ruling out a period t where the public good is not provided at all. Positive correlation between the individual benefits makes a violation of this assumption

[8] The least-renegotiated rule will now depend on α, whereas under symmetry, $q_r = \mu_q$ satisfies condition (1.22) for any α.

more realistic. In the following I will explore the possibility of benefits dropping below this threshold, by looking at the special case of perfectly correlated benefits.

1.3.2.2 Allowing for Smaller Surplus

Public good provision is guaranteed in every period, whether through BSR- or renegotiated contributions, when $h^A + h^B > 1 + 2\alpha$. For this range of benefit factors, the main value of the BSR lies in avoiding negotiation costs. For benefits below that threshold, however, it could facilitate the public good provision itself. This is the case when the benefits are jointly still greater than one, but the cost of negotiations prevents the parties from initiating them, and they instead default to zero investment.

To analyze this aspect, I consider fluctuations in joint benefits rather than in relative bargaining positions. Assume the parties' benefit factors are subject only to a common shock, and are thereby fully correlated:

$$h_t^i = \hat{h}^i + \theta_t, \tag{1.28}$$

where $\hat{h}^i$ is a constant individual benefit and θ_t refers to the common shock in period t. If we assume the latter to have a distribution F_θ, with a zero mean, we can say that $\hat{h}^i$ is the mean individual benefit for i. As deviations from it occur simultaneously, the difference between benefit factors is constant over time, and bargaining positions never change. Accordingly, the (re)negotiation contribution levels for one period will always have a certain outcome:

$$q_t = \frac{\hat{h}^A - \hat{h}^B + 1}{2}. \tag{1.29}$$

With q_t fixed, we know that for any q_r that satisfies

$$q_t - \alpha \leq q_r \leq q_t + \alpha, \tag{1.30}$$

neither party will be able to improve their payoff by entering costly renegotiation in any period t. Any such BSR would only be canceled when either or both parties have a benefit lower than their respective contribution, and they opt out to avoid negative payoff. The bargaining solution for the rule is $q_r = q_t$ (see Appendix 1.6.5 for the proof). The value V_t^i of the BSR in period t depends on the total surplus. As before, this value is defined by the extra payoff achieved with the rule, but we must now take into account the possibility that ex post negotiations do not occur: $V_t^i = \pi_t^i(q_r) - \max\{\pi_t^i(q_t), \pi_t^i(q^*)\}$. For $q_r = q_t$,

$$\pi_t^i(q_r) = \begin{cases} \frac{1}{2}(\hat{h}^A + h^B - 1) + \theta_t, & \text{for} \quad \theta_t \geq \frac{1}{2}(1 - \hat{h}^A - \hat{h}^B) \\ 0, & \text{for} \quad \theta_t < \frac{1}{2}(1 - \hat{h}^A - \hat{h}^B), \end{cases}$$

and

$$\max\{\pi_t^i(q_t), \pi_t^i(q^*)\} = \begin{cases} \frac{1}{2}(\hat{h}^A + h^B - 1) + \theta_t - \alpha, & \text{for} \quad \theta_t \geq \frac{1}{2}(1 + 2\alpha - \hat{h}^A - \hat{h}^B) \\ 0 & \text{for} \quad \theta_t < \frac{1}{2}(1 + 2\alpha - \hat{h}^A - \hat{h}^B), \end{cases}$$

so

$$V_t^i = \begin{cases} \alpha, & \text{for} \quad \theta_t \geq \frac{1}{2}(1 + 2\alpha - \hat{h}^A - \hat{h}^B) \\ \frac{1}{2}(\hat{h}^A + h^B - 1) + \theta_t, & \text{for} \quad \frac{1}{2}(1 - \hat{h}^A - \hat{h}^B) \ < \ \theta_t \ < \ \frac{1}{2}(1 + 2\alpha - \hat{h}^A - \hat{h}^B) \\ 0 & \text{for} \quad \theta_t < \frac{1}{2}(1 - \hat{h}^A - \hat{h}^B). \end{cases} \tag{1.31}$$

For larger surplus, the rule avoids costly negotiation. For a positive surplus smaller than 2α, the rule is necessary for the public good to be provided at all. When the sum of the benefits drops below one, the BSR has no added value. The increase in expected payoff EV^i that results from the rule $q_r = q_t$ is equal for both parties. Like for the ex post value V_t^i, the different ranges of θ_t have to be taken into account:

$$\begin{aligned} EV^A = EV^B = & \int_{\frac{1}{2}(1-\hat{h}_A-\hat{h}_B)}^{\frac{1}{2}(1+2\alpha-\hat{h}_A-\hat{h}_B)} \frac{\hat{h}_A + \hat{h}_B - 1}{2} + \theta_t \ \ dF_\theta(\theta_t) \\ & + \ \alpha \left[1 - F_\theta\left(\frac{1 + 2\alpha - \hat{h}^A - \hat{h}^B}{2}\right)\right]. \end{aligned} \tag{1.32}$$

The first part of the above expression shows the expected value for smaller surplus, V_t^i integrated over the intermediate rate for θ_t, as described by (1.31). The second part covers the range of greater surplus, where the value is always exactly α, and is weighted for the probability that θ_t is aboved the derived threshold.

1.3.3 Asymmetric Nash Bargaining

The analysis so far has relied on the assumption that when two parties bargain, whether over per-period contributions or over the long term agreement, they have equal power. The Nash products places equal weight on both parties' (expected) improvement over their respective disagreement payoffs. It has been mentioned that the resulting rule reflects consideration for equity. A more general interpretation would be that this bargaining solution reflects the division of power, or bargaining strength, which it assumes to be symmetric.

If, instead, we allow for inequality in bargaining power, we may have a weight γ on party A's improvement, and a weight $(1 - \gamma)$ on party B's improvement. To see what happens as γ deviates from $1/2$, we need to take into account how this changes both $\boldsymbol{q_t}$ and $\boldsymbol{q_r}$. For the ex post negotiation on contributions in t, the size of 'the pie' (total surplus) is fixed, in which case the weight γ simply tells us what share of that pie is taken by party A. The bargaining solution for q_t will maximize the weighted Nash product:

$$\max_{q_t} \quad (h_t^A - q_t)^\gamma \times (h_t^B + q_t - 1)^{1-\gamma}, \tag{1.33}$$

which gives

$$q_t = h_t^A - \gamma(h_t^A + h_t^B - 1). \tag{1.34}$$

The surplus $h_t^A - q_t$ that A obtains is the γ-share of the total surplus $h_t^A + h_t^B - 1$. The resulting payoffs are now unequal (for $\gamma \neq \frac{1}{2}$):

$$\begin{aligned} \pi_t^A(q_t) &= \gamma(h_t^A + h_t^B - 1) - \alpha \\ \pi_t^B(q_t) &= (1-\gamma)(h_t^A + h_t^B - 1) - \alpha. \end{aligned} \tag{1.35}$$

This more general expression of the negotiation payoffs also implies we need to generalize the assumption on the $h_{\min}^i$'s, to still guarantee that the parties are always willing to negotiate contributions:

$$h_{\min}^A + h_{\min}^B > 1 + \alpha \cdot \max\{\frac{1}{\gamma}, \frac{1}{1-\gamma}\}. \tag{1.36}$$

The initial assumptions, described by equations (1.1) and (1.2), implied an assumption that $\alpha < \frac{1}{2}$, which means the public good itself never costs more than the negotiations that determine how it is paid for. Marginal changes in γ, away from $\frac{1}{2}$, marginally change this implied assumption. For larger deviations, i.e., for more extreme asymmetry, the assumption becomes more restrictive and perhaps less reasonable. I will therefore first focus on the BSR under moderate asymmetry, after which I will discuss the case of extremely unequal bargaining power separately.

Having redetermined q_t as a function of the two benefit factors, the rule's value EV^i can be expressed in the same function of q_r^i and q_t^i integrated over its self-enforcing range: $EV^i = \int_{q_r-\alpha}^{q_r+\alpha} q_t^i - q_r^i + \alpha \;\; dF_q(q_t)$. A BSR that reflects the expected (re)negotiation outcome,

$$q_r = \mu_q = \mu_A - \gamma(\mu_A + \mu_B - 1) = E[q_t], \tag{1.37}$$

would minimize the probability of renegotiation, thereby maximizing the sum of BSR values. This favors the stronger bargainer relative to the rule under symmetric bargaining strength. However, it is not the bargaining solution for $\gamma \neq \frac{1}{2}$.

The bargaining solution for the BSR would also maximize a weighted product:

$$\max_{q_r} \quad (EV^A)^\gamma \times (EV^B)^{1-\gamma}. \tag{1.38}$$

Proposition 1.2. *For small to moderate deviations from symmetric bargaining strength, the BSR will favor the stronger bargainer beyond the expected advantage when negotiating per period.*

The proof is given in Appendix 1.6.6. It shows that the least-renegotiated rule, $q_r = \mu_q$, which yields equal EV^A and EV^B, does not satisfy the first-order condition for (1.38) if $\gamma \neq \frac{1}{2}$, and its derivative indicates the solution to have lower contributions for whoever has the greater bargaining power. Remember that the rule, when honored, creates a total value of 2α. The stronger bargainer will take more than half of this value, and does so by negotiating a lower q_r^i.

This means the advantage from being the stronger bargainer is amplified when the parties negotiate a BSR. In case there is a threshold BSR value below which it is not worth negotiating, such asymmetry in bargaining strength might jeopardize its establishment. Expecting an amplified disadvantage when negotiating the rule, the weaker bargainer could decide not to enter these negotiations in the first place.

It could therefore be beneficial, to both parties, if the stronger party was able to commit to not exploiting its bargaining advantage in the ex ante rule negotiations. To illustrate this, a simple example with deterministic benefit factors is provided in Appendix 1.6.7. It derives a range for γ in which the asymmetric bargaining strength prevents the establishment of a BSR. Both parties would then be better off with the least-renegotiated BSR specified in (1.37), but this is only a feasible outcome if the stronger bargainer can somehow commit to equal BSR-values *before* both parties decide whether or not to enter negotiations.

To describe what happens in case of extremely unequal bargaining strength, some more words on the nature of BSR-bargaining may be helpful. Compared to a legally binding contract, the fact that the BSR is not fully binding adds a dimension to the negotiations. As I have mentioned before, the contributions do not only determine the payoffs under the rule, but also the probability that it is actually implemented. Imagine the rule specifies a q_r that efficiently minimizes the probability of renegotiations. A marginal change in q_r will (i) increase the payoff for one party, (ii) decrease the payoff for the other party, and (iii) reduce the probability that is honored. The first two effects are zero-sum, the third effect hurts both parties. From the perspective of one individual party, starting from this initial q_r, the payoff increase from a marginal reduction in contribution will outweigh the increase in renegotiation probability.[9]

When we shift q_r further away from the efficient solution, there must be a point beyond which this is not the case anymore: a range of q_r for which the parties would both want to change the BSR in the same direction. This implies that the best possible rule for one party is not necessarily the worst possible rule for the other party, which informs the case of extremely asymmetric bargaining power: If one party has the power to simply choose q_r, this rule can still have value for the weak party, even though that party would obtain zero value from ex post negotiations.

[9]Note that the effect on this probability depends on the difference between $f_q(q_r + \alpha)$ and $f_q(q_r - \alpha)$, which is zero initially.

1.3.4 Asymmetric α

So far, the cost of negotiation α has been assumed to be the same for both parties. To relax this assumption, denote i's cost of (re)negotiation as α_i. As these costs are sunk once negotiations start, they do not affect the renegotiation outcome q_t.[10] The range of values for q_t for which the BSR is honored does change, however, for asymmetric α's. We can update the self-enforcing range, as given by (1.14), to become

$$q_r - \alpha_A \leq q_t \leq q_r + \alpha_B. \tag{1.39}$$

The expected values of the BSR in one period depend on this range, as well as on the individual negotiation cost:

$$EV_A = \int_{q_r-\alpha_A}^{q_r+\alpha_B} q_t - q_r + \alpha_A \ \ dF_q(q_t),$$

$$EV_B = \int_{q_r-\alpha_A}^{q_r+\alpha_B} q_r - q_t + \alpha_B \ \ dF_q(q_t).$$

Proposition 1.3. *Under asymmetric negotiations costs, the BSR favors the party with the lower α_i:*

$$q_r = \mu_q + \frac{\alpha_A - \alpha_B}{2}. \tag{1.40}$$

The proof is given in Appendix 1.6.8. It is shown that the BSR that results from Nash bargaining again is the least-renegotiated rule, equating $f_q()$ at the boundaries of the self-enforcing range. However, this is not achieved by the expected renegotiation outcome $q_r = \mu_q$, but requires adjustment to any asymmetry in negotiation costs:

$$f_q(q_r + \alpha_B) = f_q(q_r - \alpha_A), \tag{1.41}$$

We can interpret the lower contributions for the party with the lower negotiation cost as having two underlying causes. First, a lower α_i means that party i will need less (upwards) deviation from its expected bargaining position (in $t \geq 1$) to trigger renegotiations. The lower BSR contributions therefore reduce the renegotiation probability, benefiting both parties. Second, party i enjoys a stronger bargaining position at $t = 0$, as the outside option of not having a BSR is more costly to the other party. The expected value of this asymmetry-adjusted agreement is again identical for both parties: $EV^A = EV^B$.

1.4 Extending The Self-enforcing Range

The characterization of the burden sharing rule, both in the basic model and in the extensions of the previous section, has relied on the parties behaving in line with the one-shot

[10] The surplus is still split equally, but subtracting the negotiation costs does leave unequal payoffs. This makes the required assumption on the minimum surplus slightly stronger:
$h^A_{\min} + h^B_{\min} > 1 + 2 \cdot \max\{\alpha_A, \alpha_B\}$.

Nash equilibrium strategies, regarding when to opt out of the BSR and when to honor it. In this section, I will characterize a subgame perfect equilibrium with trigger strategies, in which the threat of punishment ensures that the parties never opt out. I will discuss how this extention of the self-enforcing range affects the pre-negotiated agreement on $\boldsymbol{q_r}$. I will then show that equilibria with a larger SER are vulnerable to renegotiation, by demonstrating that the required punishment schemes are either renegotiable, or not strong enough to deter deviations.

One can imagine the parties may cooperate across periods to extend the self-enforcing range of the rule beyond this. They could, for example, both employ a strategy where party i never opts out of the BSR as long as party $j \neq i$ has not opted out in any of the preceding periods, even when outside the range described in Lemma 1.2. Taking the basic model with solution $q_r = \mu_q$ as the starting point, this would increase the per-period expected value of the BSR to α, as the probability of opting out drops to zero. Such cooperation would be part of an agreement constructed in $t = 0$,[11] which must also specify what happens after a party does not comply at any point in time. Its sustainability depends on what one party could gain by deviating and opting out of the BSR, compared to the losses that party expects to face in subsequent periods as the other party retaliates.

What could the punishment for a deviation be? Any response involving a refusal to renegotiate contributions, or to contribute at all, in the deviation period or for a number of periods that follow, amounts to a non-credible threat,[12] and would therefore not be part of a subgame perfect equilibrium strategy. In an alternative punishment scheme that avoids this problem, party j responds to a deviation by i by renegotiating contributions in t, and for a duration of T subsequent periods, reverting to the ex post one-shot equilibrium strategy described by Lemma 1.2: opt out when $q_t^j > q_r^j + \alpha$. Party i does the same for T periods, such that the parties behave according to the one-shot Nash equilibrium for the duration of the punishment. For both parties, this reduces the ex ante value of the BSR, per punishment period, to:

$$EV^i = \int_{q_r^i - \alpha}^{q_r^i + \alpha} q_t^i - q_r^i + \alpha \;\; dF_q(q_t^i). \tag{1.42}$$

The payoff gain from opting out of the BSR in t, having observed the benefits h_i and h_j, is equal to the reduction in the own contribution, minus the cost of renegotiation. For the described punishment to deter such deviation for all possible q_t^i, the parties must choose T so that for $i \in \{A, B\}$

$$q_r^i - q_{\min}^i - \alpha \;\; \leq \;\; \sum_{t=1}^{T} \delta^t (\alpha - EV^i), \tag{1.43}$$

[11] If both parties can improve by setting up this cooperation in any $t > 0$, the same is true in $t = 0$, so it is safe to assume they would agree on this immediately when negotiating the BSR.

[12] It is still beneficial to both parties to negotiate new contributions, in period t and in all periods after.

where $q^i_{\min} = \frac{1}{2} + \frac{h^i_{\min} - h^j_{max}}{2}$.

Provided that δ and α are large enough to pick a T that satisfies (1.43), what does it mean to have a rule that is always honored? For a given q_r, it increases the value of the agreement, but it also further reduces the scope for sharing risk. On average, payoffs will be greater, but individually they will vary more from period to period, and under certain circumstances they could even drop below zero.

Qualitatively, what has been shown for both the basic model and for the extensions into several types of asymmetry does not change. For the case of risk aversion and asymmetric individual risk, the reduction in risk sharing further improves the position of the riskier party when negotiating the rule, so q_r would deviate further from μ_q. However, it should be noted that under these circumstances it is less likely that this type of cooperation across periods is an improvement. The potential for very small absolute payoffs will weigh heavily on the rule's value if the aversion to risk is strong. It might not be worth that much risk to save some more negotiation costs.

More generally, the elimination of renegotiation shifts the focus of BSR-negotiations away from efficiency. Within the limits for which the cooperative equilibrium is still subgame perfect, i.e., as long as the gains deter deviations, determining q_r becomes a matter of dividing the pie. For unequal costs to negotiation, as in section 1.3.4, the bargaining solution for the rule is unaffected: $q_r = \mu_q + \frac{\alpha_A - \alpha_B}{2}$. The pie is split equally, at $EV^A = EV^B = \frac{\alpha_A + \alpha_B}{2}$. For asymmetry in bargaining strength, there is still an amplified advantage for the stronger party, as described by section 1.3.3. When renegotiation is ruled out in equilibrium, the rule will deviate from μ_q, and more so for greater asymmetry and greater negotiation costs: $q_r = \mu_q + (1 - 2\gamma)\alpha$. Note that due to the smaller cooperation value for the weaker bargainer, a longer punishment phase T may be required to satisfy (1.43).[13]

As mentioned in section 1.2, the cooperative equilibrium described above is vulnerable to renegotiation. When the parties negotiate $\boldsymbol{q_r}$, as well as under what circumstances to stick to it, they cannot rule out meeting again in the future. They will not meet or negotiate again, as long as they never deviate from the plan. However, after a deviation, and perhaps also in the punishment phase that follows, they will be at the negotiation table once again. Since the punishment will hurt both the deviator and the punisher, both parties would prefer to restart the cooperation immediately and skip the punishment. Knowing that this penalty can be escaped by renegotiating makes the cooperation impossible in the first place, as nothing now deters a deviation. In different types of repeated games, it is conceivable that the cost to renegotiation helps avoid this issue, by preserving the threat of the punishment phase. An example of this is the Bertrand supergame considered by McCutcheon (1997). However, in the current setting, the punishment still involves the parties meeting to negotiate, at least once after a deviation, so that α will be incurred

[13] These solutions for q_r are derived in Appendix 1.6.9.1.

regardless and cannot prevent renegotiation of the punishment.

Phrased differently, the problem with extending the SER through such strategies is that expected payoffs over the periods following a deviation are Pareto-inferior to those resulting from a continued cooperation, where neither party ever opts out. The concept of (weak) renegotiation proofness (WRP) was proposed by Farrell and Maskin (1989) to rule out such equilibria. For a cooperative equilibrium to fit their concept, its punishment strategies should treat players asymmetrically, punishing the deviator but rewarding the punisher. Referring to the cooperative state as 'the normal phase' and the T periods after a deviation as 'the punishment phase', three requirements must be met:

(i) Party i must be deterred to cheat in the normal phase, i.e., the punishment must be of sufficient size.

(ii) Party j must (weakly) prefer punishing over the normal phase, so it cannot be convinced to skip the punishment.

(iii) Party i must be deterred to cheat in the punishment phase, i.e., it must be willing stick to the prescribed strategies.

To illustrate that these conditions cannot be satisfied simultaneously in the current setting, consider constructing such a punishment phase of duration T, starting from a potential deviation by party A.[14] Following the punishment construction of Farrell and Maskin (1989):

- In the normal phase, both parties will accept $\boldsymbol{q_r}$ for all possible q_t.
- After a deviation by party A, party B for T periods switches to the one-shot Nash equilibrium strategy: reject $\boldsymbol{q_r}$ when $q_t > q_r^H = q_r + \alpha$.
- During this punishment phase, party A sticks to the cooperative strategy: always accept $\boldsymbol{q_r}$.
- If party A deviates from the prescribed strategy during the punishment phase, it starts over.

Requirement (i) must hold for the largest possible deviation gain, which occurs at renegotiation outcome $q_t = q_{\min}$. At the time of the potential deviation, the expected value of the punishment amounts to foregone V_t^A, integrated over the range $q_t \in [q_r^H, q_{\max}]$, for which B now rejects q_r and forces renegotiation. The requirement is therefore

$$\sum_{t=1}^{T} \delta^t \int_{q_r^H}^{q_{max}} q_t - q_r + \alpha \;\; dF_q(q_t) \quad \geq \quad q_r - q_{\min} - \alpha, \tag{1.44}$$

[14]This is for notational convenience. The same could be illustrated for a deviation by party B.

and may be met for sufficiently large δ.

This punishment satisfies the necessary condition for the punisher to follow through. This means that at the time of negotiating in the period of the deviation, say t^d, and at the time of negotiating in any of the T periods that follow, party B cannot be convinced to skip the remaining periods of the punishment phase. For $t' \in \{t^d, t^d+1, t^d+2, ..., t^d+T\}$, the expected future reward[15] for the punisher is

$$\sum_{t=t'+1}^{T} \delta^{t-t'} \int_{q_r^H}^{q_{max}} q_t - q_r - \alpha \;\; dF_q(q_t), \tag{1.45}$$

which is nonnegative since $q_r^H = q_r + \alpha$.

The third requirement is the problematic one, and in the current setting is just a stronger version of (1.44). For all q_t, party A must be willing to stay in the punishment phase, i.e., it must weakly prefer punishing itself over deviating. Let us examine this for the first period after the deviation, since it must hold for all periods. Again, the deviation gain is at its maximum when $q_t = q_{\min}$, so the right-hand side of (1.44) is unchanged. The cost of deviating is that the punishment phase starts over. For the first $T-1$ periods after the deviation, this does not make a difference. The only difference is in the Tth period after the deviation, which turns from a 'normal' period into a 'punishment' period. The left-hand side of (1.44) now only consists of punishment in that period, giving

$$\delta^T \int_{q_r^H}^{q_{max}} q_t - q_r + \alpha \;\; dF_q(q_t) \geq \;\; q_r - q_{\min} - \alpha \tag{1.46}$$

or

$$\delta^T \int_{q_r^H}^{q_{max}} q_t - q_r^L \;\; dF_q(q_t) \geq \;\; q_r^L - q_{\min}. \tag{1.47}$$

Appendix 1.6.9.2 shows that this condition cannot be satisfied, even when $\delta = 1$ and in the extreme case that q_t is uniformly distributed on $[q_{\min}, q_{\max}]$, also for intermediate expansions of the self-enforcing range.

The successful construction of such a WRP equilibrium is not necessarily ruled out for any symmetric game, as is discussed by Farrell and Maskin (1989) and shown for the repeated prisoner's dilemma by van Damme (1989). There are several aspects of the setting considered here, that are specifically problematic for constructing strategies in such a way that the deviating party willingly punishes itself. First, the (maximum) deviation gain does not depend on the strategy of the punishing party - it is not reduced in the punishment phase relative to the normal phase. Second, in any period t, the deviation gain in that period is deterministic, as q_t is known. The punishment, however, materializes in the future and is therefore not only discounted, but also merely an expectation of a value, the range

[15] Note that at the time of negotiating, the reward in the current period has already been reaped, and the requirement is therefore such that the parties do not negotiate away from the punishment strategies in *future* periods.

for which only differs from the range of possible deviation gains by 2α. In other words, the punishment might not hurt at all, and when it does, its relative magnitude is not expected to be that great.

1.5 Conclusion

In this paper, I have studied two parties who have a long-term relationship, repeatedly co-funding a public good under changing circumstances. Even without the possibility to write a binding contract, an agreement on a burden division can be a valuable instrument to avoid repeated bargaining. This is of interest to the parties, particularly when bargaining is costly. Using a simple setting, I have shown how such a cost to negotiating can provide the long-term agreement with a self-enforcing range. By deriving the expected value of a given long-term burden sharing rule to each party, the analysis has provided some insight into what factors are of importance when the parties negotiate this rule ex ante. In the simplest, most symmetric version of the model, the bargaining solution for the burden sharing rule specifies contributions equivalent to the expected per-period negotiation outcome. This rule-design minimizes the probability of renegotiation, divides the expected value generated by the rule equally, and under risk neutrality will be achieved regardless of individual benefit variance.

In the second half of the paper, the model was extended in several directions to learn more about the consequences of asymmetries in the relationship. The analysis of negotiations between risk averse parties highlighted that while a pre-negotiated burden division can increase expected payoffs for both parties, it also increases the payoff variation as it reduces the scope for risk sharing. Moreover, if one party has greater variance in the individual public good benefit, the rule will have asymmetric impact on the parties' payoff risk. This can shift the bargaining solution away from the least-renegotiated rule, in favor of the party with greater individual risk. When there is asymmetry in negotiation costs, the parties do agree on the least-renegotiated burden sharing rule. However, under this rule, the party with the lower cost will contribute less than in the expected renegotiation outcome, which is independent of the bargaining cost. Asymmetry in bargaining power *does* change the expected negotiation outcome. The analysis revealed that the BSR will favor the stronger bargainer even further, possibly to such an extent that establishment of the rule becomes infeasible.

Importantly, it is the repetitive nature of the relation that makes it attractive to pre-negotiate a burden sharing rule. I have modeled this as an infinitely repeated game, where the future state of the world is unknown, but expectations about it are not affected by the current state. This is a strong assumption, but it has allowed the analysis to focus on the design aspects of the agreement that determine the distribution of surplus and the probability of ex-post renegotiation. While there are several possible avenues for future research based on further generalizations of the model, perhaps the most interesting option would be to relax this assumption of history independent public good benefits. Such a setting would have to incorporate the possibility of renegotiating the rule itself, and could add insights on the duration, or shelf life, of long-term agreements.

References

Bolton, Patrick and Mathias Dewatripont, *Contract Theory*, Vol. 1 of *MIT Press Books*, The MIT Press, January 2005.

Che, Yeon-Koo and Donald B. Hausch, "Cooperative Investments and the Value of Contracting," *American Economic Review*, March 1999, *89* (1), 125–147.

Chung, Tai-Yeong, "Incomplete Contracts, Specific Investments, and Risk Sharing," *The Review of Economic Studies*, 10 1991, *58* (5), 1031–1042.

Dufwenberg, Martin, Maros Servatka, and Radovan Vadovic, "Honesty and informal agreements," *Games and Economic Behavior*, 2017, *102*, 269 – 285.

Farrell, Joseph and Eric Maskin, "Renegotiation in repeated games," *Games and Economic Behavior*, 1989, *1* (4), 327 – 360.

Grossman, Sanford J. and Oliver D. Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, aug 1986, *94* (4), 691–719.

Hart, Oliver, "Hold-up, Asset Ownership, and Reference Points," *Quarterly Journal of Economics*, 2009, *124*, 267–300.

_ **and John Moore**, "Incomplete Contracts and Renegotiation," *Econometrica*, 1988, *56* (4), 755–785.

_ **and** _ , "Contracts as Reference Points," *Quarterly Journal of Economics*, 2008, *CXXIII*, 1–48.

Kalai, Ehud and Meir Smorodinsky, "Other Solutions to Nash's Bargaining Problem," *Econometrica*, 1975, *43* (3), 513–518.

Masten, Scott E., "Equity, Opportunism, and the Design of Contractual Relations," *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft*, 1988, *144* (1), pp. 180–195.

_ **and Keith J. Crocker**, "Efficient Adaptation in Long-Term Contracts: Take-or-Pay Provisions for Natural Gas," *The American Economic Review*, 1985, *75* (5), 1083–1093.

McCutcheon, Barbara, "Do Meetings in Smoke-Filled Rooms Facilitate Collusion?," *Journal of Political Economy*, 1997, *105* (2), pp. 330–350.

Miettinen, Topi, "Promises and conventions An approach to pre-play agreements," *Games and Economic Behavior*, 2013, *80*, 68 – 84.

Nash, John F., "The Bargaining Problem," *Econometrica*, 1950, *18* (2), 155–162.

_ , "Two-Person Cooperative Games," *Econometrica*, 1953, *21* (1), 128–140.

Perloff, Jeffrey M., "The Effects of Breaches of Forward Contracts due to Unanticipated Price Changes," *The Journal of Legal Studies*, 1981, *10* (2), pp. 221–235.

Polinsky, A Mitchell, "Fixed Price versus Spot Price Contracts: A Study in Risk Allocation," *Journal of Law, Economics, and Organization*, Spring 1987, *3* (1), 27–46.

Rachmilevitch, Shiran, "The Nash solution is more utilitarian than egalitarian," *Theory and Decision*, Nov 2015, *79* (3), 463–478.

Trockel, Walter, "A Walrasian approach to bargaining games," *Economics Letters*, 1996, *51* (3), 295 – 301.

_ , "In What Sense is the Nash Solution Fair?," in "Advances in Public Economics: Utility, Choice and Welfare," Springer-Verlag, 2006, pp. 17–30.

van Damme, Eric, "Renegotiation-proof equilibria in repeated prisoners' dilemma," *Journal of Economic Theory*, 1989, *47* (1), 206 – 217.

1.6 Appendix to Chapter 1

1.6.1 Proof for Lemma 1

Cooperative contribution levels maximize the Nash Product (NP):

$$\max_{\boldsymbol{q_t}} NP = \begin{cases} (h_t^A Q_t - q_t^A) \times (h_t^B Q_t - q_t^B) & \text{if } q_t^A + q_t^B, \leq 1 \\ (h_t^A - q_t^A) \times (h_t^B - q_t^B) & \text{if } q_t^A + q_t^B > 1, \end{cases}$$

where $Q_t = q_t^A + q_t^B$. It is trivial to see that the bargaining solution will not specify contributions in excess of the capacity constraint: $q_t^A + q_t^B > 1$ cannot maximize *NP*. Below that threshold, we can replace q_t^B by $(Q_t - q_t^A)$:

$$\max_{q_t^A, Q_t} (h_t^A Q_t - q_t^A) \times (h_t^B Q_t - Q_t + q_t^A)$$

Differentiating w.r.t. q_t^A gives a first-order condition for q_t^A:

$$h_t^A Q_t - q_t^A = h_t^B Q_t - Q_t + q_t^A$$

$$\iff q_t^A = \frac{(h_t^A - h_t^B)Q_t + Q_t}{2}$$

Gains from cooperation then require

$$h_t^A - \frac{(h_t^A - h_t^B)Q_t + Q_t}{2} = \frac{(h_t^A + h_t^B - 1)Q_t}{2} > 0$$

$$\iff h_t^A + h_t^B > 1.$$

To see that when this condition holds, $Q_t = 1$, express the maximization of the Nash product as

$$\max_{Q_t} \frac{(h_t^A + h_t^B - 1)Q_t}{2} \times \frac{(h_t^A + h_t^B - 1)Q_t}{2}$$

and note that

$$\frac{\partial NP}{\partial Q_t} = (h_t^A + h_t^B - 1)(h_t^A + h_t^B - 1)Q > 0.$$

1.6.2 Proof for $q_r^A + q_r^B = 1$

By Pareto optimality, the parties will not agree on a certain BSR, ex ante, if there is an alternative rule that weakly improves both their ex post payoffs, in all possible states of the world, and strictly improves them in at least one possible state of the world. In the following, I will show why this rules out any BSR that prescribes contributions such that $q_r^A + q_r^B \neq 1$.

Consider a BSR $\{q_r^A, q_r^B\}$. In any period t an agreement will only affect payoffs (relative to having no BSR) if neither party opts out. If this is the case in period t, it must hold for both i that $\pi_t^i(q_r) \geq \pi_t^i(q^*) = 0$. Suppose that under this BSR, the parties contribute below the capacity constraint: $q_r^A + q_r^B < 1$. We can construct an alternative rule by applying a small proportional increase to both parties' contributions:

$$\{q_r'^A, q_r'^B\} = \{(1+\epsilon)q_r^A, (1+\epsilon)q_r^B\},$$

which gives us the following ex post payoff for party i:

$$\pi_t^i(q_r') = h_t^i(1+\epsilon)(q_r^A + q_r^B) - (1+\epsilon)q_r^i = (1+\epsilon)\pi_t^i(q_r).$$

Whenever $\pi_t^i(q_r) \geq 0$, it must be true that $\pi_t^i(q_r') \geq \pi_t^i(q_r)$.[16] The alternative rule therefore improves payoffs in all periods where it is honored. This improvement is available for all rules where $q_r^A + q_r^B < 1$.

Suppose that, instead, the parties' BSR contributions exceed capacity: $q_r^A + q_r^B > 1$. In this case, both q_r^A and q_r^B can be reduced by a small amount, without affecting public good benefits ($h_t^i Q_r$), therefore improving both parties' payoffs.

For both cases, it may be true that the range of states of the world in which neither party opts out of the rule becomes larger under the proposed alternative, where the sum of contributions is closer to 1. For those specific states of the world, it is also true that the alternative rules improve ex post payoffs for both parties, since the payoffs they can obtain by opting out are independent of the BSR.

1.6.3 Proof of Proposition 1

To see that $q_r = E[q_t] = \mu_q$. satisfies first-order condition

$$\frac{\partial EV^A}{\partial q_r} EV^B + \frac{\partial EV^B}{\partial q_r} EV_A = 0,$$

note that

$$\frac{\partial EV^A}{\partial q_r} EV^B = \left(2\alpha f_q(q_r + \alpha) - \int_{q_r-\alpha}^{q_r+\alpha} \; dF_q(q_t)\right)\left(\int_{q_r-\alpha}^{q_r+\alpha} q_r - q_t + \alpha \; dF_q(q_t)\right)$$

and

$$\frac{\partial EV^B}{\partial q_r} EV^A = -\left(2\alpha f_q(q_r - \alpha) - \int_{q_r-\alpha}^{q_r+\alpha} \; dF_q(q_t)\right)\left(\int_{q_r-\alpha}^{q_r+\alpha} q_t - q_r + \alpha \; dF_q(q_t)\right).$$

At $q_r = E[q_t] = \mu_q$, we have

$$\frac{\partial EV^A}{\partial q_r} EV^B = \left(2\alpha f_q(\mu_q + \alpha) - \int_{\mu_q-\alpha}^{\mu_q+\alpha} \; dF_q(q_t)\right)\left(\int_{\mu_q-\alpha}^{\mu_q+\alpha} \mu_q - q_t + \alpha \; dF_q(q_t)\right)$$

[16] Note that this will be a strict inequality in at least some states, if a BSR is to have any value at all.

and

$$\frac{\partial EV^B}{\partial q_r}EV^A = -\left(2\alpha f_q(\mu_q-\alpha) - \int_{\mu_q-\alpha}^{\mu_q+\alpha} \ dF_q(q_t)\right)\left(\int_{\mu_q-\alpha}^{\mu_q+\alpha} q_t - \mu_q + \alpha \ dF_q(q_t)\right).$$

For a single-peaked distribution F_q that is symmetric around its mean μ_q, the first-order condition is satisfied.

1.6.4 Proof for one-sided risk

For the expected values we can rewrite the limits of the BSR-range as expressions of only h^A, to obtain

$$EV^A = \int_{\hat{h}^B+2q_r-1-2\alpha}^{\hat{h}^B+2q_r-1+2\alpha} u(h_t^A - q_r) - u(h_t^A - \frac{h_t^A - \hat{h}^B + 1}{2} - \alpha) \ dF_A(h^A),$$

$$EV^B = \int_{\hat{h}^B+2q_r-1-2\alpha}^{\hat{h}^B+2q_r-1+2\alpha} u(\hat{h}^B - 1 + q_r) - u(\hat{h}^B - 1 + \frac{h_t^A - \hat{h}^B + 1}{2} - \alpha) \ dF_A(h^A).$$

Therefore,

$$\begin{aligned}\frac{\partial EV_A}{\partial q_r} =& \ 2f_A(\hat{h}^B + 2q_r - 1 + 2\alpha)\left[u(\hat{h}^B + q_r - 1 + 2\alpha) - u(\hat{h}^B + q_r - 1)\right] \\ &- \int_{\hat{h}^B+2q_r-1-2\alpha}^{\hat{h}^B+2q_r-1+2\alpha} u'(h_t^A - q_r) \ dF_A(h^A)\end{aligned}$$

$$\begin{aligned}\frac{\partial EV_B}{\partial q_r} =& \ -2f_A(\hat{h}^B + 2q_r - 1 - 2\alpha)\left[u(\hat{h}^B + q_r - 1) - u(\hat{h}^B + q_r - 1 - 2\alpha)\right] \\ &+ \int_{\hat{h}^B+2q_r-1-2\alpha}^{\hat{h}^B+2q_r-1+2\alpha} u'(\hat{h}^B + q_r - 1) \ dF_A(h^A)\end{aligned}$$

At $q_r = \frac{\mu_A - \hat{h}_B + 1}{2} = E[q_t]$, the above four equations become

$$EV^A = \int_{\mu_A-2\alpha}^{\mu_A+2\alpha} u(h_t^A - \frac{\mu_A}{2} + \frac{\hat{h}_B - 1}{2}) - u(\frac{h_t^A + \hat{h}^B - 1}{2} - \alpha) \ dF_A(h^A),$$

$$EV^B = \int_{\mu_A-2\alpha}^{\mu_A+2\alpha} u(\frac{\mu_A}{2} + \frac{\hat{h}_B - 1}{2}) - u(\frac{h_t^A + \hat{h}^B - 1}{2} - \alpha) \ dF_A(h^A),$$

and,

$$\frac{\partial EV_A}{\partial q_r} = 2f_A(\mu_A + 2\alpha)\left[u(\frac{\mu_A}{2} + \frac{\hat{h}_B - 1}{2} + 2\alpha) - u(\frac{\mu_A}{2} + \frac{\hat{h}_B - 1}{2})\right]$$
$$- \int_{\mu_A - 2\alpha}^{\mu_A + 2\alpha} u'(h_t^A - \frac{\mu_A}{2} + \frac{\hat{h}_B - 1}{2})dF_A(h^A)$$

$$\frac{\partial EV_B}{\partial q_r} = -\,2f_A(\mu_A - 2\alpha)\left[u(\frac{\mu_A}{2} + \frac{\hat{h}_B - 1}{2}) - u(\frac{\mu_A}{2} + \frac{\hat{h}_B - 1}{2} - 2\alpha)\right]$$
$$+ \int_{\mu_A - 2\alpha}^{\mu_A + 2\alpha} u'(\frac{\mu_A}{2} + \frac{\hat{h}_B - 1}{2})dF_A(h^A).$$

Given the assumptions on $u()$,

$$EV_A < EV_B,$$

$$\frac{\partial EV_A}{\partial q_r} < 0, \qquad \frac{\partial EV_B}{\partial q_r} > 0,$$

$$\frac{\partial EV_A}{\partial q_r} + \frac{\partial EV_B}{\partial q_r} < 0,$$

and

$$\frac{\partial EV_A}{\partial q_r} EV_B + \frac{\partial EV_B}{\partial q_r} EV_A < 0. \tag{1.48}$$

This means that at $q_r = E[q_t]$, q_r is above the value that satisfies the condition for the Nash bargaining solution, which thus requires the BSR-contribution by A to be reduced.

1.6.5 Bargaining solution section 1.3.2

For fully correlated risk and an unrestricted range for the common shock, we have

$$EV^A = \int_{\max\{q_r - \hat{h}^A, 1 - q_r - \hat{h}^B\}}^{\frac{1}{2}(1+2\alpha-\hat{h}^A-\hat{h}^B)} \hat{h}^A + \theta_t - q_r \;\; dF_\theta(\theta_t) + \int_{\frac{1}{2}(1+2\alpha-\hat{h}^A-\hat{h}^B)}^{\infty} q_t - q_r + \alpha \;\; dF_\theta(\theta_t), \tag{1.49}$$

$$EV^B = \int_{\max\{q_r - \hat{h}^A, 1 - q_r - \hat{h}^B\}}^{\frac{1}{2}(1+2\alpha-\hat{h}^A-\hat{h}^B)} \hat{h}^B + \theta_t - 1 + q_r \;\; dF_\theta(\theta_t) + \int_{\frac{1}{2}(1+2\alpha-\hat{h}^A-\hat{h}^B)}^{\infty} q_r - q_t + \alpha \;\; dF_\theta(\theta_t). \tag{1.50}$$

Note that for any BSR that generates any value,

$$\max\{q_r - \hat{h}^A, 1 - q_r - \hat{h}^B\} \quad \leq \quad \frac{1}{2}(1 + 2\alpha - \hat{h}^A - \hat{h}^B).$$

This becomes clearer when we rewrite the above condition to become

$$q_r - \alpha \leq q_t \leq q_r + \alpha,$$

where q_t is as in (1.29). Now, if this does not hold, the BSR would *never* be honored:

1. For any $h_t^A + h_t^B \geq 1 + 2\alpha$, the BSR would be renegotiated $\rightarrow$ no value.
2. If $h_t^A + h_t^B < 1 + 2\alpha$:
   - The sum of the payoffs is at most 2α: $\quad \pi_t^A(q_t) + \pi_t^B(q_t) \leq 2\alpha$
   - The difference between the BSR-payoffs will be larger than that: $|\pi_t^A(q_t) - \pi_t^B(q_t)| \geq 2\alpha$
   - It follows that at least one of the two parties will have negative payoff under the BSR, and would opt out $\rightarrow$ no value.

To show that the the Nash product of the two expected BSR values is maximized at $q_r = q_t = \frac{\hat{h}^A - \hat{h}^B + 1}{2}$, note that at this q_r we have $EV^A = EV^B$. What remains to be shown is that at the same time $\frac{\partial EV^A}{\partial q_r} = -\frac{\partial EV^B}{\partial q_r}$.

For the derivatives we have to look at two ranges for q_r. First consider $q_r \geq q_t \iff q_r - \hat{h}^A \geq 1 - q_r - \hat{h}^B$. Then

$$\frac{\partial EV^A}{\partial q_r} = -\left[1 - F_\theta\left(q_r - \hat{h}^A\right)\right] \quad \text{vs.} \quad \frac{\partial EV^B}{\partial q_r} = -(2q_r - 2q_t) f_\theta(q_r - \hat{h}^A) + \left[1 - F_\theta\left(q_r - \hat{h}^A\right)\right].$$

In the other range, $q_r \geq q_t$, we have

$$\frac{\partial EV^A}{\partial q_r} = -(2q_r - 2q_t) f_\theta(1 - q_r - \hat{h}^B) - \left[1 - F_\theta\left(1 - q_r - \hat{h}^B\right)\right] \quad \text{vs.} \quad \frac{\partial EV^B}{\partial q_r} = \left[1 - F_\theta\left(1 - q_r - \hat{h}^B\right)\right].$$

From the above equation it is straightforward to see that the first-order condition for the Nash bargaining solution is satisfied at $q_r = q_t$, where $q_r - \hat{h}^A = 1 - q_r - \hat{h}^B$.

1.6.6 Proof of Proposition 2

Abbreviate the weighted Nash product as

$$NP(\gamma, q_r) = \quad (EV^A(q_r))^\gamma \times (EV^B(q_r))^{1-\gamma}. \tag{1.51}$$

Proposition 2 requires the following to be true:

$$\frac{\partial NP(\gamma, E[q_t])}{\partial q_r} < (>) 0 \quad \text{if} \quad \gamma > (<) \frac{1}{2},$$

where

$$\frac{\partial NP(\gamma, q_r)}{\partial q_r} = \gamma \left(\frac{EV^B}{EV^A}\right)^{1-\gamma} \frac{\partial EV^A}{\partial q_r} + (1-\gamma)\left(\frac{EV^A}{EV^B}\right)^\gamma \frac{\partial EV^B}{\partial q_r}. \tag{1.52}$$

Similar to the proof for Proposition 1, it is useful to define a distribution F_q, that is now constructed from the two benefit factors according to

$$q_t = h_t^A - \gamma(h_t^A + h_t^B - 1). \tag{1.53}$$

Relative to this q_t, the BSR-values are determined by q_r in the same way they are for symmetric bargaining strength:

$$EV^A = \int_{q_r-\alpha}^{q_r+\alpha} q_t - q_r + \alpha \;\; dF_q(q_t), \tag{1.54}$$

$$EV^B = \int_{q_r-\alpha}^{q_r+\alpha} q_r - q_t + \alpha \;\; dF_q(q_t).$$

For $q_r = E[q_t] = \mu_q$, and distributions of the random variables that are symmetric around their means, these values are equalized: $EV^A = EV^B$. This reduces (1.52) to

$$\frac{\partial NP}{\partial q_r} = \gamma\frac{\partial EV^A}{\partial q_r} + (1-\gamma)\frac{\partial EV^B}{\partial q_r} \tag{1.55}$$

For the partial derivatives we (again) obtain

$$\frac{\partial EV^A}{\partial q_r} = 2\alpha f_q(q_r + \alpha) - \int_{q_r-\alpha}^{q_r+\alpha} \;\; dF_q(q_t)$$

and

$$\frac{\partial EV^B}{\partial q_r} = -2\alpha f_q(q_r - \alpha) + \int_{q_r-\alpha}^{q_r+\alpha} \;\; dF_q(q_t).$$

For a symmetric, single peaked distribution F_q, densities for q-values inside the range $[\mu_q - \alpha, \mu_q + \alpha]$ are greater than outside. Hence, at $q_r = \mu_q$ we have $\frac{\partial EV^A}{\partial q_r} < 0$ and $\frac{\partial EV^B}{\partial q_r} > 0$, while the sum of the two is equal to zero. This is therefore only the bargaining solution if $\gamma = \frac{1}{2}$.

If not, it follows from (1.55) that the sign of the derivative, at $q_r = \mu_q$, matches that of $(1 - 2\gamma)$.

1.6.7 Example asymmetric bargaining strength

To illustrate how asymmetric bargaining strength might prevent a potentially beneficial BSR, consider the following simple example:

- Benefit factors are deterministic, assume: $h^A = h^B = \frac{3}{4}$.
- A is assumed to be the stronger bargainer: $\gamma > \frac{1}{2}$.
- Benefits are large enough for B to enter negotiations in $t \geq 1$: assume $\alpha < \frac{1-\gamma}{2}$
- $T = 2 \quad \rightarrow \quad$ There are two periods of possible public good provision, after $t = 0$ where the parties may or may not establish a BSR

- Assume that the parties do not discount future payoffs, and that there is an extra cost α_r to negotiating the BSR that is equal to the cost for per-period negotiations: $\alpha_r = \alpha$. This extra cost is the threshold anticipated value, below which a party will refuse to commence negotiations on a rule.

By (1.34), we obtain the (re)negotiation outcome in periods 1 and 2:

$$q_t = \frac{3-2\gamma}{4}.$$

Without a BSR, the surplus subject to negotiations in both periods 1 and 2 is certainly $\frac{1}{2}$. Taking into account that these negotiations are costly, we know that party B will end up with a payoff of $\pi_t^B = \frac{(1-\gamma)}{2} - \alpha$. The BSR-values in each period $t \geq 1$ are given by

$$V^A = \frac{3-2\gamma}{4} - q_r + \alpha \quad \text{and} \quad V^B = q_r - \frac{3-2\gamma}{4} + \alpha.$$

Note that, as before, if the BSR is set at $q_r = E[q_t] = q_t$, the BSR-value in period t will be equal: $V^A = V^B = \alpha$. The bargaining solution for q_r will be different, however, for unequal bargaining strength:

$$\max_{q_r} \; (2V^A)^\gamma \times (2V^B)^{1-\gamma}, \tag{1.56}$$

which gives first-order condition

$$\gamma \left(\frac{V^B}{V^A}\right)^{1-\gamma} = (1-\gamma)\left(\frac{V^A}{V^B}\right)^{\gamma}. \tag{1.57}$$

Using that $V^A + V^B = 2\alpha$, we obtain

$$V^A = 2\gamma\alpha \quad \text{and} \quad V^B = (1-\gamma)2\alpha. \tag{1.58}$$

Remember that $\gamma > \frac{1}{2}$. Given the cost $\alpha_r = \alpha$ associated with the BSR-negotiations, B will only be willing to enter those negotiations at $t = 0$ if

$$2V^B \geq \alpha \qquad \Longleftrightarrow \qquad (1-\gamma)4\alpha \geq \alpha.$$

It is straightforward to see, then, that there will not be a BSR for $\gamma > \frac{3}{4}$. It follows that party A would then be better off if able to commit to a BSR outcome more favorable to B (obviously B would also be better off). To illustrate this, the following table compares A's payoffs without BSR to those with BSR $q_r = q_t$.

Period t	**No BSR:** π^A	**BSR $q_r = q_t$:** π^A
0	0	-α
1	$\gamma/2 - \alpha$	$\gamma/2$
2	$\gamma/2 - \alpha$	$\gamma/2$
Total	$\gamma - \mathbf{2}\alpha$	$\gamma - \alpha$

1.6.8 Proof of proposition 3

To see that $q_r = \mu_q + \frac{\alpha_A - \alpha_B}{2}$ satisfies

$$\frac{\partial EV^A}{\partial q_r} EV^B = -\frac{\partial EV^B}{\partial q_r} EV^A, \tag{1.59}$$

first note that for this BSR, under the same assumptions on F_q as before,

$$EV^A = EV^B = \frac{\alpha_A + \alpha_B}{2} \int_{q_r - \alpha_A}^{q_r + \alpha_B} dF_q(q_t).$$

Second, for the partial derivatives we obtain

$$\frac{\partial EV^A}{\partial q_r} = (\alpha_A + \alpha_B) f_q(q_r + \alpha_B) - \int_{q_r - \alpha}^{q_r + \alpha} dF_q(q_t)$$

and

$$\frac{\partial EV^B}{\partial q_r} = -(\alpha_A + \alpha_B) f_q(q_r - \alpha_A) + \int_{q_r - \alpha}^{q_r + \alpha} dF_q(q_t).$$

Condition (1.59) therefore requires

$$f_q(q_r + \alpha_B) = f_q(q_r - \alpha_A), \tag{1.60}$$

Plugging in the proposed q_r makes it clear that (1.60) holds:

$$f_q(\mu_q + \frac{\alpha_A - \alpha_B}{2} + \alpha_B) = f_q(\mu_q + \frac{\alpha_A - \alpha_B}{2} - \alpha_A).$$

1.6.9 Derivations and proofs for section 1.4

1.6.9.1 The rule when never renegotiated

For the case of asymmetric negotiation costs:

$$EV^A = \int_{q_{\min}}^{q_{\max}} q_t - q_r + \alpha_A \ dF_q(q_t) = \mu_q - q_r + \alpha_A$$

$$EV^B = \int_{q_{\min}}^{q_{\max}} q_r - q_t + \alpha_B \ dF_q(q_t) = q_r - \mu_q + \alpha_B,$$

so

$$\frac{\partial EV^A}{\partial q_r} EV^B = -\frac{\partial EV^B}{\partial q_r} EV^A \quad \Longrightarrow \quad -EV^B = -EV^A.$$

This gives $q_r = \mu_q + \frac{\alpha_A - \alpha_B}{2}$.

For asymmetric bargaining, the first-order condition for the bargaining solution is

$$\frac{\partial NP(\gamma, q_r)}{\partial q_r} = \gamma \left(\frac{EV^B}{EV^A}\right)^{1-\gamma} \frac{\partial EV^A}{\partial q_r} + (1-\gamma) \left(\frac{EV^A}{EV^B}\right)^{\gamma} \frac{\partial EV^B}{\partial q_r} = 0, \tag{1.61}$$

where

$$EV^A = \mu_q - q_r + \alpha$$

$$EV^B = q_r - \mu_q + \alpha.$$

This means the first-order condition holds for

$$\frac{\gamma}{1-\gamma}(q_r - \mu_q + \alpha) = \mu_q - q_r + \alpha,$$

which is satisfied at $q_r = \mu_q + (1 - 2\gamma)\alpha$.

1.6.9.2 WRP strategies - Asymmetric punishment

To show that the following condition cannot be satisfied for the current setting, consider the LHS, and its largest possible value:

$$\delta^T \int_{q_r^H}^{q_{max}} q_t - q_r^L \;\; dF_q(q_t) \quad \geq \quad q_r^L - q_{\min}. \tag{1.62}$$

First, assume $\delta = 1$. What remains is a value for $q_t - q_r^L$, waited for the probability with which q_t takes on specific values in the range $q_t \in [q_r^H, q_{\max}]$, with the total probability that it falls in this range strictly smaller than 1: $1 - F_q(q_r^H)$. Given that the probability distribution of q_t is symmetric and unimodal, both the probability within this subrange on one end of the complete range $[q_{\min}, q_{\max}]$, and the expected q_t given that it falls within this subrange, cannot be greater than it is for a uniformly distributed q_t.

Also, let us take $q_r = \mu_q$. A deviation from this solution for the rule would create asymmetry in potential deviation gains and in the continuation payoffs under cooperation, which would only make it harder to satisfy all three requirements (i) - (iii), for all q_t, for both parties, as one of them would have a stronger incentive to deviate than the other. In other words, if the condition can not be met for $q_r = \mu_q$, it will also be impossible for different rules.

Given the uniform distribution of q_t, we can rewrite the LHS, using that the probability density is the same across all values for which the distribution has support:

$$\int_{q_r^H}^{q_{max}} q_t - q_r^L \;\; dF_q(q_t) = [1 - F_q(q_r^H)] \cdot (E(q_t | q_t \in [q_r^H, q_{\max}]) - q_r^L). \tag{1.63}$$

We can plug in more specific terms for the expected value and the probability, to rewrite the condition:

$$\frac{q_{\max} - q_r^H}{q_{\max} - q_{\min}} \cdot \left(\frac{q_{\max} + q_r^H}{2} - q_r^L\right) \quad \geq \quad q_r^L - q_{\min},$$

which can be rearranged to become

$$\frac{q_{\max} + q_r^H}{2} - q_r^L \quad \geq \quad (q_r^L - q_{\min}) \cdot \frac{q_{\max} - q_{\min}}{q_{\max} - q_r^H}.$$

We know that for a $q_r = \mu_q$, the symmetry of q_t means that $q_r^L - q_{\min} = q_{\max} - q_r^H$, which reduces the condition to

$$\frac{q_{\max} + q_r^H}{2} - q_r^L \quad \geq \quad q_{\max} - q_{\min}. \tag{1.64}$$

This is a false inequality, since $q_{\max} > q_r^H$ and $q_{\min} < q_r^L$, whenever α is small enough to have renegotiation with positive probability in the repeated one-shot Nash equilibrium. In other words, if both sides in (1.64) are equal, the deviation gain is never positive, so a rejection of q_r would never occur anyway.

1.6.9.3 Smaller expansions

In the following, I will show that the asymmetric punishment schedule also fails to meet the third requirement for smaller expansions of the self enforcing range. Instead of honoring the agreement q_r for *all* possible q_t, imagine that the parties promise to stick to the rule (in the 'normal' phase) in all periods where $q_t \in [q_{\min} + c, q_{\max} - c]$. Note that the 'full' range discussed above, is a special case of this general one, where $c = 0$. The rest of the scheme, again considering a potential deviation by party A:

- After a deviation by party A, party B switches to the one-shot Nash equilibrium strategy for T periods: reject $\boldsymbol{q_r}$ when $q_t > q_r^H = q_r + \alpha$.

- During this punishment phase, party A executes to the cooperative strategy: stick to the BSR when $q_t \in [q_{\min} + c, q_{\max}]$, opt out otherwise.

- If party A deviates from the prescribed strategy during the punishment phase, it starts over.

Requirement (i) now becomes

$$\sum_{t=1}^{T} \delta^t \int_{q_r^H}^{q_{\max}-c} q_t - q_r + \alpha \;\; dF_q(q_t) \quad \geq \quad q_r - (q_{\min} + c) - \alpha, \tag{1.65}$$

and can still be fulfilled for large enough δ. Party B still gains from punishing, fulfilling condition (ii):

$$\sum_{t=t'+1}^{T} \delta^{t-t'} \int_{q_r^H}^{q_{\max}-c} q_t - q_r - \alpha \;\; dF_q(q_t) \geq 0. \tag{1.66}$$

This brings us to requirement (iii). The most likely period to deviate, fixing q_t, is still the first period after the initial deviation. For party A to be willing to stay in punishment phase, for all q_t, it must hold that

$$\delta^T \int_{q_r^H}^{q_{\max}-c} q_t - q_r + \alpha \;\; dF_q(q_t) \geq \;\; q_r - \alpha - (q_{\min} + c), \tag{1.67}$$

The same steps can be applied in this more general case. Assume that $\delta = 1$, $q_r = \mu_q$ and that q_t is uniformly distributed over the range $[q_{\min}, q_{\max}]$. A transformation equivalent to that described by equation (1.63) allows us to rewrite the above condition as

$$\frac{q_{\max} - c - q_r^H}{q_{\max} - q_{\min}} \cdot \left(\frac{q_{\max} - c + q_r^H}{2} - q_r^L\right) \;\; \geq \;\; q_r^L - (q_{\min} + c)$$

which can be rearranged to become

$$\frac{q_{\max} - c + q_r^H}{2} - q_r^L \;\; \geq \;\; (q_r^L - (q_{\min} + c)) \cdot \frac{q_{\max} - q_{\min}}{q_{\max} - c - q_r^H}.$$

For $q_r = \mu_q$, $q_r^L - (q_{\min} + c) = q_{\max} - c - q_r^H$, which reduces the condition to

$$\frac{q_{\max} - c + q_r^H}{2} - q_r^L \;\; \geq \;\; q_{\max} - q_{\min}. \tag{1.68}$$

This is a false inequality, since $q_{\max} > q_r^H - c$ and $q_{\min} < q_r^L$, for the relevant α range. We can therefore conclude that the scheme also does not work for smaller expansions of the self-enforcing range. The largest possible gain a party can experience by deviating is smaller, but so is the penalty that is supposed to deter it.

The same is true for what one might call a hybrid scheme, where the expansion in the normal phase can be smaller, but the punishment stays more severe. The problem, again, lies in requirement (iii). The more severe the punishment, the more tempting it is to deviate from the punishment phase. This hybrid scheme is illustrated in the following.

The normal phase is as that in the scheme above: the parties stick to the BSR in all periods where $q_t \in [q_{\min} + c, q_{\max} - c]$. The rest of the scheme:

- After a deviation by party A, party B switches to the one-shot Nash equilibrium strategy for T periods: reject $\boldsymbol{q_r}$ when $q_t > q_r^H = q_r + \alpha$.

- For the duration of this punishment phase, party A switches to the *fully* cooperative strategy: always accept $\boldsymbol{q_r}$, never opt out.

- If party A deviates from the prescribed strategy during the punishment phase, it starts over.

Along similar lines as before, it is possible to meet requirements (i) and (ii). The highest posible deviation gain in the normal phase is identical to the right-hand side of equation (1.65), the LHS is now greater as the punishment is more severe, which also makes it easier to meet requirement (ii). Requirement (iii), again, proves to be the prohibitive obstacle:

$$\delta^T \left[\int_{q_r^H}^{q_{\max}-c} q_t - q_r + \alpha \;\; dF_q(q_t) + \int_{q_{\min}}^{q_{\min}+c} q_r - q_t - \alpha \;\; dF_q(q_t) \right] \geq \;\; q_r - \alpha - q_{\min}, \quad (1.69)$$

Again, assume that $\delta = 1$, $q_r = \mu_q$ and that q_t is uniformly distributed over the range $[q_{\min}, q_{\max}]$. Then the LHS becomes

$$\frac{q_{\max} - c - q_r^H}{q_{\max} - q_{\min}} \cdot \left(\frac{q_{\max} - c + q_r^H}{2} - q_r^L\right) + \frac{c}{q_{\max} - q_{\min}}\left(q_r^L - q_{\min} - \frac{c}{2}\right),$$

while the RHS is the same as in the full expansion case: $q_r^L - q_{\min}$. Differentiating the LHS with respect to c gives

$$\frac{c + q_r^L - q_{\max}}{q_{\max} - q_{\min}},$$

which is negative for the relevant values of c. This means that an increase in c will decrease the LHS of (1.69), and we know that LHS $<$ RHS for $c = 0$ (note that RHS is independent of c), ruling out the fulfillment of requirement (iii).

Chapter 2

Asymmetric Risk and Non-Binding Rules: An Experiment

Author: Arne Pieters

Abstract: In this paper, I use an experiment to study a non-binding rule on how to divide the costs for a joint project. In a repeated setting with fluctuating individual project revenues, pairs of subjects decide whether to stick to the rule or renegotiate contributions ex post. The fixed division under the rule does not adjust to changing bargaining positions, but renegotiations are costly to both partners. I test theoretical predictions on the effects of asymmetry in individual revenue risk. My results indicate that subjects who face more individual risk (ex ante) are more likely to opportunistically force renegotiations (ex post), improving their own payoff but reducing the joint earnings. I find no evidence for a direct effect of risk asymmetry, *per se*, on the degree of cooperation between partners.

JEL: C90, D81

Keywords: Non-Binding Rules, Risk Asymmetry, Cooperation, Costly Negotiation

2.1 Introduction

The costs for public goods or services, that generate benefits to more than one person or entity are often shared between several parties. Many of these public goods, services, institutions or projects require investment repeatedly (e.g., every month or every year). Over time, circumstances change, and this might change what we see as a fair division of costs. While we may stick to an existing division, we could also decide to renegotiate who pays what share of the costs. This will depend on the extent to which circumstances have changed, but also on the type of agreement we have. This paper focuses on cost-sharing agreements that are not binding, which are more likely to be renegotiated than legally binding contracts. Although there is little research on such agreements in comparison to binding contracts, various types of economic relations make frequent use of them. One can think of cost-sharing between two or more countries in funding international institutions, using non-binding agreements because they do not have a supranational court to enforce contracts. In other, more informal relations, such as those between friends or housemates, it is not impossible to write a contract but it is simply not the preferred option.

Renegotiation requires time and effort, and in some cases generates uncertainty about the public good's availability. This can be a reason for parties to avoid renegotiation, even when they would expect an improved outcome, and instead stick to a cost-sharing rule that was agreed upon earlier. While not free of cost, renegotiating the burden division also serves as a way to risk-share, by allowing adjustments to changing circumstances. If we stick to a predetermined cost division instead, risk is borne individually, which can affect the appeal of such a rule if agents have risk averse preferences.

To illustrate this, imagine you and your housemate sign up for a joint subscription to Netflix. You consume a monthly average of 50 hours of Netflix, and so does your housemate. You consumption variance is large: Your get up to 100 hours in some months, while in other months you do not consume at all. Your housemate, on the other hand, has zero consumption variance, and consistently consumes 50 hours of Netflix every month. Together, you have two options for splitting the costs: (i) each of you covers half the subscription fee, regardless of how much you use it in a certain month, or (ii) you sit down at the end of each month and jointly determine a fair cost division based on your respective consumption.

The first option may have your preference because you do not have to negotiate the cost division every single month. On the other hand, you run the risk of having to pay half the cost for a service you do not use at all. This is different for your housemate. The first option avoids negotiation, but also means she remains unaffected by the variance in your consumption: each month, she knows exactly how much she will pay for her 50 hours of Netflix.

A reduced scope for risk-sharing increases the total risk borne by a group of agents. The effect it has on individual agents depends on how much their individual circumstances vary, as illustrated by the Netflix example.[1] The asymmetric risk implications of a cost-sharing rule are mostly interesting because they could imply asymmetric bargaining positions when negotiating what the rule should be. The added risk of having to pay for a Netflix subscription you do not use is unlikely to outweigh the benefits of not having to negotiate with your housemate every month. However, you may be able to convince your housemate that this asymmetry is unfair, and negotiate a cost division where you pay less than half the costs.[2]

In this paper, I present the results of an experiment in which participants face these type of burden sharing problems. I study several aspects of a non-binding agreement on how to share costs, which I will refer to as a burden sharing rule (BSR). To investigate the elements of risk and asymmetry discussed in the previous paragraphs, I use a setting where circumstances change over time, and where for some agents they vary more than for others. Part I of the experiment studies the ex post decisions individuals make, and how they are affected by asymmetric risk. In this part, the BSR is imposed exogenously.[3] Subjects repeatedly interact with each other to jointly fund a profitable project, each time deciding between sticking to the BSR and renegotiating the cost division. In Part II, I study the bargaining positions ex ante, by testing the theoretical prediction that risk asymmetry generates asymmetry in the expected value of a given rule, if agents are risk averse. In this part, the establishment of the rule depends on subjects' willingess to pay for it.

The changing circumstances are implemented as random variations in individual project revenues, which determine the bargaining positions for potential renegotiations. A monetary cost is incurred when subjects renegotiate, and this cost determines the theoretical self-enforcing range of the BSR.[4] I find that subjects largely behave as predicted within this range, deciding to stick to the rule 87% of the time. But even when in such a good bargaining position that renegotiating promises a higher payoff, subjects choose to stick to the BSR 27% of the time. This can be seen as a cooperative action, foregoing an individual improvement to maximize total joint payoff.

The main finding of this paper is that the tendency to cooperate, which effectively extends the self-enforcing range of a BSR, is negatively affected by the level of individual risk. When in the position to improve their own payoff by renegotiating, subjects with fixed project revenues are less likely to opt out of the rule than their partners, whose revenues vary between periods. This treatment effect is significant in the first interaction between two partners, and across all interactions when excluding subjects with risk-seeking

[1] Think of risk here as the variance in what an individual pays per hour of consumption.

[2] These and other aspects of non-binding agreements are theoretically analyzed in a separate paper (Pieters (2019)).

[3] The imposed rule is ex ante 'fair', in the sense that it generates the same expected profits for both partners.

[4] This is derived as the range of relative bargaining positions for which both partners stick to the rule in the subgame-perfect Nash equilibrium.

preferences. I attribute it in part to efficiency concerns becoming more important at higher overall payoff levels, and in part to (asymmetric) consequences of extending the self-enforcing range on the payoff risk. Both these mechanisms are associated with utility that is concave in monetary payoff, corresponding to risk aversion.

In a control treatment, pairs were symmetric in risk, with fluctuating revenues for both subjects. I compare these subjects to those on the risky side of the asymmetric risk treatment. The only difference between these two groups is in the revenue variance of their parters. This allows me to study how heterogeneity in risk affects cooperation. To the best of my knowledge, experimental research that examines this specific relation does not exist. Several studies, among which those by Ahn et al. (2007) and Beckenkamp et al. (2007), have found payoff assymetry in the prisoner's dilemma to reduce cooperation.[5] In the current experiment, expected payoffs are always symmetric, but the risk allocation is not. I find no evidence for a negative effect of risk asymmetry on cooperation. In fact, the cooperative expansion of the self-enforcing range is more prevalent among asymmetric pairs. However, this difference is driven by the more cooperative behavior of the risk-free individuals. It can therefore be attributed to the lower joint risk levels of those pairs, and not to the asymmetry itself.

In related work, several experimental studies (Irlenbusch, 2004; Ben-Ner and Putterman, 2009; Kessler and Leider, 2012) find that subjects often choose to enter a non-binding agreement when they have the option, and are likely to honor it, even when they could profitably deviate. A key common element in these experiments is the ex ante formation of agreement, and the interaction between subjects that takes place at that point, which is not part of the current study. The literature mostly considers such agreements as mutual promises, that are pshychologically costly to renege on, as people prefer to be honest and trustworthy (Dufwenberg et al., 2017). It is unlikely that an exogenously imposed rule, or 'agreement', is viewed by the subject as a promise they intend to keep.

Also related is the literature on behavioral contract theory, that looks at how incomplete contracts shape outcomes beyond what they legally bind the parties to. Of particular relevance is the work by Hart and Moore (2008), who formulate what they call the Reference Point Hypothesis, based on the idea that contracts create a sense of entitlement on elements that the parties are not necessarily bound to. Several articles provide experimental evidence supporting this hypothesis, including the idea that subjects can engage in hostile behavior ex post if their expectations are not met (Fehr et al., 2011, 2015; Bartling and Schmidt, 2015)[6]. In the current experiment, each partner can unilaterally opt out of the BSR, but

[5] Cherry et al. (2005) report similar results for public good games with heterogeneous endowments, whileTan (2008) finds that heterogeneity in the marginal per capita return of the public good reduces the mean contribution level.

[6] Fehr et al. (2011) and Fehr et al. (2015) investigate this in a buyer-seller setting where sellers can shade (at no cost or gain) to punish buyers who select a low price. Bartling and Schmidt (2015) show that contracts create reference points that can affect ex post renegotiations.

renegotiations take place only if both partners agree. At that point, refusing renegotiations is never payoff-maximizing, but could serve as a costly way to punish the partner for opting out of the rule. I do not observe such actions at a high rate: when participants opt out of the BSR, their partners refuse to renegotiate about 15% of the time. The data suggests that subjects who do refuse renegotiations, do so impulsively rather than strategically, which would support the idea that deviating from the rule can generate inefficient hostility.

The second part of the experiment elicits the ex ante value subjects attach to a given BSR. These values are indicators of what their bargaining positions would be, when negotiating the specific cost division of the rule. Theory would predict that, if agents are risk averse and expected utility maximizers, their valuations should be impacted by asymmetry in revenue risk. I find no evidence to support this. Among asymmetric pairs, there is no indication that individual risk has any effect on how subjects value the rule. A possible explanation for this is that subjects lack a certain awareness or sophistication in the risk domain, that they do display in other domains (Potters et al. (2014)[7] find evidence in support of this). Another explanation could be that subjects have prosocioal preferences that extend to the risk domain. Several experimental findings, among which Bolton and Ockenfels (2010) and Brock et al. (2013), indicate that individuals do take risk of others into account. The empirical analysis of Part II provides some support for this explanation.

The rest of the paper is structured as follows. Section 2.2 presents the experimental design. Section 2.3 constructs theoretical predictions on subject decisions and discusses possible deviations. The results are presented in Section 2.4. Section 2.5 concludes.

2.2 Experimental Design and Procedures

In this section I will first describe the design of the stage game, explaining the choices subjects face and how these determine their payoffs. I will then discuss the administered treatments, the overall structure of the experiment and the practical experimental procedures.

2.2.1 Payoff Structure

Pairs of subjects can earn money by funding a project that generates revenue to both of them. Each subject is endowed with ECU 90 and the total project costs are always ECU 100, so neither partner has the means to finance the project on her own. A successfully financed project generates revenue r_i to subject i. The two partners simultaneously make decisions that determine how they pay the costs.

[7] In a study on adverse selection in risk sharing arrangements, Potters et al. (2014) find that while subjects in a fairly sophisticated manner identify differences in expected outcomes, and consequently behave strategically, this is not the case when only the risk is different between individuals.

In decision $D_{1i} \in \{0,1\}$, subject i indicates whether she is willing to stick to the burden sharing rule and contribute c_r. The BSR is always an equal split of the total costs: $c_r = 50$. As it is not binding, mutual consent ($\min\{D_{11}, D_{12}\} = 1$) is required to implement BSR-contributions. Without such consent, a second decision $D_{2i} \in \{0,1\}$ determines whether or not renegotiations take place, incurring cost θ on both partners and resulting in contribution c_i for each individual i. For the cost of negotiation, different values are implemented ($\theta \in \{5, 10, 20\}$), but it is always the same for both partners. If either partner refuses renegotiation, the project does not materialize and no profits are made. This structure can be summarized in the following payoff function:

$$\begin{aligned} \pi_i = 90 \quad &+ \quad (r_i - c_r) \cdot \min\{D_{11}, D_{12}\} \\ &+ \quad (r_i - c_i - \theta) \cdot \min\{D_{21}, D_{22}\} \cdot (1 - \min\{D_{11}, D_{12}\}). \end{aligned} \tag{2.1}$$

The division of costs that results from renegotiations depends on the individual revenues, both of which are known to both partners before the first decision is made. Rather than subjects actually bargaining over $\{c_1, c_2\}$, renegotiations are automated and always result in a cost division that splits the profit evenly between the partners.[8] This means the renegotiated contributions adjust for differences in revenue to equate the individual profits:

$$c_i = 50 + \frac{r_i - r_j}{2} \tag{2.2}$$

Each participant is assigned one of two types $\rho \in \{A, B\}$. The type determines their revenue profile, i.e., the set of possible project revenues. The average revenue is the same for both types, but for types A it is constant while for types B it varies:

- Revenue from the project for type A is fixed: $r_i = 90$
- Revenue from the project for type B varies: $r_i \sim U\{50, 70, 110, 130\}$

In every session, one of two treatments is applied with regard to the pairing of types. In AB sessions, the pairs always consist of one type A and one type B. In BB sessions, all participants are assigned type B so all pairs consist of two types B.[9]

[8] With a zero profit threat point for both partners and revenues unaffected by the division of costs (implying a zero-sum situation), this corresponds to any solution that to some extent attempts to grant equal gains to both parties (including the well known solutions proposed by Nash (1950); **?** and by Kalai and Smorodinsky (1975)).

[9] There are no sessions with only types A, as a complete lack of variability in revenues would not produce any interesting dilemmas for the participants - sticking to the BSR would always be payoff-maximizing, efficient, and equitable.

Table 2.1: Overview of sessions

Session	1	2	3	4	5	6	Total
Treatment	AB	AB	AB	BB	BB	AB	
N (A)	14	14	13	0	0	14	55
N (B)	14	14	13	28	30	14	113

2.2.2 Experiment Structure and Procedures

At the start of each experimental session, participants learn their type and are informed what type their partners in the session will be. In Part I subjects then play three rounds of five periods. Before each round they are randomly matched to a new partner, reminded of the revenue profiles and informed about that round's renegotiation cost θ. In each period the partners do the project funding task explained in subsection 2.2.1. The revenues r_i for the types B are drawn at the start of each period, and made known to both partners. The actions and outcome in one period in no way affect their options in the next period.

In Part II, establishment of the burden sharing rule is no longer automatically provided, but can be established if the subjects are willing to pay for it. This part consists of two one-shot rounds, which means the participants interact with a partner for one period only. Each round starts with a first stage that determines whether or not the burden sharing rule is established, i.e., whether the partners will have the $c_r = 50$ option. If so, the second stage is identical to the periods played in Part I. If, however, no rule is established, the pair can only fund the project by negotiating contributions (at cost θ). In this case, decision D_1 is skipped and the participants move directly to decision D_2. The payoff structure is then identical to that in any period in Part I after $\min\{D_{11}, D_{12}\} = 0$.

In the first stage of each round in Part II, the subjects know their own type and that of their partner, as well as the negotiation cost θ. In the first round $\theta = 10$ and in the second round $\theta = 20$. Based on this information, they are asked to pick the maximum amount they would be willing to pay (WTP) to establish the BSR. This is incentivized using a random price mechanism (Becker et al., 1964). One of the two partners is randomly selected, and a price p is randomly generated. If WTP $\geq p$ for the selected partner, the BSR is successfully established at price p. At the start of the second stage, the subjects observe whether or not the BSR has been established, but are not told which partner was selected or what price p was.[10] At this point the revenues r_i for the types B are drawn, and the participants continue with the project funding decisions.

Table 2.2 summarizes the structure of Parts I and II. The first part studies (ex post) how often the rule is used by pairs of subjects that are in repeated interaction with each other. Part II is aimed at measuring how much participants value the rule (ex ante) as

[10] This information was provided to the participants at the very end of the session.

Table 2.2: Session Structure

Part	Round	Periods (t)	θ	Decisions each period
I	1	$1-5$	5	D_1, D_2^*
	2	$6-10$	10	D_1, D_2^*
	3	$11-15$	20	D_1, D_2^*
II	4	16	10	WTP, D_1^*, D_2^*
	5	17	20	WTP, D_1^*, D_2^*

Note: Subjects are randomly assigned a new partner for every round. D^*: decision node not always reached.

an available option. I should note here that the order in which θ changes across rounds is the same in all sessions. The main reason to have this order the same for all subjects, is for them to have the same structured experience of how different levels of θ determine the effect and the value of the BSR, which is likely to inform their decisions in Part II.[11]

To be able to control for heterogeneity in risk aversion and social preferences, subjects perform some short incentivized decision tasks at the end of the experiment (Part III). These are a standard dictator game and the task developed by Holt and Laury (2002) to elicit risk attitudes. Appendix 2.6.1 addresses these tasks and the observed decisions in more detail. For the analysis in section 2.4, it is relevant to know that subjects were categorized as *Risk Averse, -Seeking, -Neutral* or *Inconsistent*, based on their choices in the risk task.

The experimental sessions were run in the WISO-lab of the University of Hamburg in December 2016 and January and February 2017, using z-Tree software(Fischbacher, 2007). A total of 162 subjects participated in six sessions that lasted about 90 minutes each. Participants could earn experimental currency (ECU 24 = EUR 1) in all three parts discussed above, with the final payouts ranging from EUR 9.1 to EUR 19.7.

The participants received written and spoken instructions before each part and answered several control questions to make sure the payoffs and decision structure of the game were understood. When making decisions D_1 and D_2, each subject was provided with on-screen information on the payoff-relevant variables, such as negotiation cost θ, the project revenues $\{r_i, r_j\}$ for that period, and the c_i that would result from renegotiation given those revenues.[12]

The final screen for every period summarized the decisions made and the resulting individual profit for that period. At the end of the session, one period from Part I and one

[11]For Part I, this does mean that higher costs coincide with the participants having more experience with the task. Identifying experience as a significant factor, by randomizing this order, would have required a larger number of observations (per order). Even so, I acknowledge the design cannot rule out that any behavior induced by higher negotiation costs is not partially driven by experience with the task. That said, the main point of interest in Part I is the difference in behavior between types A and B and between treatments AB and BB, and these should not be affected differently by this design feature.

[12]The instructions and examples of decision screens can be found in the Appendix.

from Part II were randomly selected for payout, as well as one of the tasks in Part III. For Part II, if the BSR had been succesfully established in the selected period, an amount p was deducted from the payoff of the selected partner. At the very end of the experiment, the participants were presented with an overview of which periods and tasks were randomly chosen to determine their payout, and subsequently how much they earned.

2.3 Predictions

This section discusses theoretical predictions for actions and decisions under standard assumptions in Part I, solving the two-player game by backwards induction to derive the individual payoff-maximizing strategies. I will then briefly discuss some behavioral elements that might lead to deviations from these strategies. The last subsection lays out predictions for the rule valuations elicited in Part II, and discusses how risk aversion might lead to different WTPs between treatments.

2.3.1 Part 1

2.3.1.1 SPNE predictions

Within a round, pairs of subjects play a finitely repeated game. Both players know the fifth period is the final period, so the subgame perfect Nash equilibrium (SPNE) for the round is simply a repeated version of the one-shot SPNE. In the following, I will denote player i's payoff from sticking to the BSR as π_i^{bsr}, and the payoff after renegotiations as π_i^{new}.

Decision 2: Renegotiate contributions or cancel the project.
If either player decides to opt out of the burden sharing rule, both players will face a choice between paying θ to renegotiate contributions and canceling the project entirely. The latter will produce zero profit and leave both partners with only their endowment of ECU 90, so the following strategy would maximize payoff.

Prediction 2.1. *Decision 2: Player i will agree to renegotiation of contributions ($D_{2i} = 1$) iff*

$$\pi_i^{new} \geq 90 \quad \iff \quad r_i + r_j \geq 100 + 2\theta \tag{2.3}$$

If (2.3) holds, the sum of revenues exceeds the project costs by more than the total costs of negotiation. In other words, the gains from cooperation exceed their costs. This condition is identical for both partners, i.e., it either holds for both of them or for neither of them.

Decision 1: Stick to the BSR or opt out.
The first condition for deciding to stick to the BSR ($c_r = 50$) is that it does not lead to a

loss, since zero loss can always be achieved by simply not funding the project. By design, the lowest possible individual revenue is 50, so this condition is always met. The second condition is that one cannot increase profit by renegotiating contributions. Whenever $r_i < r_j$, player i would contribute less than 50 ($c_i < c_r$) if contributions were renegotiated. Even so, that would only increase her profit if the reduction in contribution is greater than the cost of negotiation. That is the case if $\theta < \frac{r_i - r_j}{2}$, which is equivalent to

$$90 + r_i - 50 \quad < \quad 90 + r_i - (50 - \tfrac{r_i - r_j}{2}) - \theta,$$

or $\pi_i^{bsr} < \pi_i^{new}$. This gives us the predicted strategy for Decision 1.

Prediction 2.2. *Decision 1: Player i will stick to the BSR ($D_{1i} = 1$) iff*

$$\pi_i^{bsr} \geq \pi_i^{new} \qquad \Longleftrightarrow \qquad r_j - r_i \leq 2\theta, \tag{2.4}$$

Note that when (2.4) does not hold, condition (2.3) is satisfied for both players.[13] This means that whenever one player can increase her profit by renegotiating contributions, it is optimal for the other player to agree to renegotiations. The design of the revenue profiles ensures that, for a given θ, the ex ante probability that condition (2.4) does not hold is the same for both partners.

Further note that this assumes that subjects stick to the rule when $\pi_i^{bsr} = \pi_i^{new}$, i.e. they do not reduce their partner's profit if it does not increase their own. For the project to be financed according to the BSR, both partners need to agree. We can therefore use (2.4) to construct a prediction on the self-enforcing range of the rule:

Prediction 2.3. *The BSR will be honored ($\{D_{11}, D_{12}\} = \{1, 1\}$) iff*

$$-2\theta \leq r_1 - r_2 \leq 2\theta. \tag{2.5}$$

2.3.1.2 Cooperation

Predictions 1 and 2 describe the combination of strategies from which one cannot improve by a unilateral deviation. The players' expected payoffs would be greater if, instead, both players were to follow a strategy where $D_{1i} = 1$ under all circumstances, or at least when their partners stuck to the BSR in all previous periods. Table 2.3 shows the set of payoffs that would result from this alternative set of strategies for subjects in the *AB* sessions. Given the prohibitive level of negotiation costs in round 3 ($\theta = 20$), this is equivalent to the SPNE payoffs. For the other two rounds, the average payoff would increase by ECU 5 for both types *A* and *B*. Although the expected payoff is still equal ex ante, the

[13]By design, $\pi_i^{bsr} > 90$, so $\pi_i^{new} > \pi_i^{bsr}$ implies $\pi_i^{new} > 90$.

payoff risk now entirely befalls the B player, while the A player has a fixed payoff level. The attractiveness of this outcome would therefore not be the same between types if risk appetites are non-neutral.

Table 2.3: Payoffs under the BSR for each draw of r_B ($r_A = 90$)

$\boldsymbol{r_B}$	**50**	**70**	**110**	**130**
$\pi_{\boldsymbol{A}}$	130	130	130	130
$\pi_{\boldsymbol{B}}$	90	110	150	170

These hypothetical strategies do not constitute an equilibrium under assumptions of pure self-interest for a finite number of interactions. Even so, it is not uncommon to see cooperative behavior in experimental versions of repeated games, like the public good game or the prisoner's dilemma.[14] The extent to which this occurs depends on the type of game and how it is framed.[15] Cooperation usually diminishes towards the final period. Fischbacher and Gachter (2010) find that in public good games, cooperation unravels because on average, subjects cooperate conditional on other's cooperation, but do so imperfectly, i.e., they match contributions only partly. Most people are not complete free-riders, and the presence of free-riders is not necessary for, but accelerates, the unraveling of cooperation.

The strategic element in cooperation in the current setting is different from that in a repeated prisoners dilemma or public good game, however. First, at most one of the two partners can be clearly cooperative in any given period, in that her choice to stick to the BSR is not in line with pure self-interest. In some periods, both partners will stick to the rule even if they employ purely selfish strategies. Second, a participant might expect her partner to reciprocate cooperative behavior in a later round, but will not know whether her partner will get a chance to do so, as this depends on the draws of the project revenues. It is not entirely clear how this atypical structure might affect cooperation levels. A fairly straightforward line of reasoning is that not knowing whether and how often your partner will be able to reciprocate reduces the incentive to cooperate. However, fewer opportunities to show one's intention to cooperate could also slow down the unraveling process inherent to (imperfect) conditional cooperation.

Another point of interest is whether the AB risk asymmetry *per se* affects the tendency to cooperate. There is evidence from previous experimental studies that assymetry in payoffs (Ahn et al., 2007; Beckenkamp et al., 2007) or endowments (Cherry et al., 2005) reduces cooperation. If asymmetry in risk has a similar impact, one may observe less cooperative behavior from B's depending on the type of their partner, which would lead to differences in outcomes between AB and BB matched pairs.

[14] See e.g. Andreoni and Miller (1993) and Cooper et al. (1996).

[15] See e.g. Andreoni (1995), Sonnemans et al. (1998) for cooperation differences in positive vs. negative frames.

2.3.1.3 Other deviations from SPNE

Impulsive punishment. In Decision D_2, there is no difference between a cooperative and a non-cooperative decision, as interests are fully aligned. One could, however, decline renegotiations as a way of punishing the partner for opting out of the BSR. In early periods this could be part of a strategy in hope of future cooperation, but it could also be a less strategic response driven by a mere desire to punish. The latter would indicate that, by being the default and the only efficient option, the rule has some normative effect on behavior (similar to that found by Irlenbusch (2004)), thereby creating some sense of entitlement even when exogenously imposed. The punishment then is a sign of the type of hostility Hart and Moore (2008) described, although in this experiment that action is costly as it equally reduces one's own payoff. As θ increases, the welfare destroyed by opting out of the BSR becomes bigger, while the cost and impact of a D_2–punishment become smaller.

Distributional preferences. Per the design implemented here, the renegotiation outcome is inefficient but equitable, while sticking to the BSR is efficient but gives unequal payoffs. Which of the two maximizes individual payoff depends on the draw of revenues, as described by Prediction 2.2. If subjects are inequity averse, they will find renegotiating more attractive, counteracting the impact of its cost θ. This could reduce the self-enforcing range of a BSR relative to the SPNE prediction, if subjects opt out of the rule even without an expected improvement in their individual payoff. Also, aversion to unequal payoffs makes the type of cooperation discussed above less valuable from an ex ante perspective. A preference for efficiency would have the opposite effect. Renegotiation always reduces welfare $(\pi_i + \pi_j)$ by 2θ, which may deter subjects from opting out of the BSR even if it would increase their own payoff. This would expand the self-enforcing range of the BSR. It also makes cooperation across rounds more attractive. Both equality and efficiency might play a role in the decisions individuals make. Experimental literature suggests that efficiency is the stronger of the two, both as individual preference (Engelmann and Strobel, 2004)[16] and as a feature to coordinate on (Isoni et al., 2014). This would justify the assumption that subjects choose the BSR option when $\pi_i^{bsr} = \pi_i^{new}$. The decision structure of this experiment might favor the efficient choice, as it is the first option. In terms of outcome, however, one could argue that the power to veto the BSR option is favorable to equal payoffs.

2.3.2 Part 2

As described in section 2.2, I use a random price mechanism (Becker et al., 1964) to elicit the maximum willingness to pay for the BSR, which garantuees that the optimal strategy

[16] Further studies include Kritikos and Bolle (2001) and Charness and Rabin (2002).

for each subject is to submit the true valuation. This valuation is most likely to be based on negotiation cost θ and on the expected probability of avoiding those cost with a BSR in place. Additionally, the effects on payoff risk allocation may be taken into account.

Given the cost of negotiation, we can use the SPNE predictions stated in section 2.3.1 to construct the expected value EV of a BSR. This EV does not take into account payoff risk, but serves as a predicted WTP for risk-neutral individuals. The cost of renegotiation affects this value through two channels. Firstly, a greater θ means more wasteful negotiation costs are avoided when both partners stick to the rule. Secondly, by Prediction 2.2, a greater θ increases the probability that both partners do stick to the rule.

Table 2.4: Expected Value (EV) of BSR

	$\theta = 10$	$\theta = 20$
AB	5	20
BB	4.375	10

Table 2.4 shows the EVs, defined as the increase in expected payoff from establishing the rule.[17] Because both contributions under the rule and expected project revenues are equal for all types, this value is identical between partners, also in the asymmetric AB setting. However, risk-averse individuals may take into account not only the expected payoff but also variance in individual payoff, which for AB pairs *is* impacted asymmetrically by the BSR. When contributing according to the rule, payoff risk is not shared between the partners, as we saw in Table 2.3 in the previous section.

This asymmetry is strongest for $\theta = 20$, such that condition (2.4) is guaranteed to be met and AB pairs can always be expected to stick to the rule. Table 2.5 displays the payoffs of types A and B for every contingency r_B, comparing the situation in which contributions are always negotiated to that in which they are fixed at 50, assuming this individual was selected to pay for the rule's establishment. Note that the design incentivizes subjects to indicate the maximum WTP for which they prefer the row of π_i^{bsr} to that of π_i^{new}.

Table 2.5: Payoffs with and without BSR ($r_A = 90$, $\theta = 20$)

$\boldsymbol{r_B}$	**50**	**70**	**110**	**130**
$\pi_{\boldsymbol{A,B}}^{\boldsymbol{new}}$	90	100	120	130
$\pi_{\boldsymbol{A}}^{\boldsymbol{bsr}}$	130 – WTP	130 – WTP	130 – WTP	130 – WTP
$\pi_{\boldsymbol{B}}^{\boldsymbol{bsr}}$	90 – WTP	110 – WTP	150 – WTP	170 – WTP

A risk-neutral subject might submit a WTP of 20 both as type A and as type B, as

[17] Note that the expected values for BB pairs differ from the AB setting because they face a different (larger) set of possible revenue combinations.

she is indifferent between potential payoff sets $\{70, 90, 130, 150\}$ and $\{110, 110, 110, 110\}$. A risk-averse subject would clearly not. The prediction is therefore a type A vs. B treatment effect on WTP among risk-averse individuals.

Prediction 2.4. *Part 2 (AB pairs): Types A will value a 50/50 BSR more than types B.*

$$WTP^A \ > \ WTP^B \tag{2.6}$$

There could be reasons rooted in distributional preferences for which subjects' WTP deviates from the rule's expected value EV. As mentioned before, negotiating contributions based on revenues results in equal payoffs between the partners, while sticking to the rule generally does not. This could reduce the BSR's value to individuals who place importance on payoff equality.

2.4 Findings

2.4.1 Part 1

2.4.1.1 Summary statistics

Out of the 1,260 potential projects, 1,092 (86.7%) were succesfully co-financed. The burden sharing rule was used 596 times (47.3%), while contributions were determined through renegotiations 496 times (39.4%). To have a first look at the choices subjects made, Tables 2.6 - 2.8 summarize the observations for the outcomes at pair level and for the two individual decisions. All three tables report overall and per treatment averages, with the last two columns zooming in on subsets of the observations based on the predictions from section 2.3.1.

For the pair outcomes, this first glance does not reveal a negative impact of asymmetry in revenue profiles on the rate at which the BSR is honored. In fact, the AB pairs stick to the rule more often than the BB pairs, especially when outside the predicted self-enforcing range.

Across all individual subjects, the prediction that subject i opts out of the BSR ($D_{1i} = 0$) when $\pi_i^{new} > \pi_i^{bsr}$ was accurate 73% of the time (Table 2.7). Subjects were predicted to stick to the rule when $\pi_i^{new} \leq \pi_i^{bsr}$, and they did so in 87% of those cases. These rates do not differ much between treatments, but one difference stands out among the conditional means (last column of Table 2.7): When an individual's profit can be improved by opting out of the BSR and renegotiating, types A choose to forego this opportunity much more often than types B do.

Table 2.6: BSR statistics at Pair level.

Pairing	All Periods	Outside Range	Inside Range
AB	0.487 (825)	0.275 (433)	0.722 (392)
BB	0.446 (435)	0.228 (241)	0.716 (194)
All	0.473 (1260)	0.258 (674)	0.720 (586)

Note: Rate at which BSR was honored at pair-level ($\{D_{11}, D_{12}\} = \{1, 1\}$), divided over inside and outside the predicted self-enforcing range ($-2\theta \leq r_1 - r_2 \leq 2\theta$). Number of observations in parentheses.

Table 2.7: Statistics for Decision D_1

Pairing	Type	All Periods	$\pi^{bsr} \geq \pi^{new}$	$\pi^{bsr} < \pi^{new}$
AB	A	0.731 (825)	0.863 (613)	0.349 (212)
	B	0.718 (825)	0.896 (604)	0.231 (221)
BB	B	0.692 (870)	0.862 (629)	0.249 (241)
All	All	0.713 (2520)	0.873 (1846)	0.274 (674)

Note: Mean of observed D_1, i.e., the share of observed choices where the participant wanted to stick to the BSR. Divided over whether or not renegotiation would improve profit. Number of observations in parentheses.

Overall, subjects could safely assume their partners would not reject renegotiations in Decision D_2. The prediction for D_2, to renegotiate whenever $\pi_i^{new} > 90$, turned out to be correct about 90% of the time across all pairing and risk type treatments. The last two columns in Table 2.8 do show a difference of about 9 percentage points, indicating that subjects were more likely to decline renegotiations when it was their partner who opted out of the rule. These decisions show no difference between risk treatments.

Table 2.8: Statistics for Decision D_2

Pairing	Type	All Periods	$D_{1i} = 0$	$D_{1i} = 1$
AB	A	0.900 (402)	0.941 (220)	0.852 (182)
	B	0.891 (402)	0.930 (213)	0.847 (189)
BB	B	0.892 (424)	0.931 (233)	0.843 (191)
All	All	0.894 (1228)	0.934 (666)	0.847 (562)

Note: Mean of observed D_2, i.e. the share of observed choices where the participant was willing to renegotiate, provided this would generate a profit ($\pi^{new} > 90$). Divided over whether or not individual i opted out of the BSR in Decision D_1. Number of observations in parentheses.

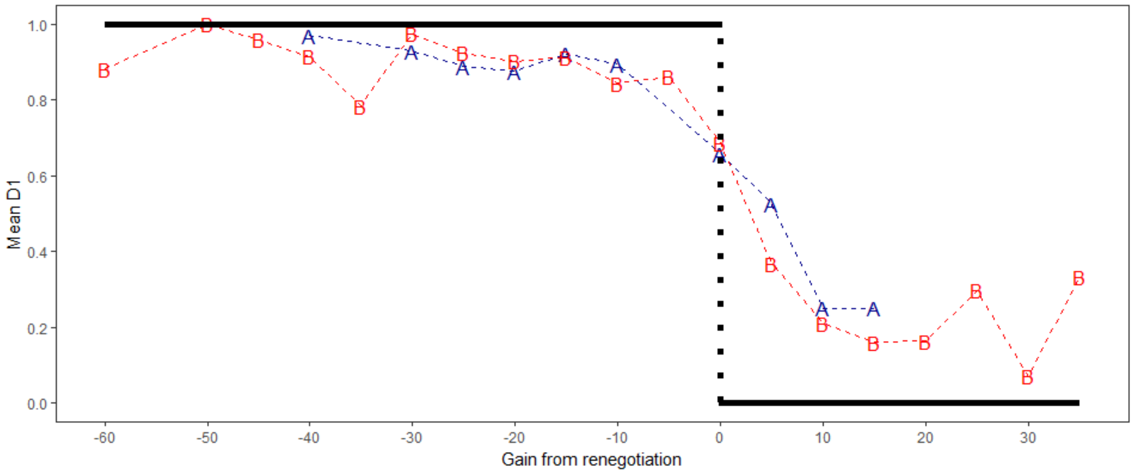
2.4.1.2 Decision 1

We saw in Table 2.7 that types A deviate from the SPNE prediction on D_1 more often than types B, refraining from opting out when renegotiation would increase the individual profit π_i. To further explore what factors are relevant for participants in the first decision, I estimate several different specifications of a random effects linear probability model.[18] Before presenting the estimations, the construction of two variables should be discussed, that I use to characterize the payoff decision individual i faces in a given period. The variable *gain* captures the increase in payoff i can expect from renegotiations ($gain_i = \pi_i^{new} - \pi_i^{bsr} = \frac{r_j - r_i}{2} - \theta$), while the variable g^+ is a binary variable that takes a value of 1 for $gain > 0$ and 0 otherwise. We can view g^+ as an indicator of whether the participant should stick to the BSR, according to Prediction 2.2.

Figure 2.1 shows the average Decision D_1 for both types, per *gain*. In the following, coefficient estimates for *gain* can be interpreted as the estimated slope of these lines, and those for g^+ as the drop as *gain* rises above zero. The latter would be the jump in the probability to opt out of the rule when doing so becomes payoff-maximizing. Note that if all subjects acted in line with Prediction 2, we would have a coefficient of zero for both *gain* and the risk type variable B, and a coefficient of exactly -1 for g^+ (with an intercept equal to 1). This predicted pattern is represented by the black line in the figure.

[18] Given the binary nature of D_1, a probit model might seem more appropriate, but does not allow for proper interpretation of interaction effects when combined with individual fixed effects (Ai and Norton (2003)). Qualitative insights on treatment effects do not change with this choice of model.

Figure 2.1: Average Decision D_1



Note: Mean of Decision D_1 for each type, for each value of *gain*–the net change in payoff from opting out. The black line represents the SPNE-prediction.

Figure 2.1 depicts average behavior for groups with varying numbers of observations, so one should be cautious when drawing conclusions just from looking at the graph. Nevertheless, the picture reveals that the observed behavior is broadly in line with SPNE, for both types A and B. It also confirms that the largest treatment differences occur when *gain* is positive. Note that, by design, the range of potential gains is smaller for both types in AB pairs than it is for participants in BB pairs, respectively $[-40, 15]$ and $[-60, 35]$, which is why the B-line covers a greater horizontal range.

The main set of estimations I will use to discuss the treatment difference in Decision D_1 is presented in Table 2.9. For these estimations the sample was restricted to observations from AB pairs.[19] There are six different specifications, subdividing the observations according to subjects' elicited risk preferences and the period within a round. The first period within a round means $t \in \{1, 6, 11\}$, i.e. a subject's first interaction with a new partner. In these periods, participants have not received any signal about their partner's strategy. Also, there are four periods with that partner still to come, so actions aimed at cooperation across periods, in hope of reciprocation later on, are most likely to surface here. For estimations (4)-(6), the *Risk Seeking* and *Inconsistent* were excluded.[20]

The last row presents estimates for the combined effect of B and $B \times g^+$, to compare type A to B conditional on $\pi^{bsr} < \pi^{new}$. The estimated combined coefficient is negative,

[19] Table 2.12 in Appendix 2.6.2 reports the estimated coefficients when BB observations were also included. It shows the *gain* coefficient for B's in AB pairs is significantly different from that for B's in BB pairs, indicating it may be more reasonable not to pool across the pairing treatments when analyzing A vs. B differences. Section 2.4.1.4 will provide an analysis of the difference between AB and BB pairs, and specifically of the behavior of types B in those pairs.

[20] The *Inconsistent* were excluded because most of them made at least one choice in line with risk-seeking preferences. Excluding only the consistent risk seekers does not change the results - the only notable change is that the combined effect of B and $B \times g^+$ in estimation (6) is no longer significant at the 10% level.

Table 2.9: Random Effects Linear Probability Estimations.
Dependent variable: D_1.

	All Risk Preferences			No Risk Seekers		
Explanatory Variables	(1) All Periods	(2) 1st Period	(3) Not 1st	(4) All Periods	(5) 1st Period	(6) Not 1st
B	0.025	0.107**	0.009	0.014	0.124***	-0.010
	(0.033)	(0.044)	(0.035)	(0.025)	(0.042)	(0.027)
g^+	-0.195**	-0.035	-0.208**	-0.171*	-0.024	-0.189*
	(0.078)	(0.162)	(0.088)	(0.090)	(0.183)	(0.101)
$B \times g^+$	-0.108	-0.325***	-0.073	-0.180**	-0.451***	-0.118
	(0.068)	(0.119)	(0.071)	(0.074)	(0.130)	(0.077)
$gain$	-0.007***	-0.007***	-0.007***	-0.007***	-0.008***	-0.007***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
$gain \times g^+$	-0.018***	-0.015	-0.019***	-0.019***	-0.012	-0.022***
	(0.006)	(0.012)	(0.007)	(0.006)	(0.013)	(0.007)
t	-0.008		-0.006	-0.010		-0.008
	(0.005)		(0.008)	(0.006)		(0.009)
Round	0.010	-0.033	0.003	0.021	-0.040*	0.016
	(0.027)	(0.023)	(0.039)	(0.033)	(0.024)	(0.048)
Observations	1,650	330	1,320	1,200	240	960
Number of ID	110	110	110	80	80	80
$B + [B \times g^+]$	-0.0830	-0.218**	-0.0638	-0.166**	-0.327***	-0.128*
	(0.063)	(0.111)	(0.066)	(0.072)	(0.124)	(0.074)

Note: Estimated effect on Prob$[D_{1i} = 1]$. Only asymmetric pairs (AB). Inclusion of triple interaction $[B \times g^+ \times gain]$ yields the same insights: the interaction term itself is never significant and no qualitative change in the other coefficients. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
$gain_i = \pi_i^{new} - \pi_i^{bsr} = \frac{r_j - r_i}{2} - \theta$.
$g^+ = 1$ iff $gain_i > 0$.

and significant when only first interactions are considered, or when risk seeking individuals are excluded. This means that when renegotiations would improve a subject's individual payoff, she is more likely to seize that opportunity and opt out of the BSR when her revenue fluctuates (B) than when it is fixed (A). The strongest effect is found when the two selection criteria are combined in estimation (5), where the coefficient indicates this probability difference between the types to be more than 32 percentage points.

The coefficients for t show a negative correlation between the period and decision D_1, but none is significantly different from zero. Table 2.13 in the Appendix reports the results of some additional estimations that further explore this correlation, by dividing the observations according to g^+, i.e., whether or not the individual could improve by opting out of the BSR and choosing $D_{1,i} = 0$. The coefficients for t are now significantly below zero, but only in those cases where opting out does improve payoff. This indicates that the cooperative action — foregoing an individual payoff improvement by sticking to the BSR — occurred less often as the partners approached the end of a round. This decrease in cooperation is especially prevalent among types A.

In the previous section, two potential motivations for such cooperative deviations from SPNE were discussed. First, subjects may place importance on their partner's payoff, and dislike the inefficiency of costly renegotiation.[21] Second, they may attempt to establish a mutually beneficial cooperation across periods where neither partner ever opts out of the BSR. For a risk-averse individual, the utility function is concave in monetary payoff, which means the relative importance of an increase in earnings is reduced as the initial level increases. It is therefore important to keep in mind that, in an AB pair, whenever the type A sticks to the BSR despite $\pi^{bsr} < \pi^{new}$, she still earns a profit of ECU 40 in that period, while the type B in the equivalent situation only earns a profit of ECU 0 or 20.

A higher initial payoff level may increase the relative importance of efficiency concerns, which would therefore be decisive more often for types A than for their partners. As illustrated in Section 2.3, a similar line of reasoning can explain why cooperation across periods, in which neither partner opts out, is less attractive to a risk-averse type B and more attractive to a risk-averse type A. The stronger treatment effects in first interactions suggests that at least some of it can be attributed to A's having more interest in establishing such cooperation.

2.4.1.3 Decision 2

If either or both partners opted out of the BSR, both would be asked to make Decision D_2: renegotiate contributions ($\pi_i = \pi_i^{new}$), or cancel the project ($\pi_i = 90$). Whenever $\pi^{new} > 90$, renegotiating is now the equitable, efficient and individual payoff-maximizing

[21] Inequity aversion was also mentioned, potentially causing deviations in the opposite directions. The observed choices here are in line with the previous finding that efficiency concerns are generally stronger than equity concerns. Appendix 2.6.3 shows this in more detail.

option. Therefore, the only reasonable motivation for refusing to negotiate is the desire to punish the partner by reducing her payoff to zero. Since it equally reduces one's own payoff, both the cost and the impact of the punishment grow with π^{new}.

We saw in Table 2.8 that, across all treatments, subjects very rarely refused to renegotiate, but did so more often when it was their partner who opted out of the BSR. This is confirmed by estimations of a random effects linear probability model (see Table 2.15 in Appendix 2.6.2), which shows a significant negative effect of the first decision D_{1i} on the second decision D_{2i}. When $D_{1i} = 1$, the partner must have opted out,[22] so this effect supports the idea that some subjects refuse renegotiations to punish their partner.

The estimations also show a positive effect of the renegotiation profit on D_{2i}. At first glance it this seems to capture the cost of punishment. However, interacting profit with D_{1i} reveals that the positive effect of profit is only significant when the subject opted out of the BSR herself ($D_{1i} = 0$). There is no indication of differences in choices between risk or pairing treatments.

When subjects do punish a non-cooperative partner by refusing renegotiations, this mostly seems to be an impulsive act rather than a strategic one. The coefficient for t is positive and weakly significant, reflecting that renegotiation is refused at a lower rate as partners approach their last interaction. Among potential punishers ($D_{1i} = 1$), this effect is insignificant. They reject renegotiations 19% of the time in the first period, and this drops to around 15% in periods 2-5. While this indicates there could be some punishers that hope to influence their partner's behavior in future periods, the majority of punishments appears to be of the impulsive kind.

2.4.1.4 Asymmetric vs. Symmetric pairing

Table 2.16 reports estimated effects on the probability that a pair of subjects funds a project using the burden sharing rule, so that $\{D_{11}, D_{12}\} = \{1, 1\}$. In line with the observations for individual decisions, the self-enforcing range (SER) consistent with SPNE strategies ($|r_1 - r_2| < 2\theta$) is the most important predictor. Additionally, both within and outside this range, the difference in revenue (labeled *gap*) has a significant negative impact. The cost of renegotiation increases the probability of a pair sticking to the rule, but this effect disappears outside the SER.

Across all specifications, the total profit a project can generate has a significant positive effect on the probability that a pair sticks to the BSR. This is consistent with the observation that individual participants place less importance on increasing their own payoff when the initial level is higher. A pairing treatment dummy and its interaction with *gap* were added to check for specific effects of the risk asymmetry. Specifications (2) and (6) show evidence of BB pairs using the BSR less often outside the SER for small revenue

[22] Otherwise one would not reach the screen for D_2.

differences, but that this rate is less responsive to further increases in *gap*.

As mentioned in section 2.3.1, previous studies have found payoff asymmetry to reduce cooperation. To compare the types B under symmetric and asymmetric risk, Table 2.14 reports estimated effects of the pairing treatment (AB vs. BB) on D_1. Specifications (1) and (3) take into account all periods, dividing them according to the SPNE prediction for the decision, and the weakly significant coefficients for BB suggest some difference in behavior. However, this seems to be driven by the periods in which the BB partners either both draw high revenues or both draw low revenues. No significant pairing treatment effect is found in specifications (2) and (4), where attention is restricted to periods that are more comparable to those occurring in the AB treatment.[23]

Based on the combined individual- and pair level analysis, there is no indication that risk asymmetry reduces cooperation. The opposite does not seem to be the case either. Although AB pairs do stick to the rule more often than BB pairs, this is driven by the types A being more cooperative. It is therefore more plausible that it is the lower overall risk level in the revenues of asymmetric pairs, rather than the asymmetry itself, that can lead to more efficient outcomes.

2.4.2 Part 2

The amounts participants were willing to pay to establish the 50/50 rule are summarized in Table 2.10. At first glance, the data does not seem to support the hypothesis that risk aversion will lead to A's valuing the BSR more than their B partners. Tables 2.17 and 2.18 in the appendix report estimates of the effects of a number of variables on the elicited WTP. Various treatment variables are included, as well as two measures obtained in Part 3: a dummy for risk aversion and the variable *give*.[24]

[23]This rules out certain combinations of r_i and g^+ that can occur in BB but not in AB pairs, e.g., when $r_i = 110$ but i can improve by opting out because $r_j = 130$.

[24]Obtained in a standard dictator game: *give* is the amount a participant gave to a randomly assigned partner, when she could allocate 100 ECU between the two of them.

Table 2.10: WTP statistics Part 2.

Pairing	Type	$\theta = 10$	$\theta = 20$
AB	EV	5	20
	A	10.7 (14.7)	12.6 (14.3)
	B	12.2 (15.9)	14.5 (15.3)
BB	EV	4.375	10
	B	10.6 (14.0)	14.1 (14.2)

Note: Average WTP in each round of Part 2, per risk type. Standard deviations in parentheses. The EV is based on expected avoided negotiation costs assuming SPNE strategies.

When confining the observations to AB pairs, I do not find any evidence for the predicted treatment effect. For $\theta = 10$, one could argue that the B's take into account the behavior by the A's in Part 1, D_1. Their greater tendency to stick to the rule, despite being able to improve their payoff by opting out, may have pushed up the rule's expected value for the B's, counteracting the unattractive impact on risk allocation. However, this cannot be the case when $\theta = 20$, since the costs of renegotiation are then so high that neither partner will ever improve by opting out. In the AB pairs, the estimated effect of being type B is always positive and insignificant, even when only including observations from round 2, and further restricting to risk averse participants who have not experienced a negative deviation from SPNE in round 3 of Part 1.[25] One possible explanation for not observing the predicted effect is that subjects are averse to risk borne by others. This would reduce the predicted WTP discrepancy between partners in AB pairs.

When controlling for risk aversion, or using it as a subsample criterion, the estimations show that the variable *give* significantly affects the elicited WTP. One can interpret *give* as the degree to which a subject cares for her partner's payoff, and it is generally greater for individuals that have a preference for equitable outcomes. We can see from the estimations in Table 2.17 that its relation to the WTP is dependent on the pairing treatment: subjects with a higher *give* value the BSR less in the AB treatment, but have a higher WTP in the BB treatment. One interpretation of this treatment-dependent relation, is that for BB pairs the rule is perceived as fairer, ex ante, as the two partners have identical sets of potential payoff outcomes. For the AB pairs, the potential payoff sets are not identical,

[25] Round 3 of Part 1 is the corresponding round in terms of negotiation cost θ.

even though the payoffs are equal in expected terms. One could imagine that individuals who care about the payoffs that others earn, are also more likely to care about the risk they bear. If so, this could explain the significant, negative coefficient for *give* among AB pairs. Although the implied prosociality in the risk domain would have to be very strong to completely neutralize the theoretically predicted valuation difference among risk averse individuals, it would make that difference less likely to be observed empirically.

2.5 Discussion and Conclusion

In this two-part experiment, I studied the use of an ex ante established burden sharing rule when ex post negotiations are costly. Groups consisted of two subjects who could both earn a revenue from their jointly funded project. Subjects were either assigned a risk-free or a risky revenue profile, meaning the project revenue was either known ex ante or drawn from a set of four possible values at the start every period. The first part of the experiment focused on the ex post willingness to stick to a rule that was available by default, while the second part measured how individuals value that same rule ex ante, by making its availability dependent on subjects' willingness to pay for it.

Overall, the burden-sharing rule functioned largely as theoretically predicted. The negotiation cost provides it with a self-enforcing range, within which subjects opt out of the rule very rarely. The inefficiency implied by the negotiation cost can extend this self-enforcing range, as some individuals refrain from forcing renegotiations even when it would increase their own payoff. Such actions can be viewed as cooperative, and in the experiment were always one-sided: in a given period, at most one of the partners was in a position to be cooperative, with no guarantees that the other would be in a position to return the favor later. Nevertheless, participants made the cooperative choice 27% of the time,[26] mostly when the foregone individual payoff gains were relatively small.

The main finding from the first part is that specifically this extention of the self-enforcing range of the BSR is sensitive to individual revenue risk. Subjects with fixed revenue, more often than their partners with varying revenue, stick to the rule when renegotiation would increase their profit. One explanation for the difference is that the efficiency of sticking to the rule is a more important feature at higher absolute payoff levels. Another explanation is that renegotiations, although costly, do allow the cost division to adjust to varying revenue levels, and thereby serve as a way to risk-share. From a risk perspective, reducing the frequency of renegotiations therefore appeals more to the subject with no individual revenue risk, implying an increased incentive to establish cooperation across periods that expands the self-enforcing range of the rule. The data offers support for both of these explanations.

In theory, the same reasoning applies to the rule itself. Assuming an aversion to risk,

[26] Cooperation defined as $[D_{1i} = 1 | \pi_i^{bsr} < \pi_i^{new}]$.

any rule that reduces renegotiation is more attractive to the agent with the lower individual risk. This is of interest because it implies that when two agents negotiate the rule, the agent with the higher individual risk has a better bargaining position and will achieve a more beneficial division of costs. Testing for this more directly in Part II, I find no evidence to support the theoretical prediction. While risk asymmetry affected the ex post choices by participants in Part I, there is no indication that it entered their ex ante valuation in Part II.

Previous studies have shown payoff asymmetry to reduce cooperation in games. The experiment presented here used a setting where, across treatments, payoffs were ex ante symmetrical, but in some treatments risk was not. Comparing pairs where both partners had a risky revenue profile to pairs where one partner's revenue was fixed, I find no indication of a particular negative effect of risk asymmetry on cooperation.

It is worth noting that the renegotiation outcome in this experiment was a mechanical function of differences in revenues, to have a clean look at the decision to stick to the burden sharing rule. It may be the case that if subjects actually bargain after opting out of the rule, the decision that precedes it affects bargaining positions by creating some hostility. One can imagine this to positively impact the self-enforcing range. Similarly, existing experimental literature provides empirical support for the idea that subjects have a behavioral preference for sticking to an agreement or promise they made, even when completely non-binding. Agreement formation was not part of the current experiment, as the burden sharing rule was exogenously imposed. It is conceivable that the self-enforcing range would be extended further if the rule was the product of a pre-play interaction between the partners. While such extra interactions are more complicated to implement and measure experimentally, they could be interesting to include in future variations to the current study.

References

Ahn, T.K., Myungsuk Lee, Lore Ruttan, and James Walker, "Asymmetric Payoffs in Simultaneous and Sequential Prisoner's Dilemma Games," *Public Choice*, 2007, *132* (3/4), 353–366.

Ai, Chunrong and Edward C. Norton, "Interaction terms in logit and probit models," *Economics Letters*, 2003, *80* (1), 123 – 129.

Andreoni, James, "Cooperation in Public-Goods Experiments: Kindness or Confusion?," *The American Economic Review*, 1995, *85* (4), 891–904.

_ **and John H. Miller**, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *The Economic Journal*, 1993, *103* (418), 570–585.

Bartling, Björn and Klaus M. Schmidt, "Reference Points, Social Norms, and Fairness in Contract Renegotiations," *Journal of the European Economic Association*, 2015, *13* (1), 98–129.

Beckenkamp, Martin, Heike Hennig-Schmidt, and Frank Maier-Rigaud, "Co-operation in Symmetric and Asymmetric Prisoner's Dilemma Games," *Working Paper Series of the Max Planck Institute for Research on Collective Goods*, 03 2007.

Becker, Gordon M., Morris H. Degroot, and Jacob Marschak, "Measuring utility by a single-response sequential method," *Behavioral Science*, 1964, *9* (3), 226–232.

Ben-Ner, Avner and Louis Putterman, "Trust, communication and contracts: An experiment," *Journal of Economic Behavior & Organization*, 2009, *70* (1), 106 – 121.

Bolton, Gary E. and Axel Ockenfels, "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment," *American Economic Review*, 2010, *100* (1), 628–633.

Brock, J. Michelle, Andreas Lange, and Erkut Y. Ozbay, "Dictating the Risk: Experimental Evidence on Giving in Risky Environments," *American Economic Review*, 2013, *103* (1), 415–437.

Charness, Gary and Matthew Rabin, "Understanding Social Preferences with Simple Tests*," *The Quarterly Journal of Economics*, 2002, *117* (3), 817–869.

Cherry, Todd L., Stephan Kroll, and Jason F. Shogren, "The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab," *Journal of Economic Behavior and Organization*, 2005, *57* (3), 357 – 365.

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross, "Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games," *Games and Economic Behavior*, 1996, *12* (2), 187 – 218.

Dufwenberg, Martin, Maros Servatka, and Radovan Vadovic, "Honesty and informal agreements," *Games and Economic Behavior*, 2017, *102*, 269 – 285.

Engelmann, Dirk and Martin Strobel, "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments," *American Economic Review*, 2004, *94* (4), 857–869.

Fehr, Ernst, Oliver Hart, and Christian Zehnder, "Contracts as Reference Points. Experimental Evidence," *American Economic Review*, 2011, *101* (2), 493–525.

_ , _ , and _ , "How do informal agreements and revision shape contractual reference points?," *Journal of the European Economic Association*, 2015, *13* (1), 1–28.

Fischbacher, Urs, "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 2007, *10* (2), 171–178.

_ **and Simon Gachter**, "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments," *American Economic Review*, March 2010, *100* (1), 541–56.

Hart, Oliver and John Moore, "Contracts as Reference Points," *Quarterly Journal of Economics*, 2008, *CXXIII*, 1–48.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *American Economic Review*, December 2002, *92* (5), 1644–1655.

Irlenbusch, Bernd, "Relying on a man's word?: An experimental study on non-binding contracts," *International Review of Law and Economics*, 2004, *24* (3), 299 – 332.

Isoni, Andrea, Anders Poulsen, Robert Sugden, and Kei Tsutsui, "Efficiency, Equality, and Labeling: An Experimental Investigation of Focal Points in Explicit Bargaining," *The American Economic Review*, 2014, *104* (10), 3256–3287.

Kalai, Ehud and Meir Smorodinsky, "Other Solutions to Nash's Bargaining Problem," *Econometrica*, 1975, *43* (3), 513–518.

Kessler, Judd B. and Stephen Leider, "Norms and Contracting," *Management Science*, 2012, *58* (1), 62–77.

Kritikos, Alexander and Friedel Bolle, "Distributional concerns: equity- or efficiency-oriented?," *Economics Letters*, 2001, *73* (3), 333 – 338.

Nash, John F., "The Bargaining Problem," *Econometrica*, 1950, *18* (2), 155–162.

Pieters, Arne, "Repeated Burden Sharing and Costly Negotiations," *Working Paper*, 2019.

Potters, Jan, Arno Riedl, and Franziska Tausch, "An experimental investigation of risk sharing and adverse selection," *Journal of Risk and Uncertainty*, 2014, *48* (2), 167–186.

Sonnemans, Joep, Arthur Schram, and Theo Offerman, "Public good provision and public bad prevention: The effect of framing," *Journal of Economic Behavior and Organization*, 1998, *34* (1), 143 – 161.

Tan, Fangfang, "Punishment in a Linear Public Good Game with Productivity Heterogeneity," *De Economist*, Sep 2008, *156* (3), 269–293.

2.6 Appendix to Chapter 2

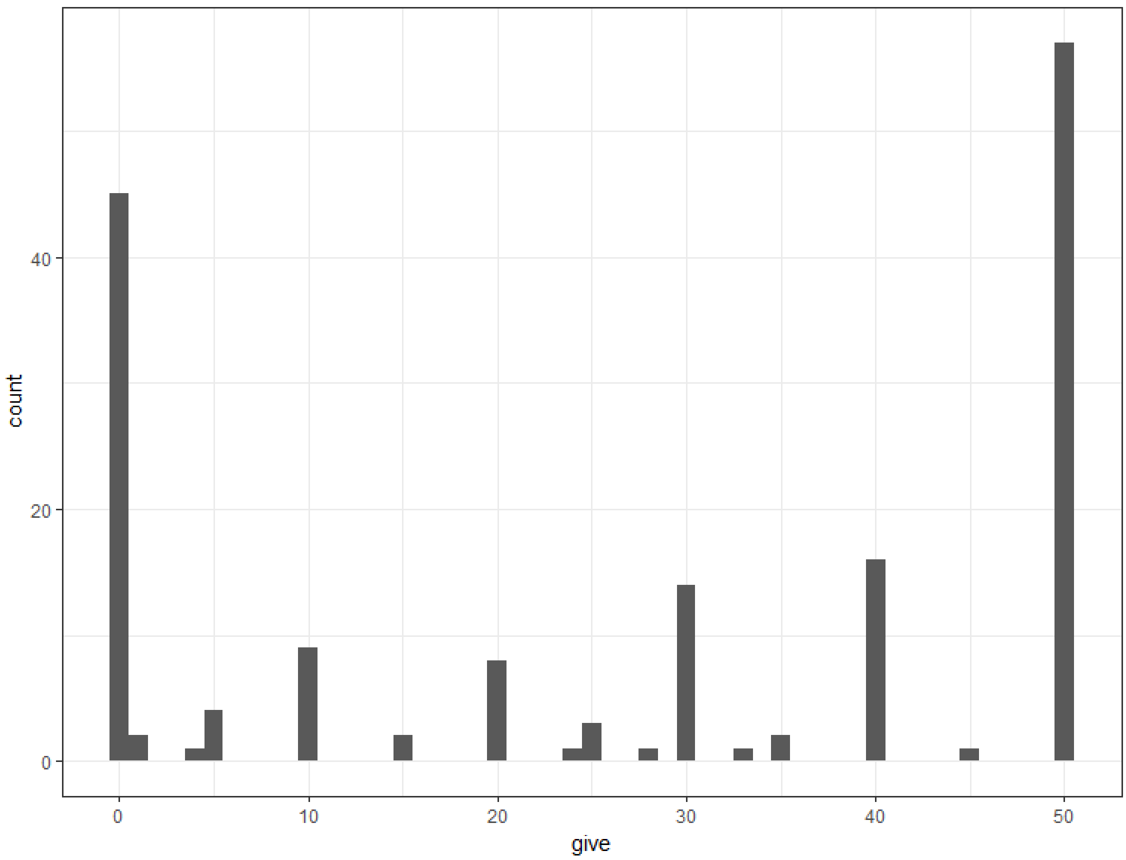
2.6.1 Part 3

After completing Parts 1 & 2 of the experiment, the subject did two short choice tasks to provide some insight in their individual preferences:

3a. Dictator Game. One task was the well-known Dictator Game. Subjects were randomly paired, and were told that one of them would be randomly selected into the role of 'Dictator', while the other would be the passive 'Receiver'. The Dictator is endowed with 100 ECU, and simply chooses how to split that amount between him or herself and the Receiver. All subjects were asked to indicate their chosen action as Dictator, only to be implemented if they were assigned that role.

Figure 2.2 shows how these choices were distributed among the 168 subjects.[27]

Figure 2.2: Dictator Game



Note: Histogram of 'give', the amount (out of 100) subjects gave to their partner.

[27] There was one subject that gave all the 100 ECU to his/her partner, the only instance where $give > 50$.

3b. Risk Preference Elicitation Task. To have an indication of the subjects' risk preferences, they made ten choices in a menu of lottery options equivalent to that introduced by Holt and Laury (2002). The choices are ordered in such a way that the risky option's expected value increases relative to the less risky option. The risk preference is inferred from the switching point, the row at which the subject switches from the safer lottery to the riskier one. The payoffs and probabilities are constructed so that individuals with preferences very close to risk neutrality choose option A (safer) in the first four decisions and option B (riskier) in the remaining six decisions.

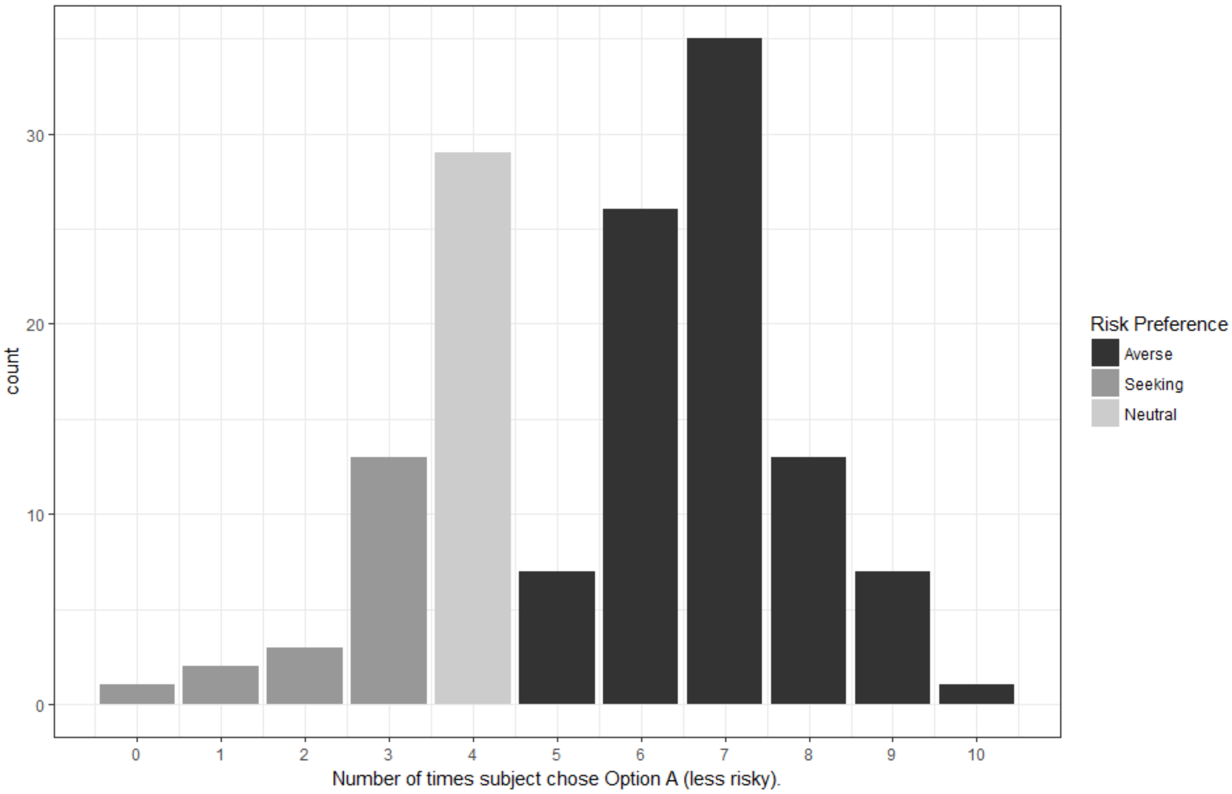
Figure 2.3 shows the distribution of switching points, table 2.11 shows how this was mapped onto risk preference categories. If a participant went back and forth between A and B, switching both before and after the fifth decision, no risk preference was inferred (label: *Inconsistent*).

Table 2.11: Risk Preferences

Switching Point $A \to B$	Label	N
(1) – (4)	*Risk Seeker*	20
(5)	*Risk Neutral*	29
(6) – (10)	*Risk Averse*	89
More than once	*Inconsistent*	30

Note: Distribution of 168 participants across risk preference categories.

Figure 2.3: HL Risk Task



2.6.2 Regression results

Table 2.12: Random Effects Linear Probability Estimations.
Dependent variable: D_1

Expl. Variables	(1)	(2)
B	0.054	0.030
	(0.047)	(0.054)
$gain$	-0.007***	-0.007***
	(0.001)	(0.001)
$B \times gain$	0.003**	0.000
	(0.001)	(0.002)
g^+	-0.336***	-0.312***
	(0.072)	(0.072)
$B \times g^+$	-0.139*	-0.122
	(0.079)	(0.088)
BB		0.020
		(0.048)
$BB \times g^+$		0.026
		(0.086)
$BB \times gain$		0.004**
		(0.002)
$g^+ \times gain$	-0.001	-0.004
	(0.003)	(0.003)
t	-0.005	-0.005
	(0.004)	(0.004)
Round	-0.004	-0.006
	(0.024)	(0.024)
Observations	2,520	2,520
Number of ID	168	168

Note: Estimated effect on Prob[$D_{1i} = 1$]. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
$gain_i = \pi_i^{new} - \pi_i^{bsr} = \frac{r_j - r_i}{2} - \theta$.
$g^+ = 1$ iff $gain_i > 0$.

Table 2.13: Random Effects Linear Probability Estimations.
Dependent variable: D_1

	Full Sample		*AB* Only		*AB* - *A*	*AB* - *B*
Explanatory	(1)	(2)	(3)	(4)	(5)	(6)
Variables	$gain > 0$	$gain \leq 0$	$gain > 0$	$gain \leq 0$	$gain > 0$	$gain > 0$
B	-0.071	0.028	-0.061	0.028		0.016
	(0.062)	(0.033)	(0.063)	(0.033)		(0.025)
BB	0.131**	-0.054**				
	(0.063)	(0.028)				
$gain$	-0.011***	-0.005***	-0.024***	-0.007***	-0.024***	-0.024***
	(0.003)	(0.001)	(0.006)	(0.001)	(0.009)	(0.008)
t	-0.018*	-0.002	-0.026**	-0.003	-0.036**	-0.020
	(0.010)	(0.005)	(0.012)	(0.005)	(0.013)	(0.017)
Round	-0.006	-0.010	0.035	-0.008	0.042	0.041
	(0.057)	(0.027)	(0.071)	(0.027)	(0.105)	(0.094)
Observations	674	1,846	433	1,217	212	221
Number of ID	168	168	110	110	55	55

Note: Estimated effect on Prob$[D_{1i} = 1]$.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Sampling criterion $\boldsymbol{gain > 0}$ implies $\boldsymbol{D_1 = 0}$ was payoff maximizing (assuming the partner would accept renegotiations). Excluding individuals with risk seeking preferences gives estimated coefficients for t of similar magnitude and p-value.

Table 2.14: Random Effects Linear Probability Estimations. Dependent variable: D_1.

Explanatory Variables	(1) $gain > 0$	(2) $gain > 0$	(3) $gain \leq 0$	(4) $gain \leq 0$
BB	0.109*	0.081	-0.054*	-0.030
	(0.064)	(0.068)	(0.027)	(0.028)
gain	-0.009***	-0.007*	-0.005***	0.001*
	(0.003)	(0.003)	(0.001)	(0.001)
t	-0.012	-0.017	-0.004	0.009**
	(0.013)	(0.013)	(0.006)	(0.005)
Round	-0.023	0.011	-0.004	-0.021
	(0.070)	(0.072)	(0.033)	(0.027)
Observations	462	443	1,233	826
Number of ID	113	112	113	113

Note: Estimated effect on Prob$[D_{1i} = 1]$. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Only types B included. Specifications (2) and (4) further restrict the sample to align revenue sets between types B in AB vs BB pairing treatments. For (2) this means $r_i \in \{50, 70\}$, in (4) only periods with large revenues are included: $r_i \in \{110, 130\}$.

Table 2.15: Random Effects Linear Probability Estimations.
Dependent variable: D_2.

Expl. Variables	(1) All	(2) All	(3) All	(4) All	(5) AB	(6) BB
$D_{1,i}$	-0.097***	-0.098***	-0.038	-0.060*	-0.021	-0.102***
	(0.021)	(0.021)	(0.051)	(0.036)	(0.067)	(0.033)
$(\pi^{new} - 90)$	0.002**	0.002***	0.003***	0.002**	0.003*	0.003*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
$D_{1i} \times (\pi^{new} - 90)$			-0.002			
			(0.001)			
Round		-0.029	-0.029		-0.021	-0.015
		(0.029)	(0.029)		(0.039)	(0.043)
t		0.009*	0.009*	0.009**	0.011*	0.006
		(0.005)	(0.005)	(0.005)	(0.006)	(0.008)
θ				-0.002		
				(0.004)		
$D_{1i} \times \theta$				-0.004		
				(0.004)		
B					0.060	
					(0.071)	
$D_{1i} \times B$					-0.149	
					(0.121)	
Observations	1,228	1,228	1,228	1,228	804	424
Number of ID	168	168	168	168	110	58
Total Effect^{+}			0.001	0.006		
			(0.001)	(0.005)		

Note: Estimated effect on Prob$[D_{2i} = 1 \mid \pi^{new} > 90]$. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
When limiting sample to observations where $D_{1,i} = 1$, the coefficients on t are still positive, but lose significance.
Total Effect^{+} estimates combined effect for:
(3) - renegotiation profit and its interaction with D_{1i},
(4) - θ and its interaction with D_{1i}

Table 2.16: Random Effects Linear Probability Estimations at pair level.

	(1)	(2)	(3)	(4)	(5)	(6)
Expl. Variables	All	All	In SER	In SER	Not in SER	Not in SER
$gap \leq 2\theta$	0.279***	0.267***				
	(0.057)	(0.057)				
BB	0.016	-0.128*	-0.031	0.006	0.074	-0.341***
	(0.036)	(0.072)	(0.059)	(0.086)	(0.053)	(0.129)
gap	-0.006***	-0.010***	-0.009***	-0.008***	-0.005***	-0.012***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)
$BB \times\ gap$		0.005**		-0.002		0.010***
		(0.002)		(0.004)		(0.003)
t	-0.008	-0.008	0.006	0.005	-0.020*	-0.019*
	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)	(0.011)
θ	0.009**	0.009**	0.015***	0.015***	-0.001	-0.000
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Total Profit	0.001***	0.001***	0.001**	0.001**	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Observations	1,233	1,233	559	559	674	674
Number of PID	252	252	182	182	193	193

Note: Dependent variable: Stick to BSR (=1 if $\{D_{11}, D_{12}\} = \{1, 1\}$).
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Only includes observations where Total Profit ¿ 0.
Total Profit $\equiv r_1 + r_2 - 100 > 0$.
$gap \equiv |r_1 - r_2|$.

Table 2.17: Random Effects Estimation.
Dependent variable: WTP.

Expl. Variables	(1) All	(2) All	(3) *AB*	(4) *AB*	(5) *AB*	(6) *AB*
EV	0.17***	0.19***	0.14***	0.16***	0.11*	
	(0.04)	(0.04)	(0.05)	(0.04)	(0.06)	
B	1.84	1.22	1.94	1.12	2.28	4.54
	(2.80)	(1.72)	(2.83)	(1.73)	(4.51)	(4.01)
give	-0.07	-0.11	-0.06	-0.10*	-0.16*	-0.20**
	(0.06)	(0.07)	(0.06)	(0.06)	(0.09)	(0.07)
RiskAv		6.66*		6.64*		
		(3.61)		(3.61)		
RiskAv × *B*		0.99		1.20		
		(4.64)		(4.89)		
BB	-4.82	-1.32				
	(3.94)	(2.58)				
BB × *give*	0.18*	0.23***				
	(0.09)	(0.09)				
RiskAv × *BB*		-7.93**				
		(3.77)				
Avg. Revenue Pt. 1	0.12	0.27*	0.34	0.33	0.36	0.42*
	(0.15)	(0.15)	(0.25)	(0.28)	(0.46)	(0.21)
Observations	336	274	220	186	120	27
Number of ID	168	137	110	93	60	27

Note: Estimated effect on WTP. Robust standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Dummy variable *RiskAv* = 1 if categorized as *Risk Averse*, 0 otherwise. Including *B* × *give* in regressions (4)-(6) shows no interaction effect.
Selection criteria in addition to pairing treatment as indicated in table:
(2)+(4): Participants excluded when Risk Preference = *Inconsistent*.
(5)+(6): Only participants with Risk Preference = *Risk Averse*.
(6): Only 2nd round in Part2 (so here OLS not Random Effects), where revenue draws are always inside the SER as $\theta = 20$. Only participants that experienced no deviations from SPNE in round 3 of Part 1.

Table 2.18: Random Effects Estimation.
Dependent variable: WTP.

Expl. Variables	(7) BB	(8) BB
EV	0.56***	0.62***
	(0.12)	(0.17)
give	0.09	0.15**
	(0.07)	(0.07)
Part 1:		
Avg. Revenue	-0.17	0.05
	(0.19)	(0.16)
Neg. Deviations	2.49	1.00
	(1.89)	(1.82)
Pos. Deviations	3.01*	3.61**
	(1.63)	(1.72)
Observations	116	58
Number of ID	58	29

Note: Estimated effect on WTP. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
"Deviations" variables count the number of positive or negative deviations from SPNE, in terms of BSR honored or not, a participant experienced in Part 1. When included in estimations (1)-(6), their estimated effects are insignificant and that of other variables unchanged. Including a dummy variable for risk aversion in specification (7) does not alter results, and its own significant is insignificant.
(7): All observations in the *BB* pairs.
(8): Only participants categorized as *Risk Averse*.

2.6.3 Distributional Choices

This section consists of several tables presenting overall statistics on subject choices in Decision 1. Assuming the partner will agree to renegotiations at Decision 2 ($D_{2j} = 1$), these choices are between two sets of payoffs. One set is that with BSR payoffs ($D_{1i} = 1$), the efficient but unequal option, and one set is that with the renegotiation payoffs ($D_{1i} = 1$), the equal but inefficient option. The round-specific negotiation costs θ and individual project revenues (R_i, R_j) determine which of the two options maximizes (i)the own payoff, (ii) the lower of the two payoffs, and the choices are sorted into table accordingly.[28]

One should be careful to directly draw conclusions from these numbers, as they are aggregations of observed behavior in a repeated strategic setting. What we can see, however, is that subjects more often give up profit to achieve the efficient outcome (Table 2.19) than they do to equalize payoffs (Table 2.22). In periods where the own payoff would not change by renegotiating ($\pi_i^{bsr} = \pi_i^{new}$, Table2.20), subjects choose the efficient option about two thirds of the time.

Note: in the table descriptions, maximin preferences refer to the choice that maximizes the smaller of the two payoffs: $\max\limits_{D_{1,i}} \min\{\pi_i, \pi_j\}$.

Table 2.19: Choices where opting out of the BSR ($D_1 = 0$) corresponds to selfish and maximin preferences.

R_i	R_j	θ	BSR Profits*	Renegotiation Profits*	Difference: $\pi^{new} - \pi^{bsr}$	$D_{1i} = 1$	N
90	110	5	(40,60)	(45,45)	5	53%	76
70	90	5	(20,40)	(25,25)	5	37%	63
90	130	5	(40,80)	(55,55)	15	25%	64
50	90	5	(0,40)	(15,15)	15	14%	72
90	130	10	(40,80)	(50,50)	10	25%	72
50	90	10	(0,40)	(10,10)	10	21%	86

*** First entry is subject's own profit, second entry is the partner's profit. These are the earnings on top of the 90 endowment.**

[28] These tables show statistics for 24 different choices in the *AB* treatment. No *BB* statistics are included because this would add another 48 choices with only a small number of observations per choice.

Table 2.20: Choices where BSR and renegotiation are equivalent in selfish and maximin preferences.

R_i	R_j	θ	BSR Profits*	Renegotiation Profits*	Difference: $\pi^{new} - \pi^{bsr}$	$D_{1i} = 1$	N
70	90	10	(20,40)	(20,20)	0	69%	48
90	110	10	(40,60)	(40,40)	0	62%	69
50	90	20	(0,40)	(0,0)	0	71%	68
90	130	20	(40,80)	(40,40)	0	70%	60

Table 2.21: Choices where BSR and renegotiation are equivalent in maximin preferences, while BSR is the selfish option.

R_i	R_j	θ	BSR Profits*	Renegotiation Profits*	Difference: $\pi^{new} - \pi^{bsr}$	$D_{1i} = 1$	N
110	90	10	(60,40)	(40,40)	-20	96%	69
90	70	10	(40,20)	(20,20)	-20	88%	48
130	90	20	(80,40)	(40,40)	-40	95%	60
90	50	20	(40,0)	(0,0)	-40	97%	68

Table 2.22: Choices where BSR is favored by maximin and selfish preferences.

R_i	R_j	θ	BSR Profits*	Renegotiation Profits*	Difference: $\pi^{new} - \pi^{bsr}$	$D_{1i} = 1$	N
70	90	20	(20,40)	(10,10)	-10	86%	71
90	110	20	(40,60)	(30,30)	-10	89%	76
90	70	20	(40,20)	(10,10)	-30	96%	71
110	90	20	(60,40)	(30,30)	-30	96%	76

Table 2.23: Choices where renegotiation is favored by maximin preferences, while BSR is the selfish option.

R_i	R_j	θ	BSR Profits*	Renegotiation Profits*	Difference: $\pi^{new} - \pi^{bsr}$	$D_{1i} = 1$	N
90	70	5	(40,20)	(25,25)	-15	92%	63
110	90	5	(60,40)	(45,45)	-15	97%	76
90	50	5	(40,0)	(15,15)	-25	89%	72
130	90	5	(80,40)	(55,55)	-25	91%	64
90	50	10	(40,0)	(10,10)	-30	91%	86
130	90	10	(80,40)	(50,50)	-30	99%	72

2.6.4 Instructions

2.6.4.1 Experimental Instructions Before First Part (English Translation)

Thank you for you participation in this economic experiment. Communication with other participants is not allowed and a violation of this rule will lead to exclusion from the experiment as well as from all payments. This experiment will consist of 4 parts. The first part is described below. You will receive the instructions for the other parts after the previous part has ended.

Payment
In each part you can earn money. Your decisions in one part are not relevant for payments from other parts of the experiment. At the end of the experiment the payments from Part 1, Part 2, Part 3 and Part 4 will be added up.
You earnings will be in Experimental Currency Units (ECU), which is converted in the end with the following exchange rate:

24 ECU = 1 EURO

PART 1

In Part 1 you will play 15 rounds. At the end of every round you will see your income in that round. At the end of the experiment one of these 15 rounds will be randomly selected to for payment. Each round has the same probability of being selected.

You play every round with a partner, who is a randomly selected participant in this room. You keep the same partner for 5 rounds. So you will have the same partner for rounds 1-5, for rounds 6-10, and for rounds 11-15.

Everyone starts each round with 90 ECU in their account. You and your partner can earn additional income by jointly financing a profitable project:

Project Revenues [29]

The project generates revenue to both partners. The revenue from the project can be 50, 70, 110 or 130 ECU, all with equal probability. It will be randomly determined in every round, independent of the revenue in previous rounds.

The information sheet gives a summary of the revenue profile.

[29] For BB treatment.

Project Revenues [30]

The project generates revenue to both partners. This revenue depends on your Profile. At the start of Part 1 you will see what your Profile is. You can have profile A or B, and you are always assigned a partner that has the other profile. All participants will keep the same profile throughout the 15 rounds.

For Profile A: The revenue from the project is a certain amount of 90 ECU.
For Profile B: The revenue from the project can be 50, 70, 110 or 130 ECU, all with equal probability. It will be randomly determined in every round, independent of the revenue in previous rounds.

The information sheet gives a summary of revenue profiles A and B.

Financing the Project[31]

The cost of the project is always 100 ECU. It will only be financed if you and your partner both pay a share of the total costs. There are two ways to determine who contributes how much:

> **Option 1: Standard Contributions (free)**: Both partners pay 50 ECU. No extra cost is incurred.
>
> **Option 2: Computer-determined Contributions (costly)**: The computer determines the contributions, such that you both make the same profit, i.e. it adjusts for differences in revenue.
> An extra cost ***C*** is incurred on both partners.

At the start of every round, both partners observe their own and each other's revenue for that round.

In **Screen 1**, you make Decision 1: you decide whether you want to stick to the **Standard Contributions** and pay 50 ECU (see Screen 1 on the information sheet).

If you both answer "Yes", 50 ECU will be deducted from both your and your partner's account, and the joint project will be automatically financed. You then receive the indicated revenue. No further costs are incurred.

- *See Profit Calculation 1 on the information sheet.*

If you do not both answer "Yes", you will move on to **Screen 2**, where both partners make Decision 2 (see Screen 2 on the information sheet).

[30] For *AB* treatment.
[31] Remainder of instructions identical between treatments.

In Decision 2 you decide whether or not you want to let the **Computer determine Contributions**. Only if you both answer "Yes", the computer will calculate contributions based on your revenues in this round. These will then be automatically deducted from both your respective accounts to finance the joint project, so that you and your partner receive the indicated revenue. Additionally, administrative costs ***C*** will be deducted from both accounts. These costs will be 5, 10 or 20 ECU, and are always the same for both partners. The information sheet shows the cost ***C*** for each round.

- *See Profit Calculation 2.*

If at least one of you chooses "No" in Decision 2, the project will not be financed, i.e. no contribution or administrative costs will be deducted from your accounts and you will not receive any revenue.

- *See Profit Calculation 3.*

The decisions made by you and your partner will not affect the choices you have in the next round.

In the "Decision Diagram" on the information sheet, all decisions and their consequences are once more illustrated. There you will also find examples for Screen 1 and 2 with descriptions. You will now go through some control questions on the computer, to make sure you understood these instructions. Please click OK on your screen to start.

2.6.4.2 Instructions Before Second Part (English Translation)

PART 2

Part 2 of this experiment is similar to Part 1, but will consist of only 2 rounds. One of the 2 rounds will be randomly selected for payout at the end of the experiment.

In this part, you will be randomly assigned a new partner for each round. You can again earn money by financing a project with your partner. The project costs and possible revenues are the same as in Part 1 (you keep the same revenue profile).

The difference with Part 1 is that Option 1 (Standard Contributions) is no longer automatically available.

At the start of each round, before you know the exact revenues for that period, you and your partner are both asked how much you would be willing to pay to make Option 1 available for that round.

If Option 1 is available you will consequently have the exact same decision as in Part 1. This means that you will see the respective project revenues for you and your partner after which you can choose between the same options: Option 1 (Standard Contributions: both partners contribute 50 ECU, if they both agree), Option 2(both partners pay $\boldsymbol{C}$ to have the computer determine contributions, if they both agree), or not finance the project at all.

If Option 1 is ***not*** *available*, you *cannot* use the Standard Contributions. In this case, after seeing the project revenues, you will directly go to Screen 2, where you decide between paying $\boldsymbol{C}$ to have the contributions be determined by the computer, and not financing the project at all.

The profit and contribution calculations are the same as in Part 1 (see information sheet). In the first round $\boldsymbol{C = 10}$, and $\boldsymbol{C = 20}$ in the second round.

Whether or not Option 1 is available, will be determined as follows:

1. You and your partner submit the maximum amount you are willing to pay to have Option 1 available.
2. The computer randomly selects one of you as the "Buyer" (both with equal probability).
3. The computer randomly generates a price $\boldsymbol{P}$.
4. Option 1 will be available only if $\boldsymbol{P}$ is below the maximum amount the Buyer has submitted:
   - If $\boldsymbol{P}$ is *below* the submitted maximum amount (or equal), the Buyer will pay $\boldsymbol{P}$, and Option 1 will be available. *The other partner pays nothing.*
   - If **P** is *above* the submitted maximum amount, the Buyer does not pay anything, and Option 1 will *not* be available.

When this procedure has taken place, you and your partner will be informed of whether or not the rule was established. At this point, you will not see which of you was selected, or what $\boldsymbol{P}$ was. You will obtain this information at the end of the experiment, if this round was selected for payment. Any amount $\boldsymbol{P}$ paid by the Buyer will then be deducted from his or her earnings.

Please read the example below and then fill in the control questions.

<u>Example:</u>

Assume you submit 10 ECU as the maximum amount you would be willing to pay, and you are selected by the computer.

- If the randomly generated price is ***P=7***, you pay 7 ECU and Option 1 will be available (since 7 ¡ 10). You will pay 7 ECU and you partner will pay nothing. If this round is selected for payment at the end of the experiment, 7 ECU will be deducted from your earnings.

- If the randomly generated price is ***P=12***, Option 1 will not be available (since 12 ¿ 10). In this case you and your partner will only be able to finance the project, if both of you agree to computer-determined contributions and pay administrative costs ***C*** for this.

2.6.4.3 Additional Information Provided

The above instructions were provided in written form and read aloud by a laboratory assistant. In addition, an information sheet summarized the payoff, revenue and decision structure of the experiment (as referenced in the instructions). These are reproduced below, and include screenshots of the afore mentioned Screen 1 and Screen 2.

Having read the instructions for Part 1, the subjects proceeded to answer several on-screen control questions before the actual experiment started, to make sure they understood what decision they were asked to make in which screen, and how their and their partner's decisions would determine their earnings. The instructions for Part 2 were followed by a pen-and-paper control exercise that made sure they understood how the random price mechanism (Becker et al. (1964)) worked.

INFORMATIONSBLATT

Erlösprofile im Überblick:

	Erlös	Wahrscheinlichkeit
IHR PROFIL: *A* Profil Ihres Partners: ***B***	90	100%
IHR PROFIL: *B*	**Erlös**	**Wahrscheinlichkeit**
	50	25%
	70	25%
Profil Ihres Partners: **A**	110	25%
	130	25%

Gewinnberechnung:

1. Wenn Sie den Standardisierten Beiträge nutzen (Option 1):

$$[\boldsymbol{Gewinn}] = [\boldsymbol{Erl\ddot{o}s}] - \mathbf{50}$$

----

2. Wenn Sie Ihre Beiträge vom Computer ermitteln lassen (Option 2):

$$[\boldsymbol{Gewinn}] = [\boldsymbol{Erl\ddot{o}s}] - [\boldsymbol{Beitrag}] - \boldsymbol{C}$$

Wobei

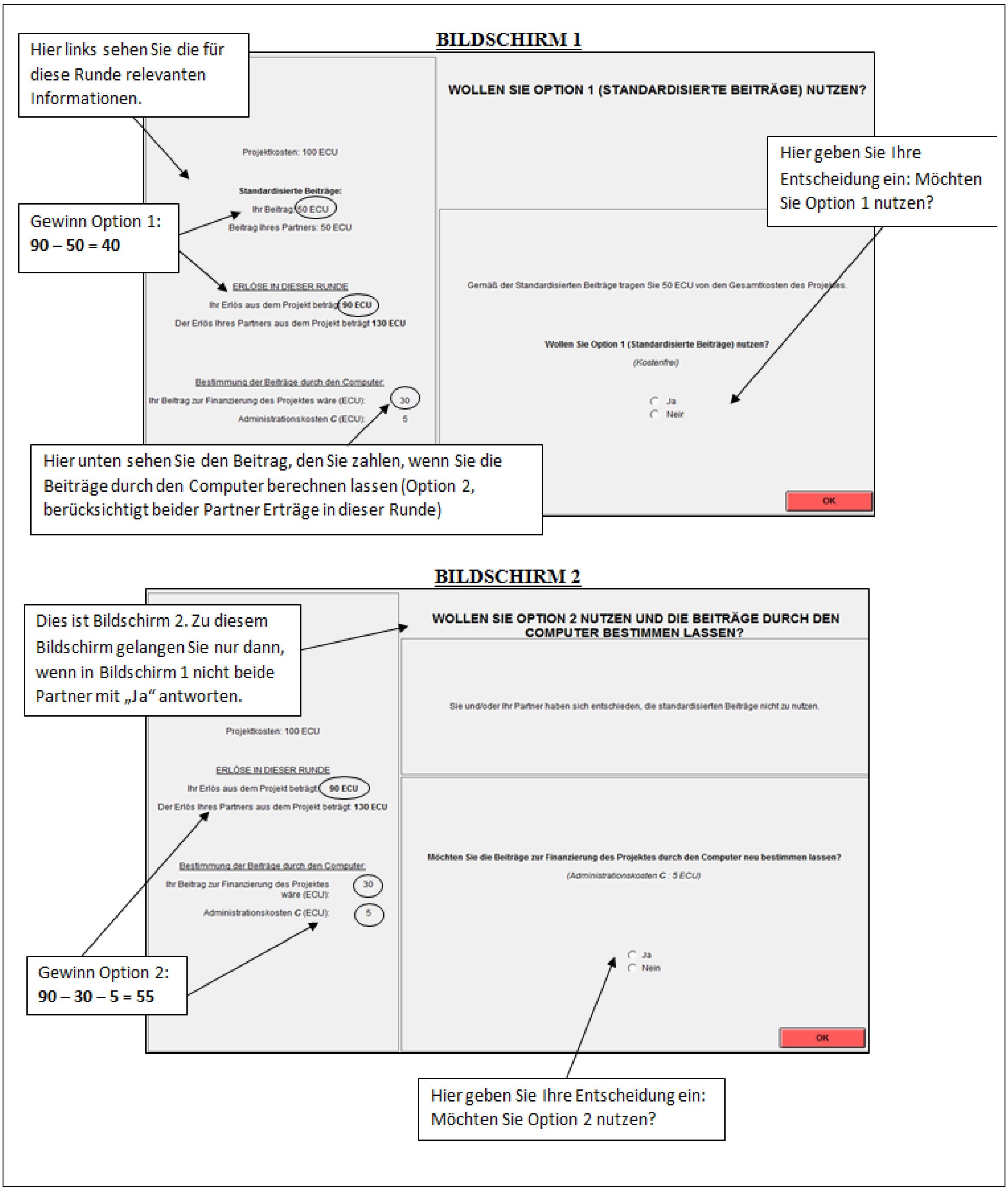
$$[\boldsymbol{Beitrag}] = \mathbf{50} + \frac{[\boldsymbol{Erl\ddot{o}s}] - [\boldsymbol{Erl\ddot{o}s\ Ihres\ Partners}]}{\mathbf{2}}$$

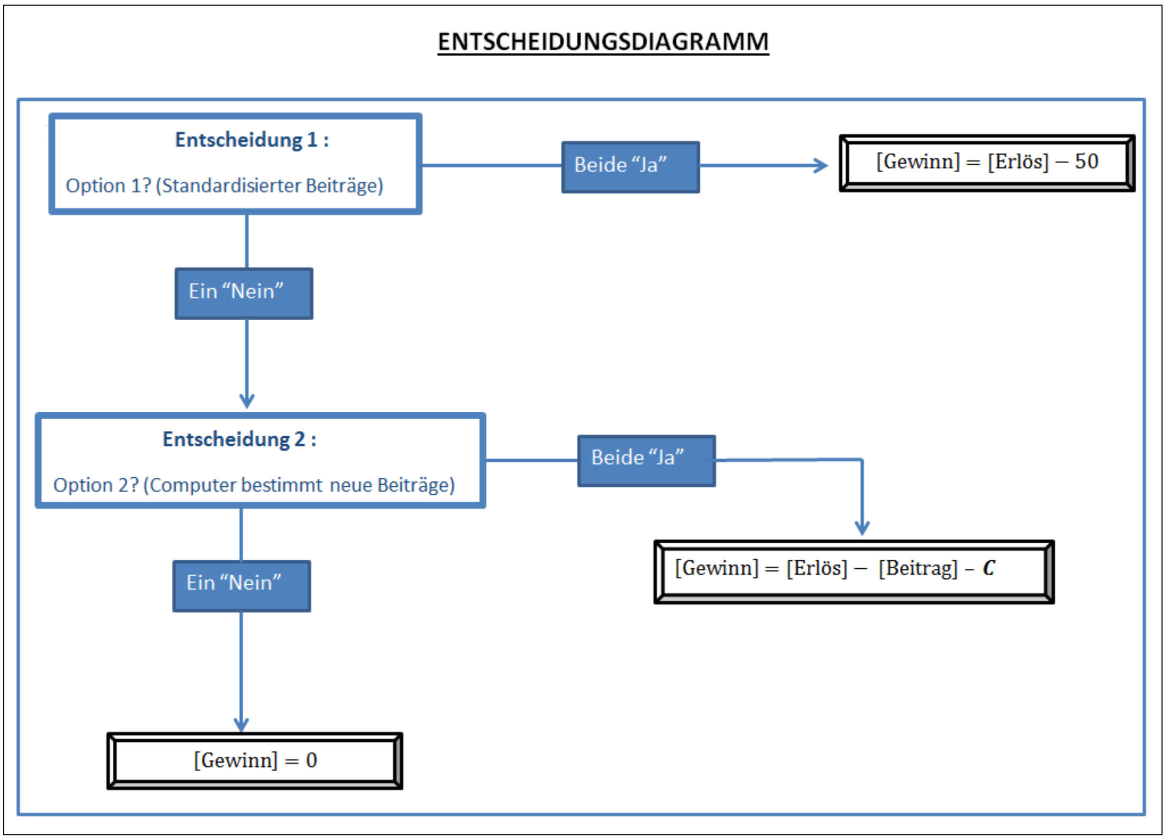
Kosten ***C***	Runde 1-5:	Runde 6-10:	Runde 11-15:
	***C* = 5**	**C = 10**	**C =20**

----

3. Wenn Sie keine der beiden Möglichkeiten nutzen und das Projekt nicht finanzieren:

$$[\boldsymbol{Gewinn}] = \mathbf{0}$$


BILDSCHIRM 1
Hier links sehen Sie die für diese Runde relevanten Informationen.
WOLLEN SIE OPTION 1 (STANDARDISIERTE BEITRÄGE) NUTZEN?
Projektkosten: 100 ECU
Standardisierte Beiträge:
Ihr Beitrag: 50 ECU
Beitrag Ihres Partners: 50 ECU
Hier geben Sie Ihre Entscheidung ein: Möchten Sie Option 1 nutzen?
Gewinn Option 1:
90 – 50 = 40
ERLÖSE IN DIESER RUNDE
Ihr Erlös aus dem Projekt beträgt 90 ECU
Der Erlös Ihres Partners aus dem Projekt beträgt 130 ECU
Gemäß der Standardisierten Beiträge tragen Sie 50 ECU von den Gesamtkosten des Projektes.
Wollen Sie Option 1 (Standardisierte Beiträge) nutzen?
(Kostenfrei)
Bestimmung der Beiträge durch den Computer:
Ihr Beitrag zur Finanzierung des Projektes wäre (ECU): 30
Administrationskosten C (ECU): 5
Ja
Nein
Hier unten sehen Sie den Beitrag, den Sie zahlen, wenn Sie die Beiträge durch den Computer berechnen lassen (Option 2, berücksichtigt beider Partner Erträge in dieser Runde)
OK
BILDSCHIRM 2
Dies ist Bildschirm 2. Zu diesem Bildschirm gelangen Sie nur dann, wenn in Bildschirm 1 nicht beide Partner mit „Ja" antworten.
WOLLEN SIE OPTION 2 NUTZEN UND DIE BEITRÄGE DURCH DEN COMPUTER BESTIMMEN LASSEN?
Sie und/oder Ihr Partner haben sich entschieden, die standardisierten Beiträge nicht zu nutzen.
Projektkosten: 100 ECU
ERLÖSE IN DIESER RUNDE
Ihr Erlös aus dem Projekt beträgt 90 ECU
Der Erlös Ihres Partners aus dem Projekt beträgt 130 ECU
Möchten Sie die Beiträge zur Finanzierung des Projektes durch den Computer neu bestimmen lassen?
(Administrationskosten C : 5 ECU)
Bestimmung der Beiträge durch den Computer:
Ihr Beitrag zur Finanzierung des Projektes wäre (ECU): 30
Administrationskosten C (ECU): 5
Ja
Nein
Gewinn Option 2:
90 – 30 – 5 = 55
OK
Hier geben Sie Ihre Entscheidung ein: Möchten Sie Option 2 nutzen?


ENTSCHEIDUNGSDIAGRAMM
Entscheidung 1 :
Option 1? (Standardisierter Beiträge)
Beide "Ja"
[Gewinn] = [Erlös] – 50
Ein "Nein"
Entscheidung 2 :
Option 2? (Computer bestimmt neue Beiträge)
Beide "Ja"
[Gewinn] = [Erlös] – [Beitrag] – C
Ein "Nein"
[Gewinn] = 0

Chapter 3

Cooperating Tomorrow: Warm Glow vs Cold Prickle revisited

Authors: Arno Apffelstaedt and Arne Pieters

Abstract: For many social dilemma's, there is a delay between the decision and its consequences. Existing experimental studies indicate that such delay can affect social decisions. We study its impact on the impure motives for contributing to a public good, by combining payoff delay with two framing treatments based on Andreoni (1995). We find that, without delay, contribution levels differ significantly between the positive and the negative externality frame. This framing effect disappears when payoffs are delayed by thirty days. Significantly reduced contributions in the negative frame suggest that the delay enables individuals to not feel bad about reducing the earnings of others. We do not observe this in the positive frame, where payoff delay results in a (non-significant) increase in contributions.

JEL: C90, D62, D64, H41,

Keywords: Public Goods, Intertemporal Choice, Impure Altruism, Cooperation

3.1 Introduction.

Economists have formally studied prosocial behavior almost exclusively in situations where decisions have immediate consequences. This is particularly true for lab experiments that study social preferences,[1] in which subjects typically receive decision-contingent payments immediately after the experiment.

In the real world, however, situations abound in which the consequences (cost and benefits) of a social dilemma decision are realized at a future point in time. Examples of situations that entail *outcome delay* range from high stake decisions—enrollment in organ donation programs, or negotiations regarding future climate change policies—to more mundane choices, such as whether to commit to writing a referee report in a few months, or to agree on a future charitable donation.

While there is a rich literature on time-inconsistent preferences when consequences only concern the decision maker herself,[2] little is known so far about how outcome delay affects choices in the social domain. The studies that do exist find mixed results: In a field experiment on charitable giving, Breman (2011) finds that donations are significantly *higher* when donors are asked to commit to future instead of immediate donations. Andreoni and Serra-Garcia (2017) confirm this finding in a lab experiment. At the same time, experiments by Kovarik (2009) and Dreber et al. (2016) find that giving in dictator games (to another participant) significantly *decreases* when outcomes are delayed. Finally, in a public good setting, Koelle and Lauer (2018) do not observe any difference between choices with future vs. immediate consequences.

A standard model of intertemporal choice—where individuals have stationary preferences, purely over final payoffs—would not predict outcome delay to have any impact on giving decisions. In such a model, all outcomes that arise at a future point in time are simply discounted by some factor δ, which does not alter the relative attractiveness of different options. Why might one nevertheless expect prosocial behavior to be affected by delay?

Patterns in prosocial behavior, observed in the field, have brought about a belief among economists that social choices are shaped by more than just preferences over payoffs. People might feel enticed to behave prosocially not only because they care about others, but, for instance, also because the act of giving generates a private emotional benefit—a "warm glow" (Andreoni, 1990).[3] Moreover, in many instances, people may behave pro-socially to avoid feelings of guilt or because they feel "social pressure" to do so (DellaVigna et al., 2012; Krupka and Weber, 2013; Name-Correa and Yildirim, 2016; Andreoni and Serra-

[1]See Vesterlund (2016) for a recent overview.

[2]See Frederick et al. (2002) for an overview. Empirical studies in monetary and other domains include Augenblick et al. (2015), Ashraf et al. (2006) and DellaVigna and Malmendier (2006).

[3]Studies that provide experimental evidence for warm-glow include Andreoni (1993); Palfrey and Prisbrey (1997); Crumpler and Grossman (2008); Ottoni-Wilhelm et al. (2017).

Garcia, 2017).[4] Because utility in these and related domains is associated more with the pro-social action itself than with its consequences for payoffs, "impure" utility elements, such as warm glow or social pressure, may not be discounted in the same way as payoff-related elements. For example, if you commit, today, to a charitable donation that will be transfered three months from now, you might feel good about that commitment immediately, even though the material costs and benefits are all in the future. If impure utility is discounted differently than payoff-related utility, outcome delay can change the observed levels of altruistic or cooperative behavior.

In this paper, we report the results of a public goods experiment designed to shed light on how impure giving motives interact with delay of outcomes. Our design builds on a well-known framing effect established by Andreoni (1995). In the standard case of immediate payoff consequences, contributions in a public good game differ significantly depending on how the game is framed: Either the player generates a *positive externality* on others' payoffs by buying into the *public* good, or she generates a *negative externality* on others' payoffs by buying into the *private* good. Because the payoff-consequences are identical under both frames, observed differences in contributions must be attributed to payoff-irrelevant (i.e., "impure") parts of the utility function. Andreoni observes a greater willingness to cooperate when the externality is positive, and offers the interpretation that the warm glow is stronger than the cold prickle: "people enjoy doing a good deed more than they enjoy not doing a bad deed".

The finding that the framing of a decision can affect contributions,[5] tells us two things: (i) impure utility must affect the decision in at least one of the frames, and (ii) one can use different frames to exogenously produce variation in the importance of this impure utility. The novelty of our design lies in the fact that we exploit this variation to study how impure altruism is affected by outcome delay, by using a between subject, 2×2 factorial design. Similar to Andreoni (1995), we use a positive and a negative frame. In addition to having two groups of subjects play the game and receive their earnings immediately, we let two new groups of subjects play the same game, but their earnings are only paid out 30 days later.

Our study tries to shed light on the role of impure utility when outcomes are delayed. If it takes the form of an intrinsic *reward* for making a cooperative decision, it may be reaped immediately upon deciding, even when payoffs materialize later. If so, the impure motive becomes more important relative to the now discounted payoff utility, and we should see an *increase* in contributions. A logical consequence would be that the framing effect on contributions, which is driven by impure utility, becomes stronger. An alternative interpretation of the impure motive is that it functions more like an intrinsic *penalty* for

[4]Here we use the term "social pressure" as an umbrella term to capture several related concepts including "social norms" (Krupka and Weber, 2013) and "social image" (Ariely et al., 2009).

[5]Later studies that confirm this finding include Fujimoto and Park (2010), Khadjavi and Lange (2015), and Dufwenberg et al. (2011)

behaving selfishly. A participant may then contribute because she feels pressured to do so, or because she tries to avoid a feeling of guilt. If this is true, it is conceivable that delaying the consequences allows her to take more emotional distance from them, reducing the importance of the impure motive. The payoff delay should then result in a *decrease* in contributions, while also reducing the framing effect.

In our experiment, participants made one decision on how much to contribute in a standard, one-shot public good game. Without delay, the two frames produced different contribution levels: our participants were significantly more cooperative in the negative externality frame than they were in the positive frame. This difference disappeared when the payout was delayed by 30 days: we observe a large, statistically significant drop in contributions under the negative frame, and a small (not significant) increase in contributions under the positive frame. Combining all treatments, we find that the delay significantly reduced the framing difference.

For the negative frame, these findings contradict the theory that prosocial decisions generate a warm glow even before the consequences materialize. Instead, they support the idea that people feel a pressure to contribute that is alleviated by the outcome delay. For the positive frame, the immediate glow theory is not contradicted, nor is it strongly confirmed. Comparing the frames, they interact quite differently with the delay treatment. It seems that, in addition to the importance of impure utility, the frames also determine the way in which this utility is discounted relative to payoff utility. This would suggest that the impure motives associated with the two frames, say warm glow and cold prickle, differ in more respects than just strength, and may even impact individuals' utility in opposite directions.

The convergence of contribution levels between frames, in the treatments with delayed payoffs, accomodates an additional interpretation of our findings. Pyschologists and behavioral economists often associate time inconsistent choices with a self-control problem. Various theories approach this as an interaction between two selves or systems within the individual's decision making process. Fudenberg and Levine (2006) model it as a game between the long-run self and the short-run self, where the latter is indifferent when it comes to decisions that affect only future outcomes. Loewenstein et al. (2015) describe how deliberative and affective processes interact in shaping behavior. Environmental stimuli, such as decision framing, can influence these processes differentially, and temporal proximity is of great importance for the affective motivation.[6]

Such theories do not make strong predictions on the direction of a framing effect or of the delay effect for a given frame, but they are consistent with the convergence of contributions we observe in our data, as emotional triggers become less important relative to rational considerations. From a methodological perspective, this observation suggests that we should be careful not to overinterpret the effects of decision framing found in static

[6]Similarly, Rogers and Bazerman (2008) describe the competing "should" self and "want" self.

lab experiments. For the extrapolation of such behavioral findings to real-world settings, our results indicate that it can be important to take into consideration the relative timing of decisions and their consequences.

In the next section, we use a simple choice model to guide our theoretical discussion on the effect of delaying payoffs. Sections 3.3 and 3.4 present the design and the results of our experiment. Section 3.5 provides a more detailed discussion on the main findings, on some additional insights in gender differences, and considers why the direction of the framing effect may be opposite to the original finding by Andreoni (1995). Section 3.6 concludes.

3.2 Theoretical Considerations

Consider an individual's decision on whether and how much to contribute to a pro-social cause (a public good, charity, etc.), seeking to maximize utility $U = U(y_t, Y_t, g)$. Let $g \geq 0$ denote the individual's contribution, y the individual's payoff, and Y the set of payoffs for all individuals that benefit from contribution g.[7] To consider the effect of positive vs. negative externality framing on the choice of g, we impose some structure on the utility function. When making a decision that has immediate payoff consequences, $t_{Payoffs} = N$ (N for "Now"), we assume that utility takes the following form:

$$U_N(y, Y, g) = u(y, Y) + v(g \,|\, \mathit{framing}). \tag{3.1}$$

The first part of utility captures concerns that relate to the payoff-consequences: The individual cares about her own payoff, and, possibly, about the payoffs of others, for example due to a concern for efficiency, aversion to inequality, or pure altruism.[8] The second part, which we allow to depend on the framing, contains what we will refer to as "impure" motives to give: the element $v(g \,|\, \mathit{framing})$ captures the individual's utility that is not related to payoffs, but to the size of her own contribution *per-se*. This may reflect, for instance, a "warm-glow" that the individual experiences when giving (see, e.g., Andreoni, 1989, 1990, 2006), or a "social pressure" to behave pro-socially (see, e.g., Ariely et al., 2009; DellaVigna et al., 2012; Krupka and Weber, 2013; Name-Correa and Yildirim, 2016; Andreoni and Serra-Garcia, 2017).[9] We assume throughout that $v'(g) > 0$ and $v'(g) \leq 0$ in both frames.[10]

[7]Following a standard public good game structure, contributions are privately costly, $\frac{\partial y}{\partial g} < 0$, but generate a positive externality for others, $\frac{\partial Y}{\partial g} > 0$.

[8]Standard references include Fehr and Schmidt (1999); Bolton and Ockenfels (2000); Engelmann and Strobel (2004).

[9]Here we use the term "social pressure" as an umbrella term to capture several related concepts including "social norms" Krupka and Weber (2013) and "social image" Ariely et al. (2009).

[10]Depending on whether one models "warm-glow" or "social pressure", the existence of $v(g)$ may be assumed to make the individual better-off ($v(g) > 0$, á la Andreoni, 1990) or worse-off ($v(g) < 0$, á la DellaVigna et al., 2012). In the former case, $v'(g) > 0$ would say that higher contributions increase the warm-glow, while in the latter, that higher contributions decrease social pressure.

To study how impure motives interact with outcome delay, we vary two aspects of the decision environment: (1) The timing of payoffs, $t_{\mathit{Payoffs}} \in \{N, F\}$ (F for "Future") and (2) The "framing" of the decision, $\mathit{framing} \in \{+, -\}$. In particular, we consider a variation in externality framing suggested by Andreoni (1995): Frame "+" highlights the positive external effect of contributing (increasing g); Frame "–" highlights the negative external effect of withholding contributions (decreasing g). Taken together, we study four scenarios (see Table 3.1). For each scenario, we study the optimal choice of g, assuming an interior solution $g^* \geq 0$, conditional on other determinants of y and Y being fixed.[11]

Table 3.1: Frames and Timings

		Frame	
		Pos	Neg
Timing	Now	g_N^+	g_N^-
	Fut	g_F^+	g_F^-

Immediate Payoffs. From Equation (3.1), when payoffs are immediate ($t_{\mathit{Payoffs}} = N$), the optimal individual contribution is governed by the first-order condition

$$\frac{\partial u(y, Y)}{\partial g} + v'(g \,|\, \mathit{framing}) = 0. \tag{3.2}$$

Because marginal payoff consequences ($\frac{\partial u(y_t, Y_t)}{\partial g}$) are identical under both frames, framing differences in behavior must be attributed to the impure part of the utility function. In particular, $g_N^+ \neq g_N^-$ if and only if $v'(g \,|\, +) \neq v'(g \,|\, -)$.[12]

Delayed Payoffs. When payoffs are delayed ($t_{\mathit{Payoffs}} = F$), payoff-dependent utility $u(y, Y)$ arrives at a later point in time. According to standard models of intertemporal choice, at the time of deciding on g, $u(y, Y)$ is then discounted with some factor $\delta_u \in [0, 1]$. The impure part of the utility function may be discounted differently, as it is less clear *when* it is experienced. Let $\delta_v \in [0, 1]$ be the discount factor on $v(g)$. Then

$$U_F(y, Y, g) = \delta_u \cdot u(y, Y) + \delta_v \cdot v(g \,|\, \mathit{framing}). \tag{3.3}$$

[11] That is, for our analysis, both y and Y are assumed to vary only in g (e.g. fixing contributions by others). We concentrate on unique solutions $g^* \geq 0$ defined by the first-order condition $\partial U/\partial g = 0$ by assuming throughout that at $g = 0$, $\partial U/\partial g \geq 0$, and for any $g \geq 0$, $\partial^2 U/\partial g^2 < 0$.

[12] This argument has already been made by Andreoni (1995) himself. Similar arguments have also made by the literature on "social pressure". Krupka and Weber (2013), for instance, argue that the difference in contributions between dictator games framed as a "give" decision and those framed as a "take" decision can be attributed to a change in what people consider to be the social norm.

The optimal contribution of the individual in the case of delayed payoffs is governed by the first-order condition

$$\delta_u \cdot \frac{\partial u(y,Y)}{\partial g} + \delta_v \cdot v'(g \mid \mathit{framing}) = 0. \tag{3.4}$$

Comparing first-order conditions (3.2) and (3.4), it is clear that if impure utility is not discounted in the same way as payoff utility ($\delta_v \neq \delta_u$), delaying payoffs will change the optimal contribution:

1. $\boldsymbol{\delta_u < \delta_v}$: If impure utility is not discounted, or discounted less than payoff-related utility, delay should *increase* contributions under both frames: $g_F^+ > g_N^+$ and $g_F^- > g_N^-$. As the frame-dependent part of utility receives more weight, any initial differences in contributions between frames should increase when payoffs are delayed: $|g_F^+ - g_F^-| > |g_N^+ - g_N^-|$.

2. $\boldsymbol{\delta_u > \delta_v}$: If impure utility is discounted more than payoff-related utility, delay should *decrease* contributions under both frames: $g_F^+ < g_N^+$ and $g_F^- < g_N^-$. As the frame-dependent part of utility receives less weight, differences in contributions between frames should decrease when payoffs are delayed: $|g_F^+ - g_F^-| < |g_N^+ - g_N^-|$.

The first case captures the idea that, since $v(g)$ is related to *deciding* how much to contribute rather than to the consequences of this decision, the individual experiences this warm glow *upon deciding* to contribute.[13] If so, it arguably arrives immediately and is not discounted ($\delta_v = 1$). This leads to an increase in the marginal utility of contribution relative to its cost, and thus to a higher optimal contribution in the case of $t_{Payoffs} = F$. Accordingly, we would expect the effect of delaying payoffs to be greatest in the frame that induces the greater impure motivation, amplifying the framing effect on contributions.

The second case reflects the opposing idea, that payoff delay allows the individual to distance herself from the impure utility $v(g)$ that shapes her behavior when the decision has immediate payoff-consequences. This may be a reasonable theory if $v(g)$ is better described as a pressure to be unselfish than as an intrinsic reward such as warm-glow—in particular when this pressure enters the utility function as a negative element ($v(g) < 0$). Theories in psychology suggest that temporal distance to an event can reduce the relative importance of negative feelings associated with that event (Trope and Liberman, 2000). This would be captured by $\delta_v < \delta_u$, implying that impure motives become relatively less important with delay, making the individual behave more selfishly. At the same time, differences between frames should become less pronounced. In the limit ($\delta_v = 0$), where contributions are entirely shaped by standard, payoff-dependent utility $u(y_t, Y_t)$, frames should have no impact on behavior.

[13] This idea was mentioned by Andreoni and Payne (2003) and is supported by the findings of Breman (2011).

3.3 Experimental Design

Participants in our experiments play a one-shot public good game in groups of four. We use the structure of a linear public good game with voluntary contributions, following the design of Andreoni (1995) for the stage game.[14] Subjects are endowed with 60 tokens, which they can freely distribute between a private good x_i and a public good g_i, labelled 'Individual Box' and 'Group Box', respectively. In terms of payoffs, the private good gives a higher individual return than the public good, but a lower total group return. Total group payoff is therefore maximized when the four players contribute all their tokens to the public good, but the unique best response for a selfish payoff-maxmizing individual is to not contribute at all.

3.3.1 Treatments: Framing and Timing

The experimental design consists of 2×2 treatments. Each treatment uses either a positive externality frame (+) or a negative externality frame (−). In the positive frame, every token in x_i earns subject i two coins, while every token in g_i earns one coin for i and one coin for each of the other three group members:

$$y_i^+ = 2x_i + g_i \quad + \quad \sum_{j \neq i} g_j \qquad \text{s.t.} \quad x_i + g_i = 60 \tag{3.5}$$

In the negative frame, a token to x_i generates a negative externality of -1 to the other three group members, while a token to g_i has no external effect. Likewise, your payoff is reduced by one coin for every token fellow group members put in their Individual Box. To keep the payoff space identical to the positive frame, every group member receives an automatic payment of 180 coins:

$$y_i^- = 2x_i + g_i \quad + \quad 180 - \sum_{j \neq i} x_j \qquad \text{s.t.} \quad x_i + g_i = 60. \tag{3.6}$$

Note that the allocation decision $\{x_i, g_i\}$ can be reduced to a single choice g_i, using that $x_i = 60 - g_i$. The payoff spaces are identical between frames: Any four token allocations will produce the same set of four final payoffs in both frames.

Appendix (3.7.2) contains screenshots of the instructions provided to participants. A key element, that differed between frames, was a summary of how a token in either the 'Individual Box' or the 'Group Box' affected incomes. These are reproduced in Tables 3.2 and 3.3.

[14] We chose a non-repeated structure to avoid certain strategic elements affecting the contribution decision. In a repeated setting, players may take into account previous actions by fellow group members, or try to affect their future actions.

Table 3.2: Positive Frame.

	INDIVIDUAL BOX		GROUP BOX	
	Your Income	Income of the Other Group Members (each)	Your Income	Income of the Other Group Members (each)
Income per Token	+ 2 Coins	0 Coins	+ 1 Coin	+ 1 Coin

Table 3.3: Negative Frame.

	INDIVIDUAL BOX		GROUP BOX	
	Your Income	Income of the Other Group Members (each)	Your Income	Income of the Other Group Members (each)
Income per Token	+ 2 Coins	- 1 Coin	+ 1 Coin	0 Coins

These frames are combined with two timing treatments, labelled *NOW* and *FUT*, referring to when participants receive the money they earn in the experiment. For the *NOW* sessions, all earnings are paid out within a day after the experiment, while for the *FUT* sessions the payment is made 30 days later. Importantly, only the time of this transaction differs between treatments: The decisions participants make in the experiment do not differ, and both in *NOW* and *FUT* they cannot be reversed at a later point. In all treatments within this 2×2 design, the outcome of the game is communicated to participants within one day of completing the experiment.

3.3.2 Procedures

All experimental sessions were conducted online, in May and October of 2018.[15] The experiment was programmed using Limesurvey,[16] and we used the online research platform Prolific for recruitment and payments.[17] The participants received £0.01 per earned 'coin'. We restrict attention to participants with English as their first language, and with a high "Prolific Rating".[18] In each session, participants went through the experiment independently, and it took them around 10 minutes from start to finish. This includes explanation of the task, control questions, the decision task itself and some post-experimental (unincentivized) questions about beliefs and personality traits. Importantly, the instructions and control questions about the payoff structure were accompanied by a simulator, in which

[15] We conducted two rounds of treatments (8 sessions) in May and one round of treatments (4 sessions) in October.

[16] Limesurvey GmbH. / LimeSurvey: An Open Source survey tool /LimeSurvey GmbH, Hamburg, Germany. URL http://www.limesurvey.org

[17] https://prolific.ac/

[18] This rules out participants who have a large number of rejections from previous studies, implying they tried to obtain payment even though they did not complete the study.

the participants could freely explore how their own contributions and those of others would determine everyone's payoffs.

Subjects earned a fixed amount of GBP 1.00 for participating, and up to GBP 3.00 extra payment in the public good game. Each participant was randomly and anonymously assigned to a group of four. The day after the sessions took place, all participants received a message reporting their own contribution decision, the total group allocation of tokens, and their individual earnings, with a reminder of when they were going to be transfered to their Prolific account. Subjects were never able to identify the decisions made by an individual fellow group member.

3.4 Results

The data summarizing the contribution decisions in each of the four treatments are presented in Table 3.4, along with some demographic information about the participants. In what follows, we first present our main results by analyzing how contributions are affected by the different treatments. We then report some additional insights from examining beliefs and gender differences. Throughout this section, we report findings on public good contribution g, by which we mean tokens put in the 'Group Box'. In most cases this is expressed as the share of the available 60 tokens that were contributed.

Table 3.4: Summary statistics

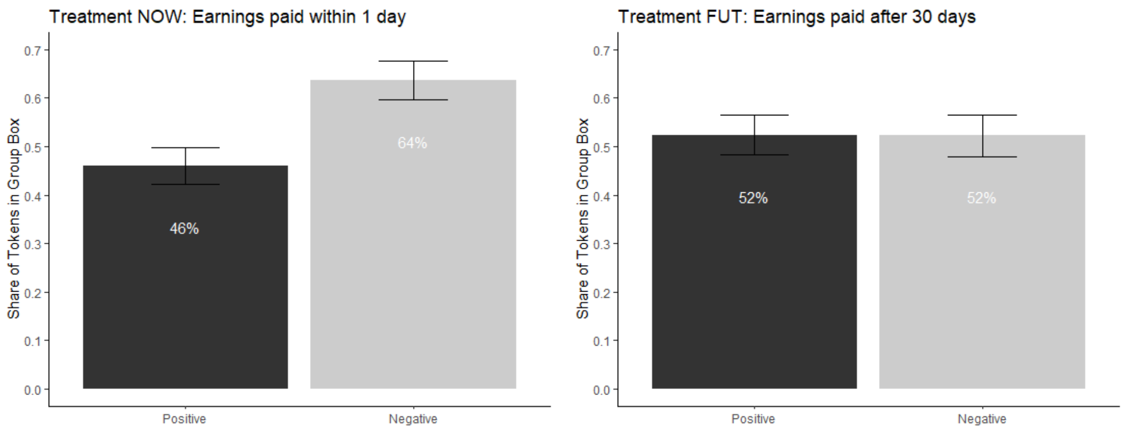
	Contributions			Demographics			
Treatment	Mean g	$g > 0$	$g \mid g > 0$	n	Female	Student	Unempl.
$NOW(+)$	27.60	79.3%	34.82	82	51%	26%	9%
	(20.92)		(17.29)				
$NOW(-)$	38.18	86.6%	44.10	82	54%	21%	12%
	(21.79)		(16.88)				
$FUT(+)$	31.44	82.5%	38.11	80	46%	15%	8%
	(22.45)		(18.83)				
$FUT(-)$	31.34	80.7%	38.82	83	53%	18%	11%
	(23.40)		(19.63)				

Note: The average number of tokens contributed, the share of contributors (at least 1 token to the Group Box), and the average among contributors. Standard errors in parentheses. The last four columns show demographics of participants per treatment.

3.4.1 Main results: Differences in contributions

Across all four treatments, the average contribution is 32.14 tokens, which amounts to 53.6% of the total token endowment. In discussing observed treatment effects, we will first consider the two treatments without delay, and then look at the effect of delaying the payment of earnings by 30 days.

Figure 3.1: Contributions per treatment



Note: Average share of tokens contributed to the public good when payout was within 1 day (left panel) and when payout was after 30 days (right panel). Error bars denote ±1 standard error.

Result 3.1. *When payoffs are not delayed (NOW), contributions depend on the framing: Contributions are higher in the negative externality frame.*

Without delay, we observe that people contribute significantly less in the positive frame than they do in the negative frame (p=.002, Mann-Whitney U). Figure 3.1 shows the mean contribution for each treatment. On average, participants contribute 46.0 % of their tokens in the $NOW(+)$ treatment, compared to 63.6% in the $NOW(-)$ treatment (+17.6 percentage points). The framing effect does not change when controlling for standard demographic characteristics, see the OLS estimations (first two columns) in Table 3.5. We observe the same difference between $NOW(+)$ and $NOW(-)$ across all sessions.

Result 3.2. *When payoffs are delayed (FUT), there is no framing effect: Contributions are identical in both frames.*

In the *FUT* treatments, average contributions are (almost) identical in both frames, see the right panel of Figure 3.1 (p=.991, Mann-Whitney U). Participants in the positive externality treatments now contribute 52.4% of their tokens, while those in the negative externality treatments contribute 52.2%. OLS estimations with and without controls (third and fourth column of Table 3.5) confirm this result: pooling the data of all four treatments, we find a significant negative effect on Neg. Frame × FUT (p=.030), which provides statistical evidence that delaying payoffs reduces the framing effect, see columns (5) and (6) of Table 3.5.

Figure 3.2: Distribution of contributions per treatment

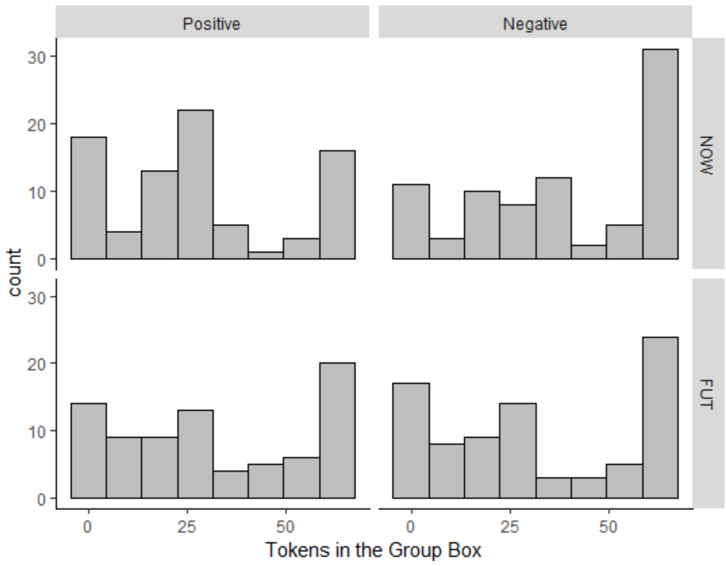


Figure 3.2 presents the entire distribution of tokens put in the Group Box for each treatment. We see that not just average contributions, but contributions in general are highly similar in the *FUT* treatments (p=.994, Kolmogorov-Smirnov) and highly significantly different in the *NOW* treatments (p=.001, Kolmogorov-Smirnov).

Result 3.3. *The effect of delay on contributions differs between frames. In the positive frame (+), delaying payoffs leads to a (non-significant) increase in contributions. In the negative frame (−), delaying payoffs leads to a decrease in contributions.*

If we examine the effect of delay on contributions for each frame separately, we see that in the positive frame (+) the delay leads to an *increase* in the average contribution from 46.0 % to 52.4%. This increase is not statistically significant (p=.481, Mann Whitney U). In the negative frame (−), on the other hand, delay leads to significant *decrease* in the average contribution from 63.6% to 51.2% (p=.040, Mann Whitney U). It follows that the effect of delay on contributions is significantly different across the two frames (as demonstrated by the significant coefficient on Neg. Frame × FUT in Table 3.5).

Table 3.5: OLS Estimations of effects on g.

	NOW		FUT		All		
Expl. Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Neg. Frame	.176***	.180***	-.002	.005	.176***	.174***	.090**
	(.056)	(.054)	(.060)	(.060)	(.058)	(.052)	(.044)
FUT					.064	.048	.047
					(.058)	(.058)	(.044)
Neg. Frame × FUT					-.178**	-.165**	-.134**
					(.082)	(.081)	(.062)
Belief Others' Contr.							.780***
							(.049)
Add. Controls	No	Yes	No	Yes	No	Yes	No
Constant	.460***	.570***	.524***	.555***	.460***	.539***	.170***
	(.039)	(.051)	(.043)	(.053)	(.041)	(.048)	(.036)
Observations	164	164	163	163	327	327	325
FUT + [N.Fr × FUT]					-.114**	-.118**	-.083*
					(.058)	(.057)	(.054)

Note: Dependent variable is the share of tokens contributed to Group Box ($g_i/60$). Additional control variables: Gender, student and unemployment status. In certain specifications the coefficients indicate that females and students contribute significantly less. This will be further discussed in Section 3.5. Standard errors in parentheses. *** p<.01, ** p<.05, * p<.1

3.4.2 Additional observations

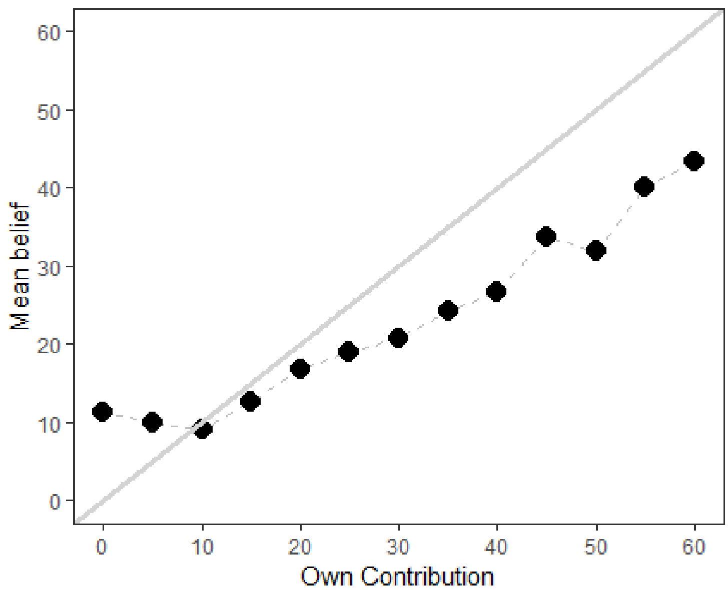
3.4.2.1 Beliefs.

After deciding how many tokens to contribute, subjects were asked to estimate the average contribution by their three fellow group members. It is important to point out that this question was not incentivized and that the stated belief is most likely not exogenous to the own contribution. One should therefore be careful not to overinterpret correlation. Nevertheless, it is interesting to see whether and to what extent these reported beliefs move along with the average contributions from one treatment to the next.

A large majority of participants report that they believe their fellow group members to contribute less tokens to the Group Box than they do themselves. The average report is 8.8, 15.4, 13.4 and 7.9 percentage points lower than the participants own contribution in treatment $NOW(+)$, $NOW(-)$, $FUT(+)$, and $FUT(-)$, respectively. Only 15.8% of all subjects believe their contribution to be below the group average, 35.4% estimate the two to be equal, and 48.8% say they contribute more than the average. Figure 3.3 shows how the average belief varies with the own contribution across all treatments. The correlation coefficient between the own contribution and the reported belief is .80.

Figure 3.4 shows the average reported belief for each treatment. While we find a similar pattern in beliefs as we do in contributions (Figure 3.1), the effects of payoff delay are much less pronounced on beliefs as they are on actual contributions. Statistically,

Figure 3.3: Beliefs vs. Own Contribution



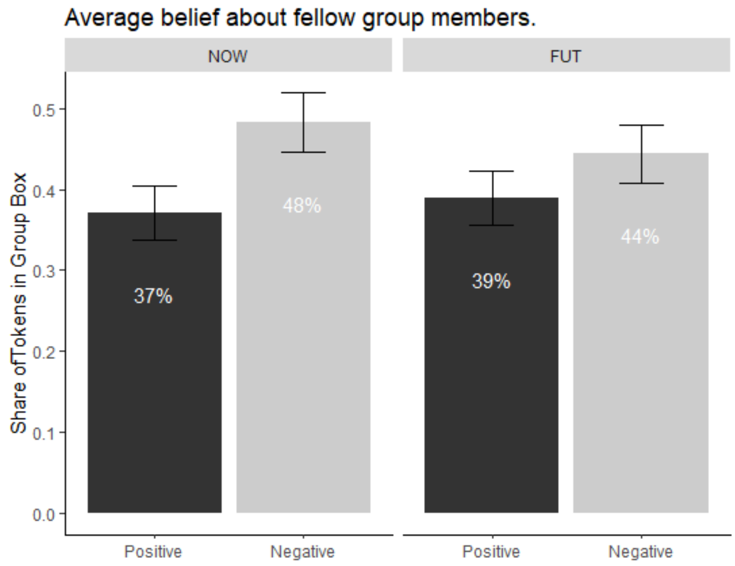
Note: Average share of tokens subjects believe their fellow group members contributed, grouped by the own contribution, rounded to the nearest multiple of 5.

when controlling for the own contribution, we do not find any significant effect of delay on beliefs. Including reported beliefs as an explanatory variable in the OLS regressions for the own contribution (see Table 3.5, column (7)) slightly reduces the coefficient on Neg.Frame and Neg.Frame × FUT, but does not affect its significance much. In summary, while we find that participants themselves react significantly to delay, we do not find evidence for a direct effect on what they believe others contribute. As far as we can tell, the importance of beliefs about the choices of others (or related concepts such as conditional cooperation), in explaining the observed delay effects, seems to be minor.

3.4.2.2 Gender differences.

We find that women contribute significantly less than men (p=.051, Mann-Whitney U). For all treatments combined, women put 49.5% of their tokens in the Group Box, while men contribute 57.8% of their tokens. Figures 3.5 and 3.6 show the averages per gender, comparing all four treatments. In the *NOW* treatments, women contribute 46.7%, which is significantly less than the 63.8% average among men (p=.003 , Mann-Whitney U). Interestingly, the gender effect disappears once payouts are delayed: across the two *FUT* treatments, average contributions are almost identical for both sexes (p=.885 , Mann-Whitney U). Tables 3.6 and 3.7 report OLS estimates that confirm these observations, also when controlling for student and unemployment status: Females contribute less in both

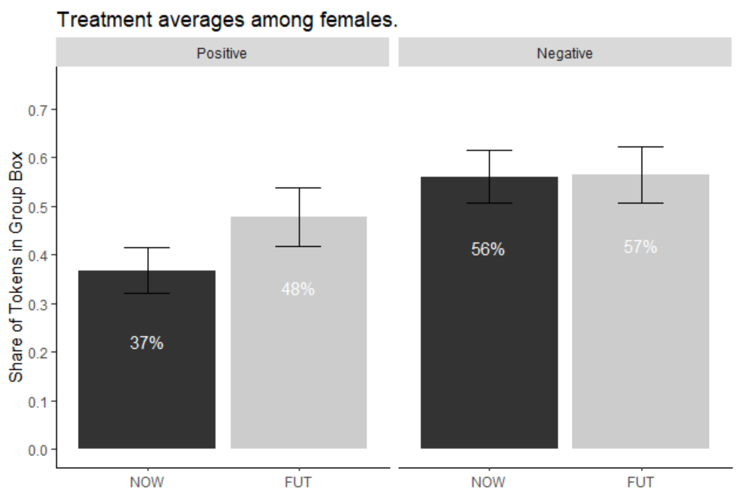
Figure 3.4: Beliefs per treatment



Note: Average share of tokens subjects believe their fellow group members contributed.

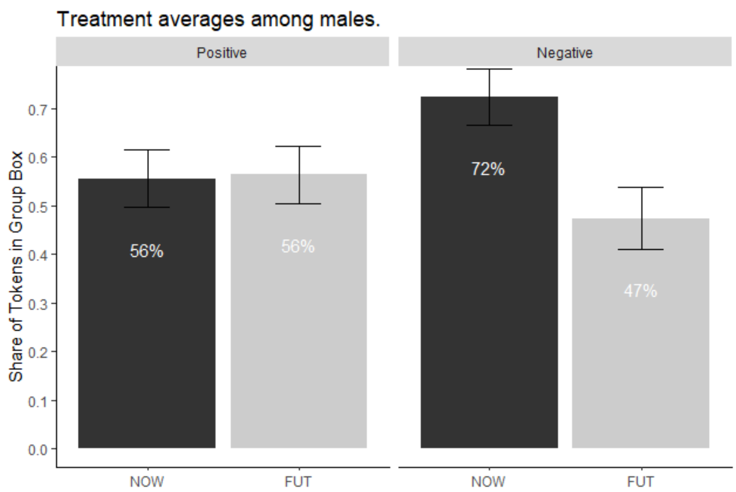
the $NOW(+)$ and the $NOW(-)$ treatment. The framing effect in the NOW treatments (Result 3.1) does not depend on gender, see column (3) of Table 3.7.

Figure 3.5: Contributions - Females



The (significant) positive interaction effect (Female $\times$ FUT) confirms that payoff delay reduces the gender difference. The treatment averages reveal that in each frame, this reduction is caused by only one of the two sexes, combining to achieve the elimination of the overall framing effect that is described by Result 3.2. In the positive frame, women contribute more in the $FUT(+)$ than in the $NOW(+)$ treatment (+10.9 percentage points, p=.245 Mann Whitney U), while for men there is no increase. We observe the opposite

Figure 3.6: Contributions - Males



in the negative frame. Contributions among females are virtually the same in treatments $NOW(-)$ and $FUT(-)$. Male contributions decrease significantly (p=.009, Mann-Whitney U), however, from 72.4% in treatment $NOW(-)$ to 47.7% in the $FUT(-)$ treatment (-14.5 percentage points).[19] It is also worth pointing out that the average male contributions in the negative frame treatments are very close to the average female contributions in the positive frame treatments, all around 56%.

3.5 Discussion

3.5.1 Main treatment effects.

We begin by discussing our main findings in light of the theoretical considerations laid out in section 3.2.

The strong and significant framing effect we find on contributions in the *NOW* treatments (see Figure 3.1, left panel) implies that—in at least one of the frames—impure motives play an important role in driving behavior when payoffs are immediate. More specifically, the difference in average contributions (17.6 percentage points) between the *NOW*(+) and the *NOW*(–) treatments can be attributed to a difference in the impure utility element $v(g\,|\,frame)$, suggesting that when payoffs are immediate, impure motives to contribute (such as warm-glow or social pressure) are stronger in the negative than in the positive frame.

As payoffs are delayed, payoff-related utility $u(y, Y)$ should be discounted ($\delta_u \in [0, 1]$). In Section 3.2, we discussed the implications of two opposing ideas in relation to the dis-

[19] Similar to their own contribution choices, the stated beliefs about contributions by fellow group members differ significantly between men and women in the *NOW* treatments (p=.001, Mann-Whitney U), but not in the *FUT* treatments.

counting of impure utility (δ_v). First, an increase in observed contributions would support the idea that impure motives become *more* important with delayed payoffs ($\delta_v > \delta_u$), for instance because the warm-glow or the social pressure is experienced instantaneously at the time of deciding ($\delta_v = 1$). However, we only observe such an increase in the positive frame, where contributions on average go up by 6.4 percentage points, and this increase is not statistically significant. Moreover, this idea is contradicted by our observation that in the negative frame, delay leads to a significant *decrease* in contributions (− 12.4 percentage points). In this frame, rather than arriving instantaneously, impure utility seems to matter *less* when deciding about a future contribution, implying it is discounted more heavily than payoffs ($\delta_v < \delta_u$). This fits with the second idea, that negative feelings associated with social pressure, or guilt, are easier to dissociate from when consequences lie in the future. While our data provide firm support for this in the negative frame, a more general (i.e., frame-*independent*) pattern for the effects of delay on the importance of impure utility does not clearly emerge.

One way to interpet our findings, specifically Result 3.3, is that the effect of delay on contributions is frame-dependent: the idea of instantaneous impure utility may apply to the positive frame, while the idea that delay enables us to ignore impure motives applies to the negative frame. This could indicate that the impure motives for pro-social behavior differ between frames in more respects than just magnitude or strength. It is plausible that negative feelings (from pressure or guilt) play a role in the negative frame, while positive motivations (such as warm glow) drive decisions in the positive frame. As a consequence, the impure utility elements could also be discounted differently ($\delta_v^+ \neq \delta_v^-$). According to (Trope and Liberman, 2000, p.876), the "effect of temporal distance on preference depends on whether the valence of the outcomes is positive or negative", and "the value of outcomes is generally discounted (diminished) over time delay, but negative outcomes undergo steeper time discounting than do positive features". Much in the spirit of "warm-glow versus cold-prickle" (Andreoni, 1995), people would use payoff delay to distance themselves (psychologically) from the negative externality in the negative frame and to embrace the positive externality in the positive frame. This hypothesis seems to be corroborated to some extent by the finding that delay leads to higher contributions in experiments on charitable giving (where it is likely that the positive externality motive dominates, see for instance, Breman, 2011; Andreoni and Serra-Garcia, 2017)[20] and to lower contributions in dictator game experiments (in which a negative feeling of pressure or guilt might be commanding, see, e.g., Kovarik, 2009; Dreber et al., 2016).

An alternative way to interpret the data is to focus on how the framing effect responds to delay, drawing more directly on Result 3.2. When payoffs are delayed, the framing

[20] Also worth mentioning here is a recent study on fundraising by Kellner et al. (2019), who find that participants donate more when the decision is only implemented with a 50% probability. Reduced probability of implementaion is obviously not the same as a time delay, but it may serve a self-serving mechanism in a similar way.

effect disappears ($g_F^+ = g_F^-$). This suggests that the frame-dependent elements of $v(g)$ are important for decisions with immediate payoff consequences, but not for decisions with delayed consequences. The implication is that impure motives become less important ($\delta_v < \delta_u$) in *both* frames. Again, there exist psychological theories that support such an interpretation of our data. Loewenstein et al. (2015, p.58), for instance, argue that "affective motivations [i.e., impure motivations] are intense when rewards and punishments are immediate but much less intense when they are temporally remote. Deliberation [i.e., the "pure" part of utility] is, in contrast, much less sensitive to immediacy." This would suggest a dual-self type of story where frames play a role for the "affective system" that makes decisions for the now, but are unimportant for the "deliberative system" making decisions for the future. As observed in our experiment, decisions taken in the *FUT* treatment would then be more 'stable' (or 'rational') than decisions taken in the *NOW* treatments.

Compared to our predictions in Section 3.2, the data fits better to the idea of frame-dependent discounting ($\delta_v^+ \neq \delta_v^-$). However, it would have been more strongly confirmed by a statistically significant positive effect of delay on contributions in the positive frame. Theories based on competing internal processes do not directly produce strong, testable, predictions (e.g., on the direction of the framing effect or the delay effect), but they do predict the reduction in framing effect we observed. While our data do not allow us to conclusively confirm either of these two theories, they *do* clearly demonstrate that the framing of the decision environment interacts with delay. Since, by design, the framing can only affect the impure part of the utility function, this implies that the relative weighting of impure motives is part of the explanation for why contribution decisions can change as consequences are delayed.

3.5.2 Gender differences.

We see in the *NOW* treatments that while both male and female contribution levels are affected by the frame, women contribute significantly less than men in both frames. When payoffs are delayed (*FUT*), gender differences disappear along with framing effects. Average contributions go up among women in the positive frame, while they go down among men in the negative frame. This suggests that the impure motivations play a role in generating the initial gender differences. A recent study by Klinowski (2018) comes to a similar conclusion. In a dictator game, he finds that women initially give more than men, but they also are more likely to retract their gift when given the opportunity to do so without detection, eliminating the gender difference in the final amount transfered. It is also consistent with the finding by Huang and Wang (2010) that frames can affect men and women differently, depending on the task domain, specifically that men show greater response to negative framing of decisions in the monetary domain.

Croson and Gneezy (2009) offer a somewhat similar interpretation of the gender differences in social preferences, but they attribute the inconsistency in experimental findings to an "increased sensitivity of women to the context of the situation" (p.461). Our findings rather suggest that *both* genders are sensitive to context, but that they can respond differently to delay. Andreoni and Vesterlund (2001) study gender differences specifically in relation to altruism, and conclude that while men's demand for altruism is highly responsive to its price, women display a more inelastic level of altruism. Accordingly, men are more likely than women to be either extremely altruistic or extremely selfish.[21]

Comparing decisions with delayed vs. immediate consequences, Andreoni and Serra-Garcia (2017) find that observed time inconsistency in giving decisions are mainly driven by women. We observe a similar pattern in the positive frame, but without any statistically significant differences. Breman (2011), in contrast, finds that men are more responsive to delay: They increase there monthly charity donation by a significantly larger amount than women do, when the change takes effect one month later vs. immediately. In our study, men are significantly more responsive to delay in the negative frame, which is arguably the frame less comparable to a charity donation experiment.

3.5.3 Direction of framing effect.

As explained in the motivation for using the two externality frames, our aim was to produce a difference that could not be explained by pure payoff considerations. This enabled us to see how this difference interacts with payoff delay. The design of the stage game is based primarily on that used by Andreoni (1995), and while we find a framing effect in the *NOW* treatment, the direction is opposite of his result. To conclude this section, we will briefly discuss this.

Over the past two decades, many experimental studies have used designs based on Andreoni (1995). The majority of these studies uses Give and Take frames, with varying results (see Cox and Stoddard (2015) for an overview).[22] Two studies (Park, 2000; Fujimoto and Park, 2010)) have used the exact same framing and instructions as the original, and both replicate the result.

While our findings confirm that the way in which a decision is framed matters, we can only speculate about what exactly flipped the result. Two differences between the current study and the original study by Andreoni (1995) stand out. Firstly, our study used a one-shot game rather than a finitely repeated game. In discussing his finding, Andreoni suggests an interaction between the frames and group dynamics: "If an action is described (..) as it is in the negative-frame condition, then the guilt from taking that action may

[21]In our data, Levene's test for equality of variances confirms that there is significantly more variation among men, but only when pooling the observations from all four treatments.

[22]The Give and Take frames both describe a positive externality, but in the initial allocation all tokens are either in the private (Give) or the group (Take) account. The difference is discussed in more detail by Cartwright (2016).

be diminished the more others do the same. In contrast, the pride one takes in choosing the action that creates the positive externality may not be diminished if others also do not choose the positive action(p.13)". If anticipated, such dynamics can create quite different strategic incentives between frames.

Secondly, we ran the experiment online rather than on a university campus, which means participants are not in physical proximity to one another as they are in a classroom, enhancing the anonimity. Perhaps a more important consequence is that only 20% of our participants were students, while in the original study all participants were economics students. Consistent with existing literature (e.g. Anderson et al. (2013); Belot et al. (2015); Gächter et al. (2004)), we do observe that students contribute less than non-students overall,[23] but in the *NOW* treatments, contributions are smaller in the positive frame both for students and for non-students.

3.6 Conclusion

In this paper, we have presented the results of an online experiment, designed to study intertemporal decisions in the social domain. We specifically considered impure motives for pro-social behavior, by combining two framing treatments with delayed or immediate payout of the experimental earnings.

Without delay, we found contributions to the public good were greater in a negative externality frame than in a positive externality frame. Delaying payments by thirty days eliminated this difference. Significantly lowered contribution levels in the negative frame suggest delay made it easier to not feel bad about reducing the earnings of others. We did not observe a similar effect in the positive frame, where the contributions saw a non-significant increase as payments were delayed.

Our experimental results demonstrate that contribution decisions can be affected by payoff delay, and, more importantly, that this interacts with how the decision is framed. For researchers and policy-makers alike, this finding warrants caution when extrapolating experimental findings on framing effects to situations in which the timing of decisions and consequences may be different.

Although we did not foresee this when designing the experiment, our findings indicate the *direction* of the delay effect may be frame-dependent. While it is clear that impure motives matter to explain delay effects, further research is needed for a deeper understanding of this differential effect of delay, which should ultimately generate more detailed insights into the impure motivations themselves.

[23] Across all treatments the difference is close to 12 percentage points, and statistically significant.

References

Anderson, Jon, Stephen V. Burks, Jeffrey Carpenter, Lorenz Götte, Karsten Maurer, Daniele Nosenzo, Ruth Potter, Kim Rocha, and Aldo Rustichini, "Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples," *Experimental Economics*, Jun 2013, *16* (2), 170–189.

Andreoni, James, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, 1989, *97* (6), 1447–1458.

_ , "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal*, 1990, *100* (401), 464–477.

_ , "An Experimental Test of the Public-Goods Crowding-Out Hypothesis," *The American Economic Review*, 1993, *83* (5), 1317–1327.

_ , "Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments*," *The Quarterly Journal of Economics*, 1995, *110* (1), 1–21.

_ , "Chapter 18 Philanthropy," in Serge-Christophe Kolm and Jean Mercier Ythier, eds., *Applications*, Vol. 2 of *Handbook of the Economics of Giving, Altruism and Reciprocity*, Elsevier, 2006, pp. 1201 – 1269.

_ **and A. Abigail Payne**, "Do Government Grants to Private Charities Crowd Out Giving or Fund-raising?," *American Economic Review*, 2003, *93* (3), 792–812.

_ **and Lise Vesterlund**, "Which is the Fair Sex? Gender Differences in Altruism," *The Quarterly Journal of Economics*, 2001, *116* (1), 293–312.

_ **and Marta Serra-Garcia**, "Time-Inconsistent Charitable Giving," 2017. NBER Working Paper No. 22824.

Ariely, Dan, Anat Bracha, and Stephan Meier, "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," *American Economic Review*, 2009, *99* (1), 544–555.

Ashraf, Nava, Dean Karlan, and Wesley Yin, "Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines*," *The Quarterly Journal of Economics*, 05 2006, *121* (2), 635–672.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger, "Working over Time: Dynamic Inconsistency in Real Effort Tasks," *The Quarterly Journal of Economics*, 2015, *130* (3), 1067–1115.

Belot, Michele, Raymond Duch, and Luis Miller, "A comprehensive comparison of students and non-students in classic experimental games," *Journal of Economic Behavior & Organization*, 2015, *113*, 26 – 33.

Bolton, Gary E. and Axel Ockenfels, "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 2000, *90* (1), 166–193.

Breman, Anna, "Give more tomorrow: Two field experiments on altruism and intertemporal choice," *Journal of Public Economics*, 2011, *95* (11-12), 1349–1357.

Cartwright, Edward, "A comment on framing effects in linear public good games," *Journal of the Economic Science Association*, May 2016, *2* (1), 73–84.

Cox, Caleb A. and Brock Stoddard, "Framing and Feedback in Social Dilemmas with Partners and Strangers," *Games*, 2015, *6* (4), 394–412.

Croson, Rachel and Uri Gneezy, "Gender Differences in Preferences," *Journal of Economic Literature*, June 2009, *47* (2), 448–74.

Crumpler, Heidi and Philip J. Grossman, "An experimental test of warm glow giving," *Journal of Public Economics*, 2008, *92* (5-6), 1011–1021.

DellaVigna, Stefano and Ulrike Malmendier, "Paying Not to Go to the Gym," *American Economic Review*, June 2006, *96* (3), 694–719.

_ , John A. List, and Ulrike Malmendier, "Testing for altruism and social pressure in charitable giving," *Quarterly Journal of Economics*, 2012, *127* (1), 1–56.

Dreber, Anna, Drew Fudenberg, David K. Levine, and David G. Rand, "Self-Control, Social Preferences and the Effect of Delayed Payments," 2016. mimeo.

Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt, "The framing of games and the psychology of play," *Games and Economic Behavior*, 2011, *73* (2), 459 – 478.

Engelmann, Dirk and Martin Strobel, "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments," *American Economic Review*, 2004, *94* (4), 857–869.

Fehr, Ernst and Klaus M. Schmidt, "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 1999, *114* (3), 817–868.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue, "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, June 2002, *40* (2), 351–401.

Fudenberg, Drew and David K. Levine, "A Dual-Self Model of Impulse Control," *American Economic Review*, December 2006, *96* (5), 1449–1476.

Fujimoto, Hiroaki and Eun-Soo Park, "Framing effects and gender differences in voluntary public goods provision experiments," *The Journal of Socio-Economics*, 2010, *39* (4), 455 – 457.

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

Huang, Yunhui and Lei Wang, "Sex differences in framing effects across task domain," *Personality and Individual Differences*, 2010, *48* (5), 649 – 653.

Kellner, Christian, David Reinstein, and Gerhard Riener, "Ex-ante commitments to "give if you win" exceed donations after a win," *Journal of Public Economics*, 2019, *169*, 109 – 127.

Khadjavi, Menusch and Andreas Lange, "Doing good or doing harm: experimental evidence on giving and taking in public good games," *Experimental Economics*, 2015, *18* (3), 432–441.

Klinowski, David, "Gender differences in giving in the Dictator Game: the role of reluctant altruism," *Journal of the Economic Science Association*, Dec 2018, *4* (2), 110–122.

Koelle, Felix and Thomas Lauer, "Delay and Cooperation," 2018. mimeo.

Kovarik, Jaromir, "Giving it now or later: Altruism and discounting," *Economics Letters*, 2009, *102* (3), 152–154.

Krupka, Erin and Roberto Weber, "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?," *Journal of the European Economic Association*, 2013, *11* (3), 495–524.

Loewenstein, George, Ted O'Donoghue, and Sudeep Bhatia, "Modeling the interplay between affect and deliberation," *Decision*, 2015, *2* (2), 55–81.

Name-Correa, Alvaro J. and Huseyin Yildirim, ""Giving" in to social pressure," *Games and Economic Behavior*, 2016, *99*, 99–116.

Ottoni-Wilhelm, Mark, Lise Vesterlund, and Huan Xie, "Why Do People Give? Testing Pure and Impure Altruism," *American Economic Review*, 2017, *107* (11), 3617–3633.

Palfrey, Thomas R. and Jeffrey E. Prisbrey, "Anomalous Behavior in Public Goods Experiments: How Much and Why?," *The American Economic Review*, 1997, *87* (5), 829–846.

Park, Eun-Soo, "Warm-glow versus cold-prickle: a further experimental study of framing effects on free-riding," *Journal of Economic Behavior & Organization*, 2000, *43* (4), 405 – 421.

Rogers, Todd and Max H. Bazerman, "Future lock-in: Future implementation increases selection of 'should' choices," *Organizational Behavior and Human Decision Processes*, 2008, *106* (1), 1 – 20.

Trope, Yaacov and Nira Liberman, "Temporal construal and time-dependent changes in preference.," *Journal of Personality and Social Psychology*, 2000, *79* (6), 876–889.

Vesterlund, Lise, "Using Experimental Methods to Understand Why and How We Give to Charity," in John H. Kagel and Alvin E. Roth, eds., *The Handbook of Experimental Economics, Volume Two*, Princeton University Press, 2016.

3.7 Appendix to Chapter 3

3.7.1 Additional Regressions

Table 3.6: OLS Estimations of effects on g - Gender Differences.

	Positive		Negative		Both Frames	
Expl. Variables	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.188**	-0.188**	-0.164**	-0.170**	-0.171***	-0.175***
	(0.079)	(0.078)	(0.083)	(0.083)	(0.058)	(0.057)
FUT	0.008	-0.003	-0.250***	-0.254***	-0.117**	-0.124**
	(0.078)	(0.078)	(0.085)	(0.085)	(0.058)	(0.058)
Female × FUT	0.102	0.096	0.255**	0.255**	0.175**	0.174**
	(0.112)	(0.112)	(0.117)	(0.116)	(0.082)	(0.081)
Ad. Controls	No	Yes	No	Yes	No	Yes
Constant	0.556***	0.591***	0.724***	0.753***	0.638***	0.671***
	(0.056)	(0.059)	(0.061)	(0.064)	(0.042)	(0.044)
Observations	162	162	165	165	327	327

Note: Dependent variable is the share of tokens contributed to Group Box ($g_i/60$). Additional control variables: Student and unemployment status. Standard errors in parentheses. *** p<.01, ** p<.05, * p<.1

Table 3.7: OLS Estimations of effects on g - Gender Differences.

	NOW	FUT	NOW	FUT
Expl. Variables	(1)	(2)	(3)	(4)
Female	-0.171***	0.004	-0.188**	-0.086
	(0.056)	(0.060)	(0.077)	(0.086)
Neg. Frame			0.168**	-0.090
			(0.079)	(0.084)
Female × Neg. Frame			0.024	0.178
			(0.109)	(0.120)
Constant	0.638***	0.521***	0.556***	0.564***
	(0.040)	(0.042)	(0.055)	(0.058)
Observations	164	163	164	163

Note: Dependent variable is the share of tokens contributed to Group Box ($g_i/60$). Standard errors in parentheses. *** p<.01, ** p<.05, * p<.1

3.7.2 Instructions

All instructions were provided onscreen. Here we will provide screenshots of the instruction screens and the choice screen.

Figure 3.7: General Instruction Screen

Choice Task: Basics

You will be in an anonymous group of 4, and each of you is presented with the same task.

You can earn additional money in this task. Your additional income will depend on your decision and on the decisions of the other 3 group members.

During the choice task, all income will be expressed in Coins. At payout, the Coins you earned will be converted to GBP, where **100 Coins = £1.00.**

The maximal amount you can earn in this task is **£3.00**. The minimal amount you can earn is **£0.60**.

Important information: Money earned in the choice task will be transferred via Prolific. The following payment schedule applies:

- All payments related to the choice task will be transferred 30 DAYS FROM NOW.
- Your base payment (£1.00) does not depend on the choice task and will be transferred tomorrow.

This is the same for all 4 group members: everyone does the task *today*, finds out the results tomorrow (see below), and will receive the money earned during the choice task *in 30 days*.

Tomorrow, you will receive an e-mail with your Earnings Report, which summarizes your choices and the choices made by the rest of the group, and informs you of your payout amount. Please note that the report will not tell you the individual decisions or individual payout amounts of the other members of your group. Your decisions and earnings are confidential.

Note: Basic instructions about payout and timing for the *FUT* treatments.

Figure 3.8: Task and payoff information – Positive

Choice Task: Details

You are in a group of 4 people.

Every group member has the same task: **Each group member gets 60 tokens and has to decide how to divide them between two boxes: the Individual Box and the Group Box.** Your final income from this task and the final income of your group members will depend on how you allocate your tokens *and* on how the other three group members allocate their tokens.

The rules are as follows:

- Every token you put in the Individual Box increases your income by 2 Coins and leaves the income of the other 3 group members unaffected.
- Every token you put in the Group Box increases your income by 1 Coin. At the same time, the income of the other three group members *is also increased* by 1 Coin each.

The same rules apply for the other 3 group members:

- Every token they put in their Individual Box does not affect your income. Every token they put in the Group Box increases your income by 1 Coin.

INDIVIDUAL BOX		
	Your Income	**Income of the Other Group Members (each)**
Income per Token	+ 2 Coins	0 Coins
Payout Date	In 30 days	-

GROUP BOX		
	Your Income	**Income of the Other Group Members (each)**
Income per Token	+ 1 Coin	+ 1 Coin
Payout Date	In 30 days	In 30 days

Notice that for each token in the Individual Box, you earn 2 Coins. If instead you put this token in the Group Box, you earn less individually, but it also increases the income of the other group members. A token in the Group Box increases the income of all group members by 1 Coin.

Final Income

To summarize, your final income will consist of:

- Income from tokens you put in your Individual Box (1 Token = + 2 Coins)
- Income from tokens you put in the Group Box (1 Token = + 1 Coin)
- Income from tokens the other three group members put in the Group Box (1 Token = + 1 Coin)

Note: Information about the task for the $FUT(+)$ treatment. Additionally, the participants were provided with a simulator to see how different token allocations, by them and their fellow group members, would result into different payoffs for them and the others.

Figure 3.9: Task and payoff information – Negative

Choice Task: Details

You are in a group of 4 people.

Every group member has the same task: **Each group member gets 60 tokens and has to decide how to divide them between two boxes: the Individual Box and the Group Box.** Your final income from this task and the final income of your fellow group members will depend on how you allocate your tokens *and* on how the other three group members allocate their tokens.

The rules are as follows:

- Every token you put in the Group Box increases your income by 1 Coin and leaves the income of the other 3 group members unaffected.
- Every token you put in your Individual Box increases your income by 2 Coins but *reduces* the income of the other 3 group members by 1 Coin each.

The same rules apply for the other 3 group members:

- Every token they put in the Group Box does not affect your income. Every token they put in their Individual Boxes reduces your income by 1 Coin.

At the end of the task, a fixed amount of 180 Coins wil be automatically added to every group member's income. As a result, no group member can end up with negative income.

INDIVIDUAL BOX		
	Your Income	**Income of the Other Group Members (each)**
Income per Token	+ 2 Coins	-1 Coin
Payout Date	Tomorrow	Tomorrow

GROUP BOX		
	Your Income	**Income of the Other Group Members (each)**
Income per Token	+ 1 Coin	0 Coins
Payout Date	Tomorrow	-

Notice that for each token in the Group Box, you earn 1 Coin. If instead you put this token in the Individual Box, you earn more individually, but it reduces the income of the other group members. A token in the Individual Box increases the income of one group member by 2 Coins and reduces the income of the other group members by 1 Coin.

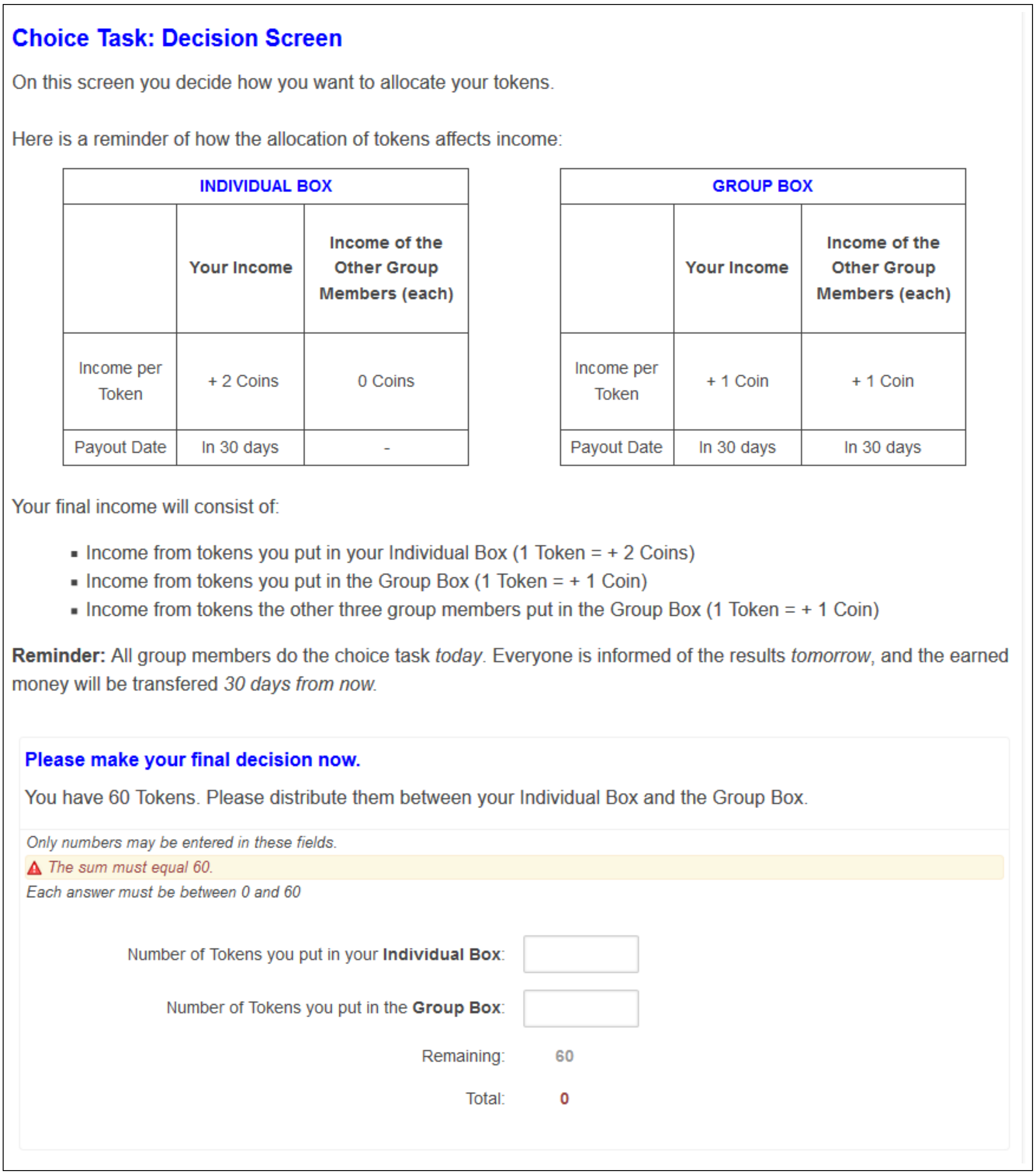
Final Income

To summarize, your final income will consist of:

- Income from tokens you put in your Individual Box (1 Token = + 2 Coins)
- Income from tokens you put in the Group Box (1 Token = + 1 Coin)
- Reductions caused by tokens the other group members put in their Individual Boxes (1 Token = -1 Coin)
- Automatic income added at the end of the task (+ 180 Coins)

Note: Information about the task for the $NOW(-)$ treatment.

Figure 3.10: Decision screen



Choice Task: Decision Screen

On this screen you decide how you want to allocate your tokens.

Here is a reminder of how the allocation of tokens affects income:

INDIVIDUAL BOX		
	Your Income	**Income of the Other Group Members (each)**
Income per Token	+ 2 Coins	0 Coins
Payout Date	In 30 days	-

GROUP BOX		
	Your Income	**Income of the Other Group Members (each)**
Income per Token	+ 1 Coin	+ 1 Coin
Payout Date	In 30 days	In 30 days

Your final income will consist of:

- Income from tokens you put in your Individual Box (1 Token = + 2 Coins)
- Income from tokens you put in the Group Box (1 Token = + 1 Coin)
- Income from tokens the other three group members put in the Group Box (1 Token = + 1 Coin)

Reminder: All group members do the choice task *today*. Everyone is informed of the results *tomorrow*, and the earned money will be transfered *30 days from now*.

Please make your final decision now.

You have 60 Tokens. Please distribute them between your Individual Box and the Group Box.

Only numbers may be entered in these fields.
The sum must equal 60.
Each answer must be between 0 and 60

Number of Tokens you put in your **Individual Box**:

Number of Tokens you put in the **Group Box**:

Remaining: 60

Total: 0

Note: Decision screen for the $FUT(+)$ treatment.

Appendix to Dissertation

List of Publications

Chapter 1: Negotiating Non-Binding Burden Sharing Agreements when Renegotiations are Costly
Author: Arne Pieters
Publication status: unpublished (journal submission planned for fall 2019)

Chapter 2: Asymmetric Risk and Non-Binding Rules: An Experiment
Author: Arne Pieters
Publication status: unpublished (journal submission planned for summer 2019)

Chapter 3: Cooperating Tomorrow: Warm Glow vs. Cold Prickly Revisited
Authors: Arne Pieters and Arno Apffelstaedt
Publication status: unpublished (journal submission planned for summer 2019)

Erklärung

Hiermit erkläre ich, Arne Pieters, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die Arbeit wurde nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

....

Ort/Datum Unterschrift Doktorand

Eidesstattliche Versicherung

Ich, Arne Pieters, versichere an Eides statt, dass ich die Dissertation mit dem Titel

"Essays on Cooperation"

selbst und bei einer Zusammenarbeit mit anderen Wissenschaftlerinnen oder Wissenschaftlern gemäß den beigefügten Darlegungen nach § 6 Abs. 6 der Promotionsordnung der Fakultät für Wirtschafts- und Sozialwissenschaften vom 18. Januar 2017 verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht benutzt.

....

Ort/Datum Unterschrift Doktorand

....

Unterschrift Verwaltung

Selbstdeklaration bei kumulativen Promotionen

Die folgende Einschätzung in Prozent über die von mir erbrachte Eigenleistung wurde mit den beteiligten Koautoren einvernehmlich abgestimmt.

Negotiating Non-Binding Burden Sharing Agreements when Renegotiations are Costly (Chapter 1)

Autor: Arne Pieters (Alleinautorschaft)

Die Eigenleistung liegt bei:

Konzeption/Planung:	100%
Durchführung:	100%
Manuskripterstellung:	100%

Asymmetric Risk and Non-Binding Rules: An Experiment (Chapter 2)

Autor: Arne Pieters (Alleinautorschaft)

Die Eigenleistung liegt bei:

Konzeption/Planung:	100%
Durchführung:	100%
Manuskripterstellung:	100%

Cooperating Tomorrow: Warm Glow vs. Cold Prickle Revisited (Chapter 3)

Autoren: Arno Apffelstaedt und Arne Pieters

Die Eigenleistung liegt bei:

Konzeption/Planung:	65%
Durchführung:	50%
Manuskripterstellung:	50%

Dabei gilt: *Konzeption/Planung*: Formulierung des grundlegenden wissenschaftlichen Problems, basierend auf bisher unbeantworteten theoretischen Fragestellungen inklusive der Zusammenfassung der generellen Fragen, die anhand von Analysen oder Experimenten/Untersuchungen beantwortbar sind. Planung der Experimente/Analysen und Formulierung der methodischen Vorgehensweise, inklusive Wahl der Methode und unabhängige methodologische Entwicklung. *Durchführung*: Grad der Einbindung in die konkreten Untersuchungen bzw. Analysen. *Manuskripterstellung*: Präsentation, Interpretation und Diskussion der erzielten Ergebnisse in Form eines wissenschaftlichen Artikels.

....	....
Ort/Datum	Unterschrift Doktorand