Essays on Climate Negotiations and the Measurement of Life Satisfaction

kumulative Dissertation

Universität Hamburg
Fakultät für Wirtschafts- und Sozialwissenschaften

Dissertation
Zur Erlangung der Würde der Doktorin der Wirtschafts- und Sozialwissenschaften
Dr. rer. pol.
(gemäß der Promotionsordnung vom 24. August 2010)

vorgelegt von
Sonja Maria Köke, M.Sc.
aus Nürnberg

Hamburg 2018

urn:nbn:de:gbv:18-87827

Vorsitzende/r: Prof. Dr. Gerd Mühlheußer

Erstgutachter/in: Prof. Dr. Andreas Lange

Zweitgutachter/in: Prof. Dr. Grischa Perino

Datum der Disputation: 30.11.2018

Acknowledgments

I want to thank Andreas Lange for being my supervisor, for all his support during my PhD, for giving me the right amount of guidance and freedom to find my way through the different facets of research, teaching and academia in its broadest sense. I also want to thank Grischa Perino for being my second supervisor and daring to explore a completely new topic with me. I want to thank Andreas Nicklisch for the inspiring co-authorship of the experimental paper and for introducing me to the world of experiments. Thanks also to all the wonderful colleagues at Hamburg University. Finally yet importantly, I want to thank my parents, brothers and sisters, my partner Maikel de Vries, my dancing partners, roommates and friends. You were there for me whenever I needed you during this sometimes stormy but very enriching project of my dissertation.

Contents

1 Introduction **1**

2 Negotiating Environmental Agreements under Ratification Constraints **9**

2.1 Introduction . . . 9
2.2 The theoretical model . . . 14
2.2.1 The agents, their preferences and the temporal structure . . . 14
2.2.2 The non-cooperative policy (S4) . . . 17
2.2.3 The ratification decision (S3) . . . 17
2.2.4 Choosing the content of the agreement (S2) . . . 19
2.2.5 Decision to join the negotiation table (S1) . . . 23
2.2.6 Alternative model specification: ratifiers setting unilateral policy . 29
2.3 Conclusion . . . 33
2.4 References . . . 36

3 Is Adversity a School of Wisdom? Experimental Evidence on Cooperative Protection Against Stochastic Losses **39**

3.1 Introduction . . . 39
3.2 Experimental Design and Predictions . . . 43
3.2.1 Experimental Treatments . . . 43
3.2.2 Predictions . . . 46
3.2.3 Experimental Procedure . . . 49
3.3 Results . . . 50
3.3.1 Average Treatment Differences . . . 50
3.3.2 Dynamics of Individual Behavior . . . 54
3.4 Explaining the Behavioral Dynamics . . . 58
3.5 Conclusion . . . 61
3.6 References . . . 62

4 For "better" or "worse": a new life-satisfaction measure able to elicit preference rankings **65**

4.1 Introduction . . . 65
4.2 Preference rankings and the level measure of life satisfaction . . . 69

4.2.1 Reporting functions and measurement of life satisfaction 70
4.2.2 Within-subject variation of reporting functions 70
4.2.3 Reporting life satisfaction when its drivers follow a trend 73
4.3 Experimental evidence on the level measure 74
4.3.1 The experiment 74
4.3.2 Recollection ability and precision 75
4.3.3 Reassessment ability 77
4.4 The ranking measure 78
4.4.1 When does the ranking measure capture preferences? 78
4.4.2 Advantages of the ranking measure 79
4.5 Comparison of measures: empirical evidence 80
4.5.1 Data 80
4.5.2 Consistency between the two life satisfaction measures 81
4.5.3 Drivers of life satisfaction measures 82
4.5.4 Adaptation 84
4.5.5 Trends in socio-economic variables and changes in reported LS . . 85
4.6 Conclusion and Discussion 90
4.7 References 92

5 Valuation of non-market goods: the life satisfaction approach under methodological examination 97

5.1 Introduction 97
5.2 The Life Satisfaction Approach to Valuation 101
5.2.1 The level measure of life satisfaction 101
5.2.2 The ranking measure of life satisfaction 104
5.3 Empirical Analysis 106
5.3.1 Data 107
5.3.2 Results 109
5.4 Discussion and conclusion 117
5.5 References 118

A Appendix Chapter 2 121

A.1 Proofs 121

B Appendix Chapter 3 129

B.1 Derivation of conditions for cooperative equilibria 129
B.2 Experimental Instructions for the DamRed Treatment (English translation) 131

C Appendix Chapter 4 135

C.1 Description of variables 135
C.2 Additional regression results 141

D Appendix Chapter 5 **153**
D.1 Summary statistics 153
D.2 Results of Powdthavee (2008) 155

E Formal requirements **161**
E.1 Zusammenfassung 161
E.2 Abstract 165
E.3 List of publications from this thesis 169

List of Figures

2.1 *Representatives setting unilateral policy: Maximal stable coalition n^* and corresponding abatement level q^* (relative to b^R/c) under certainty as a function of λ ($b^P = \lambda b^R$).* 27

2.2 *Pivotal agents setting unilateral policy: Maximal stable coalition n^* and corresponding abatement level q^* (relative to b^R/c) under certainty as a function of λ ($b^P = \lambda b^R$).* 27

3.1 *Minimal cooperation level Q^{min} required to stabilize cooperation as a function of risk aversion σ for CRRA preferences ($u(\pi) = \pi^{1-\sigma}/(1-\sigma)$). Parameters as used in experiment ($\delta = 0.2$, $p_0 = 0.5$, $x = 0.1$, $D_0 = 20$, $d = 4$, $E = 25$, $c = 5$).* 48

3.2 *Mean cooperation frequency per period by treatment.* 51

3.3 *Mean cooperation in the first period of all supergames across treatment conditions* 53

B.1 Decision screen for taking the action in a round in Part 1 of the experiment. 134

B.2 Decision screen for the risk-assessment task in Part 2 of the experiment. . 134

D.1 Distribution of life satisfaction measures depending on levels of *Real household income per year and capita* (BHPS 1997-2009). Blue is median, red is the mean. 154

D.2 Histogram of *Real household income per capita* in 1000 GBP (1997). 166,214 observations. Values below 0 and above 100,000 GBP were excluded (5,389 and 87 observations respectively). 155

D.3 *Table 2 from Powdthavee (2008).* 156

D.4 *Table 3 from Powdthavee (2008).* 157

D.5 *Table 5 from Powdthavee (2008).* 158

D.6 *Table 6 from Powdthavee (2008).* 159

List of Tables

3.1 *Summary of damage size $D^{Treat}(Q^t)$ and damage probability $p^{Treat}(Q^t)$ for the respective treatments ProbRed, ProbRed$^+$, DamRed, and CertDam.* 45

3.2 *Average cooperation rates by treatments over the entire experiment (left panel), over the first periods of all supergames (middle panel), and over the last periods of all supergames (right panel), tests refer to two-sided Wilcoxon Mann-Whitney rank sum tests, *** indicates significance at a $p < 0.01$ level, ** at a $p < 0.05$ level and * at a $p < 0.1$ level.* 50

3.3 *Left panel: linear regression of cooperation behavior in the first period, right panel: random effects regression of cooperation behavior in all periods of the experiment; coefficients are reported along standard errors in parenthesis (errors are clustered at the matching group level); *** indicates significance at a $p < 0.01$ level, ** at a $p < 0.05$ level and * at a $p < 0.1$ level. obs reports the number of observations while n reports the number of subjects; models' fitness are assessed by F-test and Wald-Chi^2-tests.* . . 52

3.4 *Random-effects linear regression of time trends for individual cooperation decision q_i^t; coefficients are reported along standard errors in parenthesis (errors are clustered at the matching group level); *** indicates significance at a $p < 0.01$ level, ** at a $p < 0.05$ level and * at a $p < 0.1$ level. obs reports the number of observation while n report the number of subjects; model's fitness is assessed by a Wald-Chi^2-test.* 54

3.5 *Mean q_i^{t+1} given q_i^t and the occurrence of the damage s^t; tests refer to two-sided Wilcoxon Mann-Whitney rank sum tests, *** indicates significance at a $p < 0.01$ level, ** at a $p < 0.05$ level and * at a $p < 0.1$ level.* 55

3.6 *Estimation results for an Arellano-Bond panel regressions with dependent variable q_i^{t+1}; coefficients are reported along standard errors in parenthesis; *** indicates significance at a $p < 0.01$ level, ** at a $p < 0.05$ level and * at a $p < 0.1$ level. Standard errors are clustered at the matching group level. obs reports the number of observation while n reports the number of subjects; models' fitness are assessed by Wald-Chi^2-tests.* 57

4.1 Order of questions asked in Lightspeed survey. 400 participants per treatment. 75

4.2 Self-assessed and real accuracy of recollection of LM1. LM1recall and accuracy conditional on answering LM1. 76

4.3 Comparison of RM and signdLM. Standard error in brackets. Lightspeed based on treatments Control and RecallExPost. Difference from zero is reported at * $p < 0.05$, **$p < 0.01$, ***$p < 0.001$ significance levels. 81

4.4 BHPS: Regressions on key socio-economic variables (financial situation). Variable description in Appendix C.1. Results from pooled OLS and fixed effects regressions including year fixed effects. 86
4.5 BHPS: Regressions on key socio-economic variables (income). Variable description in Appendix C.1. Results from pooled OLS and fixed effects regressions including year fixed effects (not reported). 87
4.6 GESIS: Regressions on key socio-economic variables. Variable description in Appendix C.1. Results from pooled OLS and fixed effects regressions including year fixed effects (not reported). 88
4.7 Deviation between measures on direction of change and trends in financial situation (BHPS). Description of variables given in Appendix C.1. Results from pooled OLS regression including year fixed effects. 89

5.1 Frequency of social interaction with friends/family and neighbours 108
5.2 Frequency of Health status and Days in hospital 108
5.3 Frequency of signdLM and RM 108
5.4 Frequencies of Real household income per year and capita by LM 109
5.5 Valuation results with the level measure and the ranking measure of life satisfaction (BHPS 1997-2009) 112
5.6 Valuation results for different household income variables (BHPS 1997-2009) 114
5.7 Valuation results for different personal income variables (BHPS 1997-2009) 115
5.8 Valuation results for different logartihmic income variables (BHPS 1997-2009) 116

A.1 *Stable coalition sizes* n^**, optimal decisions of* $\hat{q}^*$ *and* $\hat{r}^*$*, and expected utility gains* $\Delta E[U^R]/U^{RN}$ *for representatives (*$\lambda = E(b^P)/b^R \in \{1.1, 1, 0.25\}$ *(*$\sigma^2_\theta = 0$*), normal distribution of country-specific shocks* $\sigma^2_\epsilon \in \{0, 10, 20, 100\}$*,* $b^R = 50$*,* $c = 50$*).* 125

C.1 Summary Statistics of BHPS variables 135
C.2 Summary Statistics of GESIS variables 136
C.3 Summary Statistics of Lightspeed Variables 138
C.4 Lightspeed: Ordered logit regressions on changes in perceived changes in socio-economic circumstances (Treatments: Control and RecallExPost). . 141
C.5 BHPS: Regressions on key socio-economic variables (financial situation) without lags. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects. 142
C.6 BHPS: Regressions on key socio-economic variables (income) without lags. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects. 143
C.7 BHPS: Regressions on key socio-economic variables (financial situation) including only the three first years of the panel, making it comparable to the current time horizon of GESIS. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects. 144

C.8 BHPS: Regressions on key socio-economic variables (income) including only the three first years of the panel, making it comparable to the current time horizon of GESIS. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects. 145
C.9 GESIS: Regressions on key socio-economic variables (income). Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects. 146
C.10 BHPS: Regressions on key socio-economic variables (financial situation). Variable description in Appendix A.1. Results from ordered logistic regressions including year fixed effects. 147
C.11 Deviation between measures on direction of change and trends in financial situation using leads instead of lags to capture expected trends (BHPS). Description of variables given in Appendix A.1. Results from pooled OLS regression including year fixed effects. 148
C.12 BHPS: Regressions on key socio-economic variables (log of income). Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects. 149
C.13 BHPS: Regressions on key socio-economic variables (financial situation) excluding observations where the current LM is either 1 or 7. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects. 150
C.14 BHPS: Regressions on key socio-economic variables (financial situation) excluding observations where the lag of LM is either 1 or 7. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects. 151

D.1 Summary Statistics of BHPS Variables 1997-2003 (2001 missing) 153
D.2 Personal income per month by LM 154

Chapter 1

Introduction

Protecting our natural environment is getting more and more important to modern societies. Most environmental goods and services are not efficiently provided through individual decisions on markets due to different types of market failure, and thus political intervention is required to reach the socially optimal amount of provision, depletion, pollution, conservation and mitigation. A growing population and a growing affluence over the last decades have made this need for intervention more evident and pressing since the negative consequences get increasingly obvious. For instance, destroying the ozone layer through CFCs, cutting down the (rain) forests, overfishing the oceans, particulate matter at a local level and greenhouse gas emissions at the global level to only name a few phenomena that come with growing population and growing economic activity, have considerable negative impact on welfare and thus call for action. Some of those problems are easier others more difficult to solve. All of them involve some type of social dilemma that needs to be understood, since individual and social interests are not well aligned. Two key lines of investigation in environmental economics are firstly to understand the structure of the social dilemma at hand in order to find tools that can then reduce or eliminate the dilemma and secondly to measure the magnitude of discrepancy between the actual situation and the situation where the dilemma is resolved. Take for example the Pigouvian tax (Pigou 1932, Mas-Colell et al. 1995). It is important to understand the structure of externalities as a social dilemma in order to see that implementing a tax, which forces market participants to internalize the external effects they cause, can reach an efficient outcome. However, in order to implement the tax, we also need methods that provide us with reliable and valid estimates of the size of the externality, in order to know how large the tax has to be. Environmental economics has been making a lot of progress on these two types of questions in different applications in recent years. The understanding of the underlying dilemmas such as external effects and public goods have been extended and refined. Increasingly adequate tools for different types of environmental protection as well as improved methods for identifying the value of environmental goods and services have been developed. Both theoretical and empirical approaches have been used to answer those pressing questions.

This thesis consists of four main chapters. Two of them are about better understanding the structures of the environmental problem at hand and thus the first type of question described above. They are dealing with the challenge of providing public goods, analyzing how two types of uncertainty affect the amount of public good provision using game theory and laboratory experiments. The other two chapters address the challenge of measuring life satisfaction and of using it for valuation of non-market goods. They thus make a methodological contribution to answering the question of “how much environmental protection do we need?” using panel data analysis. In what follows I give an overview over the two parts and the four papers of the thesis.

Part 1: Public good provision and two types of uncertainty

International climate negotiations are known to be a long and difficult process. The conferences of the parties (COP) of the United Nations Framework Convention on Climate Change (UNFCCC) have last year taken place for the 23rd time and still the goal of achieving not more than a 2 °C increase of global average temperature seems far from achieved when looking at actions taken (e.g. Rogelj et al. 2017). Not only is the reduction of greenhouse gases a public good, but it also has a global and inter-temporal scale, making it even harder to reach the optimal provision. The literature on coalition formation has largely shown the impossibility of forming a stable coalition that is large and at the same time incentivizes its members to substantial abatement efforts (e.g. Barrett 1994, Marrouch and Chaudhuri 2016). Those models typically model the participating countries as single players (with few exceptions - e.g. Marchiori et al. 2017). However, there are many agents involved in negotiations and in the implementation of an agreement. For instance, the person signing and the body ratifying the agreement are not identical in many countries. Due to the sequence, not all signing countries thus end up being members if some of them fail to ratify. There is thus more uncertainty in the decision process in reality, which is not captured in those existing models. Borrowing from the political sciences (Putnam 1988), the first paper (chpater 2) thus introduces a ratification process into a coalition formation model, allowing for ratification failure. It explores the effect such ratification constraints and ratification uncertainty have on the process and outcome of negotiations. By incorporating uncertain preferences of the pivotal voter in the ratification stage, this paper makes several important contributions: (i) it contributes to a more realistic modeling of the temporal structure of international agreements, (ii) it shows a surprising, yet intuitive non-monotonic relationship between the optimal commitment level and the variance of ratification decisions, (iii) it identifies reasons to expect a larger number of countries to join international negotiations than predicted by most of the coalition formation literature based on a representative agent model. Ratification constraints thereby can improve the welfare gains from stable international agreements. Game theory seems to be a useful tool to analyze environmental agreements since there is no international government to enforce the socially optimal outcome. The success of international negotiations thus hinges on the self-enforcing properties of the structure of the agreement. While the payoff-structure assumed in such a game-theoretic setup is quite general, it still captures the main properties of the social dilemma at hand and is thus helpful in identifying general insights into the functioning of international negotiations on the provision of global public goods.

While the first paper analyzes uncertainty stemming from the process of international decision-making, the second paper explores uncertainty structures in the public good itself. In the context of climate negotiations, the outcome when emitting a specific amount of greenhouse gases is not certain. The contributions to the public good might reduce the probability of a storm hitting or it might reduce the strength of the storm and thus the size of the damages caused. The second paper analyzes in a repeated pubic good game whether in a laboratory experiment participants are more willing to invest in reducing the size or to invest in reducing the probability of a probabilistic damage. It thus compares situations where contributions to the public good can reduce the size to a situation where it reduces the probability of a probabilistic damage. It does so in an indefinitely repeated game, which is implemented in the lab using a probabilistic ending of the game after each round. The results show that cooperation is larger and more stable when it affects the probability rather than the size of the adverse event. The paper provides crucial insights into behavioral adaptation: defecting players are more likely to switch to cooperation after experiencing an adverse event, while existing cooperation is reinforced when the losses do not occur. This behavior is consistent with simple learning dynamics based on ex post evaluations of the chosen strategy.

Laboratory experiments are a useful tool to identify the effect of interest here. Especially in a repeated game with the probabilistic structure we analyze here, the optimal (expected payoff maximizing) behavior over time is not straightforward anymore. The outcome does not only depend on the own choice but also on other group-members' choices and the probabilistic outcome of the damage and the ending of the game. When situations become complex, people tend to use "rules of thumb" to decide what to do (e.g. Baumol and Quandt 1964, Camerer and Ho 1999). Such rules of thumb can take many different forms. And as we find, they depend on the outcome of the damage event even though the damage event does not carry any new information and should therefore not lead to changing behavior. An experiment can thus help to identify behavioral patterns which cannot be predicted from standard game theory.

Part 2: Using life satisfaction to represent preference rankings and to infer willingness to pay for non-market goods

Life satisfaction has alongside with happiness gained increasing attention in economics over the last decades (since Easterlin 1979) as subjective measures of well-being. Utility, the individual level welfare concept in economic theory, is usually thought of as being not observable and that it can only be inferred from choices (e.g. Mas-Colell et al. 1995). It is an ordinal concept, giving higher numbers to more preferred situations than to less preferred ones. Instead of the traditional approach of observing behavior and deducing from there what people's preferences are, subjective well-being asks respondents of a survey to state their satisfaction or happiness on a categorical scale which often is then interpreted cardinally and understood as a proxy for utility. When regressing life satisfaction on different variables describing the situation of the person, correlates of life satisfaction can be inferred. Thus, the elements of the utility function can be detected and even the size of the impact and thus the functional form of utility (e.g. Clark et al. 2008). In this interpretation, life satisfaction is also used for valuation of non-market goods (e.g. Welsch and Ferreira 2014). Using the estimated utility function, the marginal rate of substitution between income and the non-market good is computed, which is then interpreted as willingness to pay. As described above, environmental goods often are not traded on markets or at least in an insufficient amount, such that utility cannot be observed from actual behavior (demand on markets). The government is thus interested in knowing how much utility the population derives from specific non-market goods in order to know how much provision would be optimal. There is a variety of methods that can be used for valuation of non-market goods (e.g. Freeman III 2003). There are revealed preference and stated preference approaches. Revealed preference approaches infer from observed demand for related goods how much value individuals attribute to the good in question. E.g. in the hedonic pricing approach (see e.g. Rosen 1974) it is possible to infer from housing price differentials how much more people are willing to pay for a house depending on its distance to a park, thereby revealing their willingness to pay for access to the park even though no house owner actually buys the park. Stated preference approaches instead ask people directly for their willingness to pay. E.g. with the contingent valuation method (see e.g. Mitchell and Carson 2013) respondents to a survey would be asked how much they are willing to pay for having a park built next to their house. The life satisfaction approach to valuation is somehow between revealed and stated preferences. It uses the life satisfaction question which is a "stated utility" and relates the answer of respondents to their income level and the amount of the non-market good in question to compute the marginal rate of substitution between income and the non-market good, which can be interpreted as the marginal willingness to pay. The willingness to pay is thus indirectly inferred, which is more in line with "revealed preferences from stated utility". Here it is not the willingness to pay that is stated by the respondent but the variables from which the willingness to pay is constructed. Such valuation methods are necessary in cost-benefit analysis (e.g. Atkinson and Mourato 2006) when deciding on public projects or regulations e.g. when building a dam there are explicit costs and benefits and there are costs and benefits which need to be estimated because the underlying values are less obvious. The different valuation methods all have advantages and disadvantages. It is therefore important to use the different approaches and to compare the results. It is also important to better understand and improve each of them.

The third paper investigates theoretically and empirically how well subjective measures of life satisfaction are able to measure preference rankings as used in economic choice or welfare theory. In order to overcome some of the detected shortcomings, an alternative measure (ranking measure) is proposed and empirically compared to the results of the currently used level measure. The paper provides empirical evidence that intra-personal changes in life-satisfaction are a biased measure of preference rankings and, as a remedy, propose the ranking measure where subjects state whether their life has become "better" or "worse". Three representative data sets show: intra-personal changes in satisfaction levels are dominated by noise and are less well explained by socio-economic variables than the ranking measure; the deviation between the two measures is systematic and adaptation (e.g. to income and unemployment) is only observed for changes in satisfaction levels but not for rankings, indicating that adaptation is driven by the elicitation method rather than changes in preferences.

The fourth paper applies the ranking measure of life satisfaction to valuation of non-market goods in order to test its performance. The life satisfaction approach to valuation is increasingly used. It calculates the marginal willingness to pay (MWTP) as the marginal rate of substitution between income and the non-market good, keeping life satisfaction constant. It often suffers from unrealistically small income coefficients and therefore large MWTP. This paper investigates two possible sources for biases in the method: person-fixed effects and adaptation of the reporting function. Fixed effects are accounted for by fixed effects or first differenced models. To avoid changes in the reporting function, the ranking measure of life satisfaction is suggested, which avoids the need for adaptation of the reporting function (Köke and Perino 2017). The paper finds that when moving from pooled OLS, to fixed effects, to first differences with the LM to the RM, the inferred marginal willingness to pay is reduced to an increasingly realistic level. This methodological examination helps better understand the merits and limitations of the approach.

The four chapters are thus very different. They can all be interpreted in an environmental economics context but are not restricted to that. The first two can more generally be interpreted for all kinds of public good provision and the life satisfaction approach to valuation can be used for a variety of non-market goods. The four chapters cover a variety of methods: game theory, laboratory experiments, panel data analysis. This is reflective of the diversity of approaches used in (environmental) economics. Since all methods have their advantages but also their shortcomings, the use of a diverse set of methods helps better understand the problems at hand.

References:

Atkinson, G., and Mourato, S. (2006). *Cost-benefit analysis and the environment: recent developments*. OECD, Paris.

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 878-894.

Baumol, W. J., and Quandt, R. E. (1964). Rules of thumb and optimally imperfect decisions. *The American economic review*, 54(2), 23-46.

Camerer, C.F. and T.H. Ho (1999). Experience-weighted attraction learning in normal form games. *Econometrica* 67, 837-874.

Clark, A. E., Frijters, P., and Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, 46(1), 95-144.

Easterlin, R. (1974). Does economic growth improve the human lot? Some empirical evidence. In P. David and M. Reder (Eds.), *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz* (pp. 98-125). New York and London: Academic Press.

Freeman III, A.M. (2003). *The Measurement of Environmental Resource Values: Theory and Methods*. Resources for the Future, Washington.

Marchiori, C., Dietz, S., and Tavoni, A. (2017). Domestic politics and the formation of international environmental agreements. *Journal of Environmental Economics and Management*, 81, 115-131.

Marrouch, W., and Chaudhuri, A. R. (2016). International Environmental Agreements: Doomed to Fail or Destined to Succeed? A Review of the Literature. *International Review of Environmental and Resource Economics*, 9(3–4), 245-319.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic theory* (Vol. 1). New York: Oxford university press.

Mitchell, R. C., and Carson, R. T. (2013). *Using surveys to value public goods: the contingent valuation method*. Rff Press.

Pigou, A.C. (1932). *The Economics of Welfare*. 4th ed., London: Macmillan and Co.

Rogelj, J., Popp, A., Calvin, K. V., Luderer, G., Emmerling, J., Gernaat, D., ... and Krey, V. (2018). Scenarios towards limiting global mean temperature increase below 1.5° C. *Nature Climate Change*, 8(4), 325.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1), 34-55.

Sanderson B., C. Tebaldi, and B. O'Neill (2016), What would it take to achieve the Paris temperature targets?, *Geophys. Res. Lett.*, 43, 7133–7142.

Welsch, H., and Ferreira, S. (2014). Environment, well-being, and experienced preference. *International Review of Environmental and Resource Economics*, 2013, 7: 205–239.

Chapter 2

Negotiating Environmental Agreements under Ratification Constraints

Co-author: Andreas Lange

Published as: Köke, Sonja and Lange, Andreas (2017). Negotiating Environmental Agreements under Ratification Constraints. *Journal of Environmental Economics and Management*, Volume 83, pp. 90-106.

2.1 Introduction

Achieving meaningful cooperation on global public good provision like greenhouse gas emission reductions remains an important challenge. The Kyoto protocol and its potential successor is a prominent showcase of both the tedious negotiation process as well as the difficulties or failure in several countries to ratify an already signed agreement.

Such problems in formulating successful international agreements hardly come as a surprise for economists. A large strand of literature on self-enforcing environmental agreements (e.g., Carraro and Siniscalco, 1993; Barrett, 1994; Hoel and Schneider, 1997; Kolstad and Ulph, 2008) derives generally pessimistic predictions for voluntary cooperation. It relies on a multi-stage model where countries first decide whether or not to join a climate coalition, and, second, coalition members maximize their collective payoff by choosing their emission reductions. This literature models countries as unitary actors and assumes that a negotiated agreement automatically enters into force. In reality, the temporal structure is more complex: most negotiated agreements must also be ratified by the legislatures of the respective countries. For example, it took 8 years for the Kyoto protocol to enter into force after it had been signed in 1997. Most prominently, the U.S. did not ratify this protocol. While ratification constraints may strategically be used to improve bargaining positions (Schelling 1960), a country's ratification may also be subject to political economy considerations *after* the agreement has been signed.

In this paper, we investigate the impact of ratification constraints and ratification uncertainty on the optimal terms of an international environmental agreement. Importantly, we consider countries as consisting of a plurality of players, those involved in the negotiation and those involved in the ratification processes. For example, ratification of international agreements may require acceptance through additional domestic agents within countries. This may involve an approving majority of a different legislative chamber (see U.S. Senate) or passing a public referendum on the ratification of international agreements. An alternative motivation for negotiators and ratifiers to differ is provided through potential changes in government. Due to the time required by negotiations and subsequent ratification processes, both the ratification process and the formulation of the unilateral policy might be in the hands of a player which potentially differs from the negotiating representative. To capture the range of these motivating examples, we consider two cases which vary *who* sets the unilateral policy: (i) the negotiating representative or (ii) the agent who is decisive for the ratification decision. Common to both, the ratification decision is in the hands of a player whose preferences may differ from those of the negotiating representative.

By relaxing the frequent assumption made in the literature on international environmental agreements that countries act as unitary actors, we connect this literature with public choice approaches. While we frame our discussion in terms of environmental agreements, our modeling framework can equally be applied to other international policy field as, e.g., discussions on disarmament, child labor issues, or contributions to international organization, where international agreements are negotiated on the provision of public goods and later have to be accepted by some domestic legislature.

The interaction between different players involved in the negotiation and ratification processes has been described by Putnam (1988). He formulates a metaphor of two tables at which the negotiator of a country sits: the national and the international one. The negotiator's task is to reconcile the interests at both tables, that is, to identify negotiation outcomes that are preferred to the status quo by all relevant players. If none of those exist, the agreement fails. Putnam suggests a two-level game consisting of a negotiation level (international table) and a ratification level (national table) and describes different aspects that influence the relation between those two levels. Most approaches of formalizing this metaphor have concentrated on a game with two negotiators and one ratifier (e.g., Kıbrıs, 2012; Iida, 1993; Iida, 1996; Kroll and Shogren, 2008; Humphreys, 2007). Negotiators first decide on the content of the agreement which is then presented to the ratifier in the ratification stage. Ratifiers either accept or reject according to their own preferences. Naturally, negotiators need to take those ratification constraints into account. Such ratification constraints can be certain (Humphreys, 2007; Hug, 2009; Kroll and Shogren, 2008) or uncertain due to asymmetric information (Iida, 1993; Iida, 1996; Kıbrıs, 2012; Milner and Rosendorff, 1997). Barrett (1998) describes the applicability of two-level games to international environmental agreements without providing a formal model. In the same spirit, Wagner (2001) notes that the theoretical literature on international environmental agreements has largely disregarded the interaction between domestic policy constraints and the formulation of international environmental agreements. A recent paper by Marchiori et al. (2017) incorporates lobbying into the coalition formation literature, but does not consider a ratification stage.

In the tradition of the public choice approaches on two-level games, we employ a model that allows for ratification failure after an agreement has been signed. We introduce this feature into a standard model of international environmental agreements. Specifically, we study the role of ratification constraints and uncertainty on two important design elements of international environmental agreements: (i) the commitment level and (ii) a minimum participation requirement. We then continue to investigate the implication for the incentives of countries to join negotiations and contribute to the literature on coalition formation by investigating the welfare implications of ratification constraints.

Minimum participation clauses are included in several international agreements. The Kyoto protocol, for example required a double trigger before entering into force: it had to be ratified by at least 55 parties to the UNFCCC, incorporating Annex I parties which accounted for at least 55% of total Annex I carbon emissions in 1990. A few theoretical studies analyze the impact of such minimum participation rules on the success of international agreements (Carraro et al., 2003; Carraro et al., 2009; Rutz, 2001; Weikard et al., 2015). Within the literature on environmental agreements, our paper is related to Baker (2005), Kolstad (2007), Ulph and Maddison (1997), and Ulph and Ulph (1996) who all incorporate uncertainty and learning about the cost and benefits of climate change into a coalition formation framework. However, these studies do neither consider ratification processes nor do they consider minimum participation rules. Black et al. (1993) investigate minimum participation rules under ratification uncertainty. By concentrating on the participation decision, they do not consider the impact of uncertainty on the commitment level which turns out to be of crucial importance in our study. In a standard IEA setting, Finus and Pintassilgo (2013) analyze the effect of negotiators' full, partial and no learning about their own benefit parameters. For this, the benefit parameter is revealed either before the game or after joining the coalition or only after the abatement has been negotiated. They also analyze different kinds of uncertainty, two of which have the same structure as we use in our two cases, consisting of different correlations of the parameter realizations across countries. The main difference, however, is that in our case, uncertainty applies to a second agent, whose preferences are not relevant for the joining and negotiation decision but for a third stage of ratification, involving only the option to accept or reject being a member of a given contract. An analysis of the set of acceptable agreements (consisting of an abatement level and a minimum participation rule) for such a ratifying agent has been done by Courtois and Haeringer (2012). In line with the literature on minimum participation, they find that it is optimal to require full participation. In this paper, we explore the robustness of this results by allowing the negotiator and the ratifier to not share the same preferences.

Our combination of public choice approaches with international environmental agreements yields several interesting and novel insights. Extending the standard coalition formation literature, we find that ratification constraints generally lead to more optimistic predictions on the success chances of international agreements. For example, ratifying agents may be less environmentally oriented, e.g., if a larger majority is required for ratifying an international agreement than it is for formulating unilateral policies. As a consequence, the commitment level in the agreement is reduced in order to increase the chances of ratification. Joining the international agreement therefore becomes less costly which increases the number of countries within the coalition. The driving force is similar to Finus and Maus (2008) who show that a partial internalization increases the stable coalition size. Importantly, this increased participation even leads to improvements in expected payoffs from the perspective of representatives. We therefore conclude that a smaller willingness to accept climate policy by domestic agents is beneficial for the success chances of international agreements.

We further show that ratification uncertainty can additionally enhance the prospects of cooperation via a similar channel: for any given set of negotiating countries, the negotiators may try to reduce the commitment level of the agreement in order to increase the chances of the agreement to be ratified and to enter into force. Simultaneously, it can be optimal to only require partial participation for an agreement to enter into force if country-specific shocks exist: while a full participation requirement stabilizes a grand coalition under certainty (Carraro et al., 2009), it is unlikely to trigger an agreement to enter into force if successful ratification is subject to risk in the respective countries. Uncertainty thereby can lower the costs of joining negotiations: (i) a smaller abatement level is required from ratifying countries. (ii) Countries may not be subject to the agreement even if they join the negotiations due to the partial minimum participation threshold. Combined this can enlarge the stable coalition and possibly lead to (expected) welfare gains. For large variance, however, negotiators return to the level that would be implemented without ratification constraints as the probability of ratification is rather insensitive to the abatement level required by the agreement. We thereby find that ratification uncertainty has a non-monotonic impact on the commitment level within negotiations, on the size of the stable coalition, and on expected welfare gains.

Throughout the paper, we focus on ex ante symmetric countries such that the agreement specifies an identical commitment level for all countries.[1] Even with this caveat, however, we believe that our results point to an interesting and important extension of the current literature. The remainder of the paper is structured as follows. We formulate the basic model in Section 2.2.1. We then turn to the non-cooperative case in Section 2.2.2 before discussing the probability of ratification in Section 2.2.3 and the optimal design in Section 2.2.4 when negotiating representatives also set unilateral policy. Section 2.2.5 discusses the incentives of countries to join the negotiation table and the welfare implications. In Section 2.2.6 we discuss a variant of the model where the ratifying agent also sets the unilateral policy. We conclude in Section 2.3.

2.2 The theoretical model

2.2.1 The agents, their preferences and the temporal structure

Let the number of countries be denoted by N. Each country must choose its abatement level of a global pollutant q_i ($i = 1, \ldots, N$). Abatement induces increasing and convex abatement costs that are assumed to be identical across countries: $C(q_i)$ ($C'(\cdot) > 0$, $C''(\cdot) \geq 0$). While starting with this general cost function, we later concentrate on quadratic specification, $C(q_i) = \frac{c}{2}q_i^2$ ($c > 0$), that has frequently been used in the literature on IEAs. Abatement creates linear benefits $B_i(Q) = b_iQ$, where $Q = \sum_i q_i$ denotes the aggregate abatement level and $b_i > 0$ is a country-specific benefit parameter. As usual, the linearity assumption on abatement benefits facilitates the analytical treatment by allowing for a separation of reaction functions be the different countries. The welfare of a country i from abatement related activities is therefore given by $b_iQ - C(q_i)$.

We use the benefit parameter b_i to incorporate different ways how the political power on (i) negotiating the agreement, (ii) the ratification of an agreement, and (iii) the formulation of unilateral policy is distributed. Specifically, we introduce two agents

[1] An extension would involve a multi-dimensional negotiation space as commitment levels for all respective parties need to be specified. For this an extension of the literature on two-level games is necessary. Kıbrıs (2012), Iida (1993) and Iida (1996) consider a single dimension, while Hug (2009) considers two-dimensional negotiation space. Humphreys (2007) claims to provide a model which can account for public goods by introducing a multi-dimensional policy-variable. However, he does not introduce the specific preference structure of public goods. Kroll and Shogren (2008) study a bilateral public good setting where own and foreign contributions are negotiated.

per country who have different roles in the political process and who have (potentially) different preferences for abatement levels. (i) We denote by b^R the benefit preference of the representative (R, in charge of the negotiation) and define it to be identical in every country. (ii) By b_i^P, we then denote the benefit preferences of the pivotal agent in the ratification process (P, whose preferences are decisive in the ratification decision) in country i.

For the most part of the paper, we assume that the representative in the negotiations can also formulate the unilateral policy, i.e. preference b^R is decisive for the choice of the abatement level in case that country i is not bound by an international agreement. In Section 2.2.6 we additionally explore a setting where the pivotal player deciding about the ratification also formulates the unilateral policy.

Preferences b_i^P may be subject to uncertainty, i.e. b_i^P may be unknown to the representative at the time when the agreement is negotiated. We assume that it is given by

$$b_i^P = \theta + \epsilon_i \tag{2.1}$$

Here, θ refers to a (possibly uncertain) component of the preference parameter that is common to all countries and ϵ_i reflects a country-specific additive shock. We denote the cumulative distribution function of θ by $H(\cdot)$ and assume that it has a mean of λb^R and variance σ_θ^2. That is, we allow for differences in expected abatement benefits between pivotal agents' and negotiators': if $\lambda > 1$ ($\lambda < 1$) the ratifiers are expected to be more (less) "green" than the negotiators.[2] Country-specific shocks ϵ_i are assumed to be independently and identically distributed with a cumulative distribution function $G(\cdot)$ with mean zero and variance σ_ϵ^2. We further assume that both θ and ϵ_i are continuous random variables, i.e. have continuous density functions H' and G'. We assume the densities to uniformly converge to zero if the respective variance gets large, i.e. $\lim_{\sigma_\theta^2 \to \infty} \sup_\theta H'(\theta) = \lim_{\sigma_\epsilon^2 \to \infty} \sup_\theta G'(\epsilon) = 0$. This assumption is, for example, satisfied for uniform and normal distributions. We finally assume that domestic and foreign representatives have identical beliefs about the preferences of country i's pivotal agent,

[2] Such a difference could result, for example, if a larger majority is required for the ratification of international agreements than for the formulation of domestic (unilateral) policies. In the USA, for example, the less green Senate (ratifier) was required to agree to the negotiated outcome. Additionally, strategic delegation to increase the bargaining power within negotiations may lead to different preferences between negotiators and ratifiers (e.g., Harstad, 2010; Buchholz et al., 2005; Siqueira, 2003).

i.e. H and G are common knowledge.[3]

We may interpret the uncertainty with respect to θ or ϵ_i as originating from the temporal features of ratification processes: for the case of public referenda to agree to the negotiated agreement, the perception of other policy fields may interfere with the decision at hand and thereby impact the ratification decision, but not necessarily unilateral policy. Potential governmental changes may induce uncertainty to the specifics of ratification decision *and* unilateral policy. Exogenous shocks may impact single countries only (e.g., through local events like flooding) or enfold more global impact on preferences (e.g., nuclear accidents as in Chernobyl or Fukushima, events affecting global economy).

In the following, we study the impact a ratification process has on the terms of an international agreement. More specifically we are interested in the impact of λ, σ_θ^2 and σ_ϵ^2 on the outcome of an agreement in terms of participation and commitment as well as in terms of expected welfare.

The temporal structure of the model comprises four stages:

(S1) Representatives decide whether to join the negotiation process or not. The number of joining representatives is denoted by $n \leq N$.

(S2) The agreement is negotiated by the n representatives. We concentrate on a simple two-component agreement that has the following structure: Firstly, it specifies a minimum number $\hat{r}$ of countries that have to ratify the agreement before the agreement enters into force. Secondly, it specifies a commitment level $\hat{q}$ which ratifying countries have to deliver in case that the agreement becomes active.

(S3) Pivotal agents of the n negotiating countries decide whether they ratify the proposed agreement. We assume that all uncertainty with respect to all b_i^P is resolved *before* the ratification processes in the countries start. Further, we assume that ratification takes place sequentially. The order is randomly determined *after* the agreement is formulated and *before* the ratification process starts. All countries are therefore identical at the time of the negotiations and fully informed about asymmetries at the time of ratification. We denote the number of ratifying countries by

[3] For a theoretical treatment of strategic use of uncertainty in case of asymmetric information between domestic and foreign representatives, see Iida (1996).

r.

(S4) If $r \geq \hat{r}$ countries ratify the agreement, it enters into force. In this case, ratifying countries are bound to abate $q_i \geq \hat{q}$, while all other countries do not face any obligations and can formulate their desired unilateral policy. In case that less than $\hat{r}$ countries ratify, no country faces any obligation as the agreement does not enter into force.

Our assumption of ex ante symmetric countries simplifies the analysis as it allows to solve the model depending on the *number* of countries who negotiate an agreement rather than considering their identity. We use this assumption to derive our main qualitative points on the impact of a ratification stage and embedded uncertainties.[4]

In the following subsections, the model is solved by backward induction. Sections 2.2.2 through 2.2.5 concentrate on the variant where the representative can also set the unilateral policy. Section 2.2.6 explores the alternative setting where the unilateral policy is set by the pivotal agent.

2.2.2 The non-cooperative policy (S4)

We first consider the unilateral decision of a country which is not bound by an agreement. The representative sets the unilateral policy to solve:

$$\max_{q_i} b^R(q_i + Q_{-i}) - C(q_i) \qquad \Leftrightarrow \qquad C'(q_i) = b^R \tag{2.2}$$

where $Q_{-i} = Q - q_i$. The linear benefits imply a dominant strategy, i.e. a non-cooperative policy that is independent of the negotiation outcome and ratification decision. The solution is denoted by q^N.

2.2.3 The ratification decision (S3)

We now derive the conditions under which countries ratify the agreement. Under our assumptions, ratification decisions occur sequentially and under full information about

[4] Extensions of our model to capture ex ante heterogeneities would require numerical solutions, but furthermore would require additional assumptions on the sequence of the ratification process.

all (other) countries' realizations of preferences b_i^P. Taking into account that countries not ratifying the agreement choose q^N (see (2.2)), a country may ratify the agreement only if its ratification is needed to achieve the minimum participation threshold $\hat{r}$. If at least $\hat{r}$ other countries ratify, a country will not do so since its ratification does not affect the contributions of any other country and therefore the additional costs to the country will be larger than its additional benefits as long as $q > q^N$. As a consequence, no more than $\hat{r}$ countries will ever ratify the agreement.[5]

Assume that a country is decisive for reaching the minimum participation threshold. It then may consider ratifying only if this (weakly) increases its payoff relative to the case when it does not ratify, i.e. when the agreement does not enter into force and all countries fall back to their non-cooperative level q^N. For $\hat{q} \geq q^N$, this holds if:

$$\begin{aligned} & b_i^P \left[\hat{r}\hat{q} + (N-\hat{r})q^N\right] - C(\hat{q}) \geq b_i^P N q^N - C(q^N) \\ \Leftrightarrow \quad & b_i^P \geq \beta^R := \frac{C(\hat{q}) - C(q^N)}{\hat{r}(\hat{q} - q^N)}. \end{aligned} \tag{2.3}$$

Here, $\beta^R = \beta^R(\hat{q}, \hat{r})$ denotes the minimum benefit parameter that makes ratification of a given agreement $(\hat{q}, \hat{r})$ worthwhile to a decisive country.

Under the assumed sequential ratification process, it thus follows that in a subgame-perfect equilibrium the agreement will be ratified by the last $\hat{r}$ countries for which condition (2.3) holds. We hereby assume that a country for which (2.3) holds with equality will choose to ratify if belonging to these $\hat{r}$ countries. The agreement thus enters into force if at least $\hat{r}$ of the n negotiating countries have $b_i^P = \theta + \epsilon_i \geq \beta^P$. The probability of entry into force is thus given by

$$\pi = \pi(\hat{q}, \hat{r}, n) = \int \left[\sum_{r=\hat{r}}^{n} \binom{n}{r} (1 - G(\beta^R - \theta))^r G(\beta^R - \theta)^{n-r}\right] dH(\theta) \tag{2.4}$$

Equation (2.4) is obtained from the binomial distribution noting that the probability of any individual country crossing the ratification threshold β^R for a specific realization of θ is given by $1 - G(\beta^R - \theta)$. In absence of individual shocks, i.e. if $\sigma_\epsilon^2 = 0$, equation

[5]Note that different results may occur if commitment levels were permitted to depend on the number of ratifying countries in a more complex way, i.e. if the commitment level would be increasing in r even for $r > \hat{r}$. Our assumption that the commitment level does not depend on the number of ratifying countries reflects the style in which the Kyoto protocol was set up.

(2.4) reduces to $\pi = 1 - H(\beta^R)$.

The following Lemma establishes how the probability of the agreement entering into force depends on the design of the agreement, i.e. on commitment level $\hat{q}$ and minimum participation $\hat{r}$:

Lemma 1 *The probability π of the agreement entering into force is (weakly) decreasing in the chosen commitment level $\hat{q}$. For a sufficiently small variance of the country-specific shocks ϵ_i, the probability (weakly) increases in the minimum participation threshold $\hat{r}$. In general, the probability π may be decreasing or increasing in $\hat{r}$.*

The proof is given in the Appendix. Intuitively, a larger commitment level increases benefits linearly while costs are convex such that a larger benefit parameter is needed to make ratification worthwhile. A larger participation threshold has two separate effects: (i) it increases the gains from the agreement for an individual country (as more countries will contribute) and thereby increases the willingness to ratify, but (ii) requires more countries to succeed in their ratification process which may lower the probability of reaching this threshold if country-specific shocks are significant. This latter effect is absent if no such country-specific shocks exist since then either all countries or no country would be able to ratify.

2.2.4 Choosing the content of the agreement (S2)

We will now use the insights about ratification behavior from the previous section to study the choice of an agreement, i.e. $\hat{q}$ and $\hat{r}$, by the representatives of the n negotiating countries.

Under our assumptions on preference parameters and the random sequence of countries' ratification processes, all negotiating countries are ex ante symmetric. We therefore can assume that the n representatives sitting at the negotiation table maximize the sum of their expected payoffs by choosing $\hat{q}$ and $\hat{r}$. This optimization is equivalent to each country's representative maximizing his expected payoff gains above the non-cooperative equilibrium. We explicitly do not talk about *socially optimal* outcomes in this paper, since it would not be clear whose preferences determine the social optimum, those of the ratifiers or those of the negotiators or some other agents. Therefore, the *optimal agreement* always denotes the outcome of the representatives' choices when maximizing their own expected utility. Within the set of agreements that apply a uniform contribution level $\hat{q}$ to all countries, the expected payoff gains for the n negotiating countries are given by:

$$\Delta W^R(\hat{q}, \hat{r}, n) \quad = \quad \pi(\hat{q}, \hat{r}, n)\hat{r}\left[nb^R(\hat{q} - q^N) - (C(\hat{q}) - C(q^N))\right] \tag{2.5}$$

When the agreement enters into force, it makes $\hat{r}$ countries change their abatement level, leading to a cost difference, but also creating additional benefits to all n negotiating countries. The gains relative to the non-cooperative solution do not depend on the random draw as the unilateral policy is set by the representatives,[6] but must be weighted with the probability π of the agreement entering into force.

Negotiating representatives maximize (2.5) by *jointly* choosing $\hat{q}$ and $\hat{r}$. We discuss the determinants of these jointly determined optimal choices below.

We first note that (2.5) immediately implies that the optimal $\hat{r}$ maximizes the expected number of ratifying countries for the optimal $\hat{q}$, i.e. optimality requires

$$\max_{\hat{r}} \pi(\hat{q}, \hat{r}, n)\hat{r} \tag{2.6}$$

With (2.4) we know that the probability is always increasing in $\hat{r}$ if country-specific shocks are of minor importance (since β^R decreasing in $\hat{r}$). It is therefore optimal to choose $\hat{r} = n$ in this case. When country-specific shocks exist, however, requiring ratification by all countries may decrease the probability of the agreement entering into force. As such, requiring only partial participation under uncertainty may outweigh the expected loss in contributions from one less cooperator.

To gain insights into the optimal choice of $\hat{q}$, we consider the marginal impact of changes in $\hat{q}$ on expected welfare gains ΔW^R for representatives:

$$\frac{\partial \Delta W^R}{\partial \hat{q}} \quad = \quad \pi\hat{r}\left[nb^R - C'(\hat{q})\right] + \frac{\partial \pi}{\partial \hat{q}}\hat{r}\left[nb^R(\hat{q} - q^N) - (C(\hat{q}) - C(q^N))\right] \tag{2.7}$$

[6] This will be different in Section 2.2.6 where ratifiers set the unilateral policy.

The first part reflects the optimal commitment level q^{*R} from the perspective of the representatives when there is no ratification uncertainty, given by $C'(\hat{q}^{*R}) = nb^R$. Then, the representatives would like to internalize all mutual benefits. However, they also need to consider the impact of $\hat{q}$ on the probability that the agreement enters into force. This is given by the second part of (2.7). Note that the term in the brackets is positive as representatives would only formulate an agreement that makes them better off than could be achieved unilaterally (with q^N). This shows that representatives may reduce the commitment level below their personally preferred q^{*R} in order to increase the probability that the agreement enters into force as $\partial\pi/\partial\hat{q} \leq 0$.

Note, however, that a compromise on the abatement level may already be necessary without uncertainty ($\sigma_\epsilon^2 = \sigma_\theta^2 = 0$) if the pivotal agents' preferences differ sufficiently from those of representatives. In order to fully assess the impact of ratification constraints, we therefore first explore the case of general λ in absence of uncertainty. That is, we allow for differences in preferences between representatives and pivotal agents. We already have seen that $\hat{r} = n$ is optimal in this case. Furthermore, from equation (2.7), it is obvious that representatives would choose $C'(\hat{q}) = nb^R$ if this level were accepted by ratifiers, i.e. if $\beta^R(q^{*R}, n) \leq \lambda b^R$. Otherwise, $\hat{q}$ must be chosen at level $\beta^R(\hat{q}, n) = \lambda b^R$, to ensure ratification. However, such a level would not be binding if $\lambda < 1/n$ as then representatives would unilaterally abate more, i.e. $\hat{q} < q^N$ ($\lim_{\hat{q}\downarrow q^N} \beta^R(\hat{q}, n) = C'(q^N)/n = b^R/n > \lambda b^R$). As a consequence, no meaningful agreement going beyond unilateral action can be concluded under certainty if $\lambda < 1/n$. In this case, uncertainty can only be beneficial as a larger abatement level might be accepted for some realizations of ratifiers' preferences. In general, the impact of ratification uncertainty on the commitment thus crucially depends on the quantitative impact on the probability of ratification in the respective countries. We can show the following proposition:

Proposition 1 *For any given number n of negotiating countries:*

(i) under certainty, the optimal agreement requires full participation, $\hat{r} = n$. The commitment level specified by the optimal agreement is given by $\hat{q} = q^{*R}$ if $b^P = \lambda b^R \geq \beta^R(q^{*R}, n)$, while otherwise $\hat{q} < q^{*R}$ results. For $\lambda < 1/n$, no meaningful agreement exists and unilateral policy, $q = q^N$ prevails.

(ii) *under ratification uncertainty, the commitment level in international agreements is chosen (weakly) below the level preferred by the n representatives ($\hat{q} \leq q^{*R}$). In absence of individual shocks ($\sigma_\epsilon^2 = 0$), full participation is optimal ($\hat{r} = n$), while partial participation ($\hat{r} < n$) can result in presence of individual shocks ($\sigma_\epsilon^2 > 0$).*

(iii) *for small variance levels σ_θ^2 or σ_ϵ^2, uncertainty reduces optimal emission levels relative to the case of certainty ($\sigma_\theta^2 = \sigma_\epsilon^2 = 0$) if $\lambda > 1/n$. Uncertainty always increases the commitment level relative to the certainty case if $\lambda \leq 1/n$. In all cases, the optimal commitment level converges to q^{*R} if $\sigma_\theta^2 + \sigma_\epsilon^2 \to \infty$.*

The proof is given in the Appendix. Proposition 1 states that the benefits from abatement will generally only be partially internalized from the perspective of the n representatives who negotiate the agreement. Intuitively, abatement requirement are reduced to increase the chances of the agreement entering into force. Proposition 1 also states that the abatement level will never exceed q^{*R}. A lower commitment level thus results even if the ratifiers have very strong preferences for abatement. This is because the ratifiers cannot renegotiate and instead are assumed to ratify any agreement that makes them better off than the status quo. The extent to which representatives may lower the commitment below their preferred level depends on the variance: for $\lambda \leq 1/n$, only uncertainty allows meaningful agreements going beyond unilateral action with some positive probability. For $\lambda > 1/n$, small variances ($\sigma_\epsilon^2 + \sigma_\theta^2$) lead to a commitment level below the level under certainty as the effect that the ratification probability increases when lowering the abatement requirement dominates. However, if the variance of pivotal agents' preference, i.e. if $\sigma_\theta^2 + \sigma_\epsilon^2$ is very large, $\partial\pi/\partial\hat{q}$ is close to zero such that condition (2.7) leads to a commitment level close to q^{*R}. Simultaneously, ratification uncertainty can impact the optimal minimum ratification threshold $\hat{r}$: while negotiators will always require full participation $\hat{r} = n$ in absence of ratification uncertainty or when only global preference realizations are uncertain, typically partial ratification will be required for individual shocks. In fact, for symmetric distributions of individual shocks, we show in the proof of proposition 1 that the optimal minimum participation level converges to $\bar{r} = \arg\max_{\hat{r}} \hat{r} \sum_{r=\hat{r}}^{n} \binom{n}{r}$ when $\sigma_\epsilon^2 \to \infty$. As such, ratification uncertainty can give a public choice motivation to formulating minimum participation clauses in international agreement that do not require full participation.

The proposition directly leads to the following corollary:

Corollary 1 *For any given number n of negotiating countries:*

(i) Considering $\hat{q}$ as a function of variance of the respective distributions, the variance of ratifiers' preferences has a non-monotonic impact on the optimal commitment level $\hat{q}$ if $\lambda > 1/n$.

(ii) If $\lambda b^R < \beta^R(q^{*R}, n)$, the optimal commitment level under ratification uncertainty with sufficiently large variance is larger than under certainty ($\sigma_\theta^2 = \sigma_\epsilon^2 = 0$).

At a first glance, the non-monotonic impact of uncertainty may appear surprising. However, the intuition is straightforward. Representatives may choose a smaller commitment level in order to increase the chances for the agreement to be ratified and to enter into force. As long as the variance is relatively small, a reduction in commitment effort can substantially increase the ratification probability which needs to be traded-off against the negative impact on payoffs that result once the agreement enters into force. If, however, the variance of pivotal agents' preferences is large, the marginal impact of such a reduction on the probability of ratification is very small. As a consequence, the representatives return to the commitment level they prefer in absence of ratification constraints.

2.2.5 Decision to join the negotiation table (S1)

We now consider the first stage of the game: the coalition formation stage where we explore the incentives of countries' representatives to join the negotiations and thereby endogenize the number of negotiating countries $n \leq N$. In the tradition of the coalition formation literature (e.g., Barrett, 1994; Kolstad and Ulph, 2008), we employ the concepts of internal and external stability (see, e.g., d'Aspremont et al., 1983; Carraro and Siniscalco, 1993; Barrett, 1994). That is, no country must have an incentive to leave the negotiation table nor should an additional country have an incentive to join. Given our ex ante symmetric framework, coalitions are fully characterized by the number of their members $n \leq N$. Anticipating the terms of the agreement for any given n, i.e. $(\hat{q}(n), \hat{r}(n))$ as characterized in the last section, *internal stability* of a coalition of size

$n \leq N$ is given if the expected payoff of a country (from the perspective of representatives) exceeds its expected payoff when leaving the negotiations, while the other $n-1$ countries reformulate the agreement:

$$\begin{aligned} & \hat{\pi}(n)\left[\hat{r}(n)b^R(\hat{q}(n)-q^N)-\frac{\hat{r}(n)}{n}(C(\hat{q}(n))-C(q^N))\right] \\ \geq \; & \hat{\pi}(n-1)\left[\hat{r}(n-1)b^R(\hat{q}(n-1)-q^N)\right]. \end{aligned} \tag{2.8}$$

A coalition of n negotiating countries is *externally stable*, if no country outside the negotiation table has an incentive to join, thereby increasing the number of negotiating countries to $n+1$ and facing potential commitments. That is, it must hold that

$$\begin{aligned} & \hat{\pi}(n+1)\left[\hat{r}(n+1)b^R(\hat{q}(n+1)-q^N)-\frac{\hat{r}(n+1)}{n+1}(C(\hat{q}(n+1))-C(q^N))\right] \\ \leq \; & \hat{\pi}(n)\left[\hat{r}(n)b^R(\hat{q}(n)-q^N)\right]. \end{aligned} \tag{2.9}$$

Even though conditions (2.8) and (2.9) may be fulfilled for different n, we concentrate on discussing the largest internally stable coalition, i.e. the largest n for which (2.8) holds. For this, external stability is given by definition. The size of the largest internally stable coalition is denoted by n^*.

In the literature that relies on a representative agent model, joining a coalition is equivalent to facing the respective commitments. In our setting, ratification may be subject to uncertainty such that a country that joins the negotiations does not necessarily face obligations: the agreement may fail to enter into force or the country may fail to ratify. So, the single country has not only to anticipate the impact its joining has on the terms of the agreement, i.e. $\hat{q}(n)$ and $\hat{r}(n)$, but also on the ratification probability $\hat{\pi}(n) = \pi(\hat{q}(n), \hat{r}(n), n)$.

The left-hand side of (2.8) reflects the potential gains above the fully non-cooperative solution if the country negotiates within a group of n countries: the agreement enters into force with probability $\hat{\pi}(n)$ and then triggers $\hat{r}(n)$ countries to increase their abatement level from q^N to $\hat{q}(n)$. Conditional on the agreement entering into force, the individual country belongs to the group of ratifying countries and faces additional costs only with probability $\hat{r}(n)/n$. The right-hand side corresponds to the gains above the fully non-cooperative solution if the country does not enter negotiations and does therefore not belong to a group of $n-1$ negotiating countries.

In order to obtain analytical results, we use the quadratic cost function $C(q_i) = \frac{c}{2}q_i^2$ which is commonly used in a wide range of the coalition formation literature. We first consider the case where no uncertainty exists ($\sigma_\theta^2 = \sigma_\epsilon^2 = 0$), before discussing potential impacts of uncertainty.

Under certainty, we know $\hat{r} = n$. For the quadratic specification of the cost function, conditions (2.2) implies

$$q^N = \frac{b^R}{c}, \tag{2.10}$$

while (2.3) can be rewritten as $b_i^P = \lambda b^R \geq \beta^R := (c/2n)(\hat{q} + q^N)$. This implies the following effective commitment level under certainty:

$$\hat{q}(n) = \begin{cases} n\frac{b^R}{c} & \text{if} \quad \lambda \geq \frac{n+1}{2n} \\ (2n\lambda - 1)\frac{b^R}{c} & \text{if} \quad \frac{1}{n} < \lambda < \frac{n+1}{2n} \\ \frac{b^R}{c} & \text{if} \quad \lambda \leq \frac{1}{n} \end{cases} \tag{2.11}$$

Using this, we can derive closed form solutions for the maximal size of a stable coalition n^* and the resulting commitment level $q^* = \hat{q}(n^*)$ as a function of λ (see Appendix). We obtain the following result:

Proposition 2 *If representatives with environmental preference b^R face ratification constraints for international agreements driven by agents with a preference $b^P = \lambda b^R$, then the following holds for quadratic abatement costs:*

(i) The size of the stable coalition is decreasing in λ. Abstracting from integer problems, it is given by

$$n^* = \begin{cases} 3 & \textit{if} \quad \lambda \geq \frac{3}{4} \\ \frac{4\lambda+3}{4\lambda-1} & \textit{if} \quad \frac{1+\sqrt{17}}{8} \leq \lambda < \frac{3}{4} \\ \frac{3}{2\lambda} + \frac{1}{\lambda}\sqrt{\frac{5}{4} - \lambda} & \textit{if} \quad \lambda < \frac{1+\sqrt{17}}{8} \end{cases} \tag{2.12}$$

(ii) The commitment level of a country within the coalition is decreasing in λ and given

by:

$$q^* = \hat{q}(n^*) = \begin{cases} n^* \frac{b^R}{c} & \textit{if} \quad \lambda \geq \frac{3}{4} \\ \frac{4\lambda+3}{4\lambda-1} \frac{b^R}{c} & \textit{if} \quad \frac{1+\sqrt{17}}{8} \leq \lambda < \frac{3}{4} \\ 2(1+\sqrt{\frac{5}{4}-\lambda}) \frac{b^R}{c} & \textit{if} \quad \lambda < \frac{1+\sqrt{17}}{8} \end{cases} \tag{2.13}$$

We already know from Proposition 1 that $\hat{q}(n)$ for a fixed n is increasing in λ. Conversely, if the ratifying pivotal agent perceives environmental damages to be smaller, the representatives may have to reduce the commitment level in order to ensure ratification. However, lowering the commitment level for any given n, makes it less costly for countries to join negotiations and to form a coalition. As a result, the number of countries in a stable coalition, n^*, is becoming larger when λ is getting smaller. For example, the solution $n^* = 3$ which is well-known in the literature and which results without ratification constraints is only stable for $\lambda \geq 3/4$. For $\lambda \downarrow 0$, n^* approaches infinity such that an arbitrarily large coalition can be stabilized. That is, for any $n \geq 3$, there exist a $\bar{\lambda} > 0$ such that for $\lambda \leq \bar{\lambda}$ a coalition of size n is stable. Interestingly, a decrease in λ leads to an increase in effective commitment level $q^* = \hat{q}(n^*)$ even though it results in decreases in $\hat{q}(n)$ for any given n. That is, the increase in the coalition size dominates. The impact of λ on stable coalition size and commitment level are illustrated in Figure 2.1.

Our result corresponds to findings by Finus and Maus (2008) who show that a partial internalization of the negotiating countries' benefits can generate larger coalition sizes. In our setting, the threat by the ratifying agent causes such a partial internalization and therefore gives a public choice motivation to partial internalization of mutual benefits of negotiators.

As a result of the larger coalition size for smaller λ, the welfare gain above the non-cooperative solution for a member of the coalition, i.e. $b_i n^*(q^* - q^N) - (C(q^*) - C(q^N))$ also increases. Note that this holds from the perspective of both the representative ($b_i = b^R$) as well as for any $b_i \geq \lambda b^R$.[7] Any players who are not inside the coalition, clearly also benefit (gains: $b_i n^*(q^* - q^N)$). We therefore again obtain a result similar to Finus and

[7]Note that in the relevant range, $b_i n^*(q^* - q^N) - (C(q^*) - C(q^N)) = b_i n^*(q^* - q^N)(1 - \lambda b^R/b_i)$ increases for falling λ as both q^* and n^* increase.

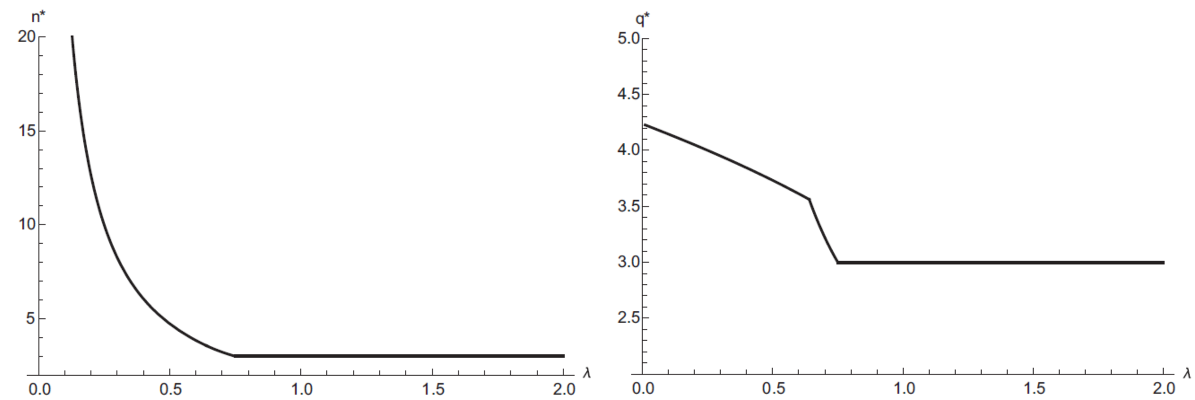

Figure 2.1: *Representatives setting unilateral policy: Maximal stable coalition* n^* *and corresponding abatement level* q^* *(relative to* b^R/c*) under certainty as a function of* λ *(*$b^P = \lambda b^R$*).*

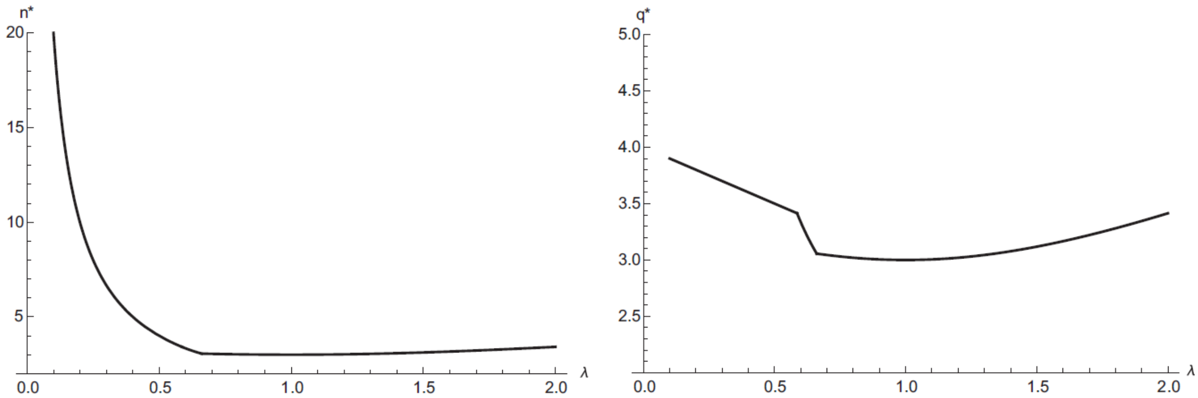

Figure 2.2: *Pivotal agents setting unilateral policy: Maximal stable coalition* n^* *and corresponding abatement level* q^* *(relative to* b^R/c*) under certainty as a function of* λ *(*$b^P = \lambda b^R$*).*

Maus (2008) based on ratification constraints: a lower environmental preference b^P of agents who are responsible for ratification decisions forces representatives to agree on a less stringent commitment level (direct effect). This partial internalization induces more countries to enter the negotiations and to ratify the agreement and thereby increases payoffs for representatives.

Proposition 3 *If representatives with environmental preference* b^R *face ratification constraints for international agreements driven by agents with a (certain) preference* $b^P = \lambda b^R$*, then anticipating the formation of the stable coalition of maximal size for quadratic abatement costs, the welfare evaluated by representatives is decreasing in* λ*.*

So far we have focused on the case without ratification uncertainty to identify the effects of ratification constraints. We now discuss how uncertainty impacts the stability of coalitions and the terms of an agreement. Specifically, we prove the following result in the Appendix:

Proposition 4 *Uncertainties within the ratification process, i.e. $\sigma_\epsilon^2 + \sigma_\theta^2 > 0$ can increase the number of countries joining the negotiation table, n^*, the effective commitment level $q^* = \hat{q}(n^*)$, the minimum participation threshold $r^* = \hat{r}(n^*)$, the expected number of ratifying countries $\hat{\pi}(n^*)\hat{r}(n^*)$, the expected total abatement $\hat{\pi}(n^*)\hat{r}(n^*)\hat{q}(n^*)$, and the expected payoff gain to representatives.*

Due to the complexity of the probability function as seen in the derivations in the last section, closed form solutions are generally impossible. To show the claim in Proposition 4, it is sufficient to rely on a proof by example.

The easiest way of seeing the potentially beneficial effects of uncertainty is to consider a uniform distribution for θ with sufficiently small variance $\sigma_\theta^2 > 0$, while keeping $\sigma_\epsilon^2 = 0$, thus also implying $\hat{r} = n$. Lets assume $H(\theta) = (\theta - \lambda b^R + d)/(2d)$ for $\theta \in [\lambda b^R - d, \lambda b^R + d]$. Here, it is obvious from $\pi = 1 - H(\beta^R)$ that $\pi = 1$ at $\beta^R = \lambda b^R - d$ and $\partial\pi/\partial\hat{q} = -(1/2d)\partial\beta^R/\partial\hat{q}$ for $\beta^R \in (\lambda b^R - d, \lambda b^R + d)$. As a consequence, the $\partial\pi/\partial\hat{q}$ converges to negative infinity in this range when d becomes small such that it is optimal to reduce $\hat{q}$ to satisfy $\beta^R = \lambda b^R - d$ for sufficiently small d. That is, for small variances under a uniform distribution, ratification is still secured and the decision corresponds to a situation under certainty where $b^P = \lambda b^R - d$. As such a small increase in variance, i.e. increasing d from zero, has the same effect as lowering the preference parameter b^P. Proposition 4 is thus a direct consequence of Proposition 2.

In order to show that this result does neither hinge upon the assumption of a uniform distribution nor on considering uncertainties only with respect to the preference component that this joint to all countries, the Appendix provides simulation results for normally distributed country-specific shocks, i.e. $\sigma_\epsilon^2 = 0 > 0$, while keeping $\sigma_\theta^2 = 0$. While here ratification is not guaranteed for positive variance, our simulations again identify an initially beneficial effect of increasing variance as stated in Proposition 4.[8]

[8]The simulations clearly rely on very specific parameter draws such that no generalization is feasible.

For large levels of uncertainty, however, negotiators return to proposing full internalization as has been shown in the previous section and, as a consequence, a small number of countries joins the negotiations. In fact, a similar level of stable coalitions ($n^* = 2$ or $n^* = 3$) results as under the well-known result without ratification uncertainty, while the agreement may fail to enter into force (see Appendix). As such, large levels of uncertainty worsen the chances of meaningful cooperation and of achieving payoff gains relative to the non-cooperative solution.

2.2.6 Alternative model specification: ratifiers setting unilateral policy

We now additionally explore a setting where the pivotal player deciding about the ratification can also formulate the unilateral policy. We here immediately concentrate on quadratic cost functions and do not allow for country-specific shocks, i.e. $\sigma_\epsilon^2 = 0$.[9] As such, the preference of the agent deciding about the unilateral policy, b^P, does not differ between countries. Similar to (2.10), the unilateral policy is thus given by

$$q^N = \frac{b^P}{c}. \tag{2.14}$$

Considering the ratification decisions for a treaty $(\hat{q}, \hat{r})$, again a country only may ratify if it is necessary for triggering the agreement to enter into force, i.e. given our sequential ratification decision no more than $\hat{r}$ countries will ratify. Again, condition (2.3) applies, but now the unilateral abatement level also depends on b^P. Given our quadratic cost specification, condition (2.3) can be rewritten as

$$b^P \hat{r} \geq \frac{c}{2}(\hat{q} + q^N) = \frac{c}{2}\hat{q} + \frac{b^P}{2},$$

It was impossible (for us) to prove that the non-monotonic impact of variance holds in general for normal distributions for all parameter settings.

[9]The extension allowing for country-specific shocks is relevant, but poses several problems: (i) as countries differ ex post and also their unilateral policy may differ, it is not decisive how many countries ratify, but who ratifies in order to calculate the gains from cooperation. (ii) Thereby also the incentives of a country to ratify depend on the parameter draws of other players: a country would not ratify if its own benefit b_i from abatement is small, but also would not ratify if other countries have high parameter draws b_j as they would unilaterally also abate a lot such that triggering the entry into force of the agreement may not be worthwhile (in the extreme case, other countries would unilaterally exceed the abatement required by the agreement). (iii) When only partial ratification is required ($\hat{r} < n$) this implies that the set of ratifying countries does not necessarily comprise the ones with largest benefits b_i. While being realistic, the formal analysis of such a setting is beyond the scope of this paper in which we desire to point out *possible* beneficial impacts of ratification constraints for prospects of international cooperation.

such that we obtain the following condition for ratification:

$$b^P \geq \beta^P := \frac{\hat{q}c}{2\hat{r}-1}, \tag{2.15}$$

where β^P denotes the minimum benefit parameter that makes ratification of a given agreement $(\hat{q}, \hat{r})$ worthwhile to a decisive country.

Under the assumed sequential ratification process, it thus follows that an equilibrium exists where exactly $\hat{r}$ countries ratify if condition (2.15) hold. The agreement thus enters into force if $\theta \geq \beta^P$, i.e. with probability $\pi^P = \pi^P(\hat{q}, \hat{r}, n) = 1 - H(\beta^P)$. Again this probability of the agreement entering into force, π^P, is (weakly) decreasing in the chosen commitment level $\hat{q}$ and increasing in $\hat{r}$.

When choosing the terms of the agreement, $\hat{q}$ and $\hat{r}$, the representatives anticipate $q^N = b^P/c = \theta/c$ and maximize expected payoff gains:

$$\Delta W^P(\hat{q}, \hat{r}, n) = \int_{\beta_P}^{\hat{q}c} \hat{r}\left[nb^R(\hat{q} - \theta/c) - (C(\hat{q}) - C(\theta/c))\right] dH(\theta) \tag{2.16}$$

The upper limit in the integral reflects that for $b^P/c = \theta/c > \hat{q}$, the unilateral policy would go beyond the required $\hat{q}$ such that – as in the case where ratification fails – the agreement gives no gains above the non-cooperative solution that fully relies on unilateral policy. When the agreement enters into force and $b^P/c < \hat{q}$, it makes $\hat{r}$ countries change their abatement level, leading to a cost difference, but also creating additional benefits to all n negotiating countries.

Identical to the other model specification, we immediately obtain an optimal minimum participation $\hat{r} = n$ as country-specific shocks are absent: the probability of entry into force as well as the gains given that the agreement is ratified are increasing in $\hat{r}$ (integrand in (2.16) positive and $\partial\beta^P/\partial\hat{r} < 0$). Noting $\hat{r} = n$ in (2.16), the marginal impact $\hat{q}$ on ΔW^P is given by:

$$\begin{aligned} \frac{\partial \Delta W^P}{\partial \hat{q}} &= \int_{\beta_P}^{\hat{q}c} n\left[nb^R - C'(\hat{q})\right] dH(\theta) \\ &\quad - \frac{\partial \beta_P}{\partial \hat{q}} n\left[nb^R(\hat{q} - \beta_P/c) - (C(\hat{q}) - C(\beta_P/c))\right] H'(\beta^P) \end{aligned} \tag{2.17}$$

which – since $\frac{\partial \beta_P}{\partial \hat{q}} > 0$ – immediately implies that the optimal $C'(\hat{q}) \leq nb^R$. That is, the abatement level is generally chosen below the abatement level which the representatives would like to implement to fully internalize all mutual benefits.

Like before, an abatement level below q^{*R} may already result under certainty. While representatives would choose $C'(\hat{q}) = nb^R$, i.e. $\hat{q} = q^{*R}$, if this level is accepted by ratifiers, they otherwise choose $\beta^P = \lambda b^R$, i.e. $\hat{q} = (2n-1)\lambda b^R/c$. Therefore, under certainty we obtain:

$$\hat{q}(n) \quad = \quad \begin{cases} \frac{nb^R}{c} & \text{if} \quad \lambda \geq \frac{n}{2(n-1)} \\ \frac{(2n-1)\lambda b^R}{c} & \text{if} \quad \frac{n}{2n-1} \leq \lambda < \frac{n}{2n-1} \end{cases} \tag{2.18}$$

with unilateral policy $q^N = b^N/c = \lambda b^R/c$ voluntarily going beyond this level if $\lambda > n$.

For the stage in which representatives decide whether or not to enter the negotiations, we proceed as before. We first consider the case under certainty, before highlighting potential effects of increasing variance.

We again consider the maximal size of a stable coalition. As before, this is given by

$$nb^R(\hat{q}(n) - q^N) - (C(\hat{q}(n)) - C(q^N)) = (n-1)b^R(\hat{q}(n-1) - q^N). \tag{2.19}$$

In the Appendix, we prove the following result for the maximal size of stable coalitions:

Proposition 5 *If representatives with environmental preference b^R face ratification constraints for international agreements driven by agents with a (certain) preference $b^P = \lambda b^R$ who also determine unilateral policy, the following holds for quadratic abatement costs::*

(i) The maximal size of a stable coalition is given by

$$n^* = \begin{cases} 2 + \sqrt{\lambda^2 - 2\lambda + 2} & \text{if} \quad \lambda \geq 0.66067 \\ \frac{5\lambda + ?\sqrt{\lambda^3 - 2\lambda^2 + 2\lambda}}{4\lambda - 1} & \text{if} \quad 2 - \sqrt{2} \leq \lambda < 0.66067 \\ \frac{2}{\lambda} & \text{if} \quad \lambda < 2 - \sqrt{2} \end{cases} \tag{2.20}$$

It is decreasing in λ for $\lambda < 1$, but increases again for $\lambda > 1$. At $\lambda = 1$, the *minimal size is given by 3 countries in the coalition, while the maximal size becomes arbitrarily large for* $\lambda \to 0$ *and* $\lambda \to \infty$.

(ii) The commitment level of a country within the coalition is given by:

$$q^* = \hat{q}(n^*) = \begin{cases} \frac{b^R}{c}(2 + \sqrt{\lambda^2 - 2\lambda + 2}) & if \quad \lambda \geq 0.66067 \\ \frac{b^R}{c}\frac{5\lambda + ?\sqrt{\lambda^3 - 2\lambda^2 + 2\lambda}}{4\lambda - 1} & if \quad 2 - \sqrt{2} \leq \lambda < 0.66067 \\ \frac{b^R}{c}(4 - \lambda) & if \quad \lambda < 2 - \sqrt{2} \end{cases} \tag{2.21}$$

It is decreasing in λ *for* $\lambda < 1$ *but increases again for* $\lambda > 1$.

Differently from the case where representatives set the unilateral policy, Proposition 5 shows that increases in the coalition size and commitment level also arise for large λ. The reason is that the unilateral policy is more demanding for $\lambda > 1$ than if representatives would set it. As such joining the agreement becomes less costly. Considering (2.19):

$$\begin{aligned} & \frac{\partial}{\partial q^N}\left[nb^R(\hat{q}(n) - q^N) - (C(\hat{q}(n)) - C(q^N)) - (n-1)b^R(\hat{q}(n-1) - q^N)\right] \\ = & \underbrace{C'(q^N)}_{\lambda b^R} - b^R = (\lambda - 1)b^R. \end{aligned}$$

That is, representatives have larger incentives to join a coalition if unilateral policy is more demanding than what they themselves would implement. This results in larger coalition sizes. Conversely, for $\lambda < 1$ the unilateral policy becomes less demanding and suboptimal from the perspective of representatives which also increases their desire to form a coalition. If λ becomes small ($\lambda < 2 - \sqrt{2}$) another effect kicks in: here representatives are forced to reduce the commitment below *their* optimal level in order to secure ratification. This in turn, additionally increases the incentives to form a coalition. Similarly to the case discussed in Section 2.2.5, the increase in the coalition size dominates the decrease in abatement for a given n, such that the effective abatement q^* increases when λ decreases for $\lambda < 1$. The relationships between λ and the stable coalition size as well as the resulting commitment level are illustrated in Figure 2.2.

Turning to welfare considerations, it is obvious that the welfare from the perspective of representatives inside the coalition, i.e. $b^R n^* q^* - C(q^*)$,

$$\frac{d(b^R n^* q^* - C(q^*))}{d\lambda} = (b^R n^* - C'(q^*))\frac{dq^*}{d\lambda} + b^R q^* \frac{dn^*}{d\lambda}$$

follows the behavior of n^* for $\lambda \geq 2-\sqrt{2}$ as $b^R n^* = C'(q^*)$ in this range. For $\lambda < 2-\sqrt{2}$, we have $dq^*/d\lambda < 0$ and $dn^*/d\lambda < 0$ as well as $b^R n^* > C'(q^*)$ such that again welfare follows the same trend as n^* and q^*.

Proposition 6 *If representatives with environmental preference b^R face ratification constraints for international agreements driven by agents with a (certain) preference $b^P = \lambda b^R$ who also can set the unilateral policy, then anticipating the formation of the stable coalition of maximal size for quadratic abatement costs, the welfare evaluated by representatives is decreasing in λ for $\lambda < 1$ and increasing for $\lambda > 1$.*

We thus obtain the result that it would be worthwhile for representatives in terms of abatement, coalition size, and welfare to (collectively) delegate the ratification decision and the power to set unilateral policy. In other words, while for $\lambda = 1$ we confirm the pessimistic results from the literature, while more optimistic results prevail for any other $\lambda \neq 1$.

We finally also discuss potential effects of uncertainty. For illustrative purposes, we again rely on uniform distribution for θ: $H(\theta) = (\theta - \lambda b^R + d)/(2d)$ for $\theta \in [\lambda b^R - d, \lambda b^R + d]$. Just as in the previous section, for sufficiently small variance, i.e. small d, it is optimal to reduce $\hat{q}$ to satisfy $\beta^P = \lambda b^R - d$ when d is sufficiently small (H' large in (2.17)). This implies again that for small variances under a uniform distribution, ratification is certain and the equilibrium properties correspond to a situation under certainty where $b^P = \lambda b^R - d$. From Propositions 3 and 5, we thus see that a small variance of the preferences of pivotal agents is beneficial if $\lambda \leq 1$, but detrimental to coalition size, abatement, and welfare if $\lambda > 1$. This crucially differs from Section 2.2.5, where a small variance of uniformly distributed θ always had a positive effect.[10]

2.3 Conclusion

In this paper we analyzed the impact of ratification constraints on the optimal design of an international environmental agreement (IEA). By introducing an agent who has veto

[10] Note that for large variance, (2.16) implies that the expected gains from the agreement approach zero: the chance of the parameter b^P being in the range $[\beta^P, \hat{q}c]$ becomes infinitesimally small. We therefore do not consider this limiting case as worthwhile investigating.

power in the ratification decision of his country, we took a first step in combining the literature on IEAs with the one on two-level games of ratification. We thereby contribute to a more realistic modeling of the temporal structure of international agreements.

We considered two variants of the model: in the first, the unilateral policy is set by the representative who also negotiates. In the second, the ratifying agent also can set unilateral policy which can be motivated by a potential change in government. Across both variants of the model, we identified channels through which the presence of a ratification stage may lead to more optimistic results than obtained by the standard literature (e.g., Barrett, 1994).

First, domestic ratifiers who have a smaller benefit parameter than their representatives trigger negotiators to formulate less demanding commitment levels. This in turn induces more countries to join negotiations and through a larger coalition size increases expected payoffs. We thereby delivered a public choice motivation to a partial internalization strategy as laid out in Finus and Maus (2008). Our findings also correspond to Marchiori et al. (2017) who explicitly allow for lobby groups in individual countries: assuming that governmental incentives are partly driven by extracting money from the respective lobby groups, they find that both a weaker environmental lobby and a stronger business lobby improve the prospects of cooperation. Qualitatively, ratification constraints in our paper have an effect similar to a weaker environmental or a stronger business lobby group, even though ratification constraints do not necessarily affect the unilateral action.

Second, when ratifying agents can also set unilateral policy, again less environmentally concerned ratifiers induce less ambitious commitment levels and thereby larger participation in coalitions. Here, an additional and qualitatively different channel exists which can also lead to more optimistic results on international cooperation: if ratifying agents are more environmentally oriented than representatives, they unilaterally would implement a stricter policy. This implies that the differential burden from joining the agreement becomes less costly such that more countries join the negotiations, leading again to welfare gains. This result shows the importance of closely investigating the political processes that impact the position of countries within international negotiations and those that impact the unilateral outside options.

Third, we showed that uncertainty with respect to the ratification stage can additionally increase the incentives of countries to join the negotiation table and thereby may lead to increases in expected payoffs. For large variance of ratifiers' preferences, however, changes in the commitment level do not substantially impact the probability of ratification such that negotiators return to implementing their own preferred commitment level in the IEA. While thereby demanding an abatement similar to the one that results without ratification constraints, the likelihood that the agreement actually enters into force is reduced. With this we establish a surprising, yet intuitive non-monotonic relationship between the prospects of international cooperation and the variance of ratification decisions. These results differ from the literature which allows for uncertainty on the benefit parameters and typically identifies a negligible effect on prospects of cooperation (e.g., full learning vs. no learning in Finus and Pintassilgo, 2013). This suggest that uncertainty within the ratification stage is crucially different from other consideration of uncertain preferences. One difference is that in our model (ex ante) uncertainty only applies to the ratifying agent, but not to the preferences of the agent who decides on joining the negotiations and on choosing the terms of the agreement. Another difference is that the ratification constraint in our model changes the set of cooperating countries: depending on the realization of the benefit parameter, a country essentially decides either to stay in the coalition or to leave. In models allowing for uncertainty regarding the benefit parameters, an exit after learning the benefit parameter is typically not modelled. It would be worthwhile exploring a dynamic game approach with uncertain parameters where the composition of cooperating countries may change over time.

Overall, our results show the importance of including a ratification stage for a better understanding of international cooperation. To simplify the derivation of our results, we concentrated on the case of (ex ante) symmetric countries and on simple agreements that only specify the commitment level and the participation threshold. Future research may better account for heterogeneities by allowing for heterogeneous preferences and also for country-specific commitments within the international agreement. Such heterogeneities would also motivate a departure from the random sequence of ratification decisions of the respective countries that underlies our investigation. This particularly applies in settings where unilateral policies can also be set by the ratifying agents: while one may intuitively think that countries with a high environmental preference should ratify first, this may in fact be suboptimal. Instead, countries with a high environmental preference would engage in substantial abatement even if the agreement fails to enter into force. The additional abatement through their ratification may therefore provide only little incentives for other countries to join (ratify) the agreement. Conversely, countries with a low realization of environmental preference would substantially increase their abatement activities when ratifying the agreement. While this is costly, it also provides strong incentives to other countries to ratify the agreement. As such, the impacts of heterogeneities with respect to both ex ante expectations and ex post realizations of country-specific shocks on the ratification process and, hence, on the formulation of international agreements are yet to be determined.

Naturally, this paper provides only a first step towards a better understanding of the importance of explicitly addressing the temporal stages of international agreements: negotiations need to be initiated, before an agreement is formulated, and later put up for ratification before it enters into force. We hope that a wide range of fruitful extensions of the coalition formation framework will be addressed by future research.

2.4 References

Baker, E., 2005. Uncertainty and learning in a strategic environment: Global climate change. *Resource and Energy Economics*, 27(1), pp. 19-40.

Barrett, S., 1994. Self-enforcing international environmental agreements. *Oxford Economic Papers*, 46 (Special Issue on Environmental Economics), pp. 878-894.

Barrett, S., 1998. On the theory and diplomacy of environmental treaty-making. *Environmental and Resource Economics*, 11(3-4), pp. 317-333.

Black, J., Levi, M.D. and De Meza, D., 1993. Creating a good atmosphere: minimum participation for tackling the 'greenhouse effect'. *Economica*, 60(239), pp. 281-293.

Buchholz, W., Haupt, A. and Peters, W., 2005. International environmental agreements and strategic voting. *Scandinavian Journal of Economics*, 107, pp. 175?95.

Carraro, C., Eyckmans, J. and Finus, M., 2006. Optimal transfers and participation decisions in international environmental agreements. *The Review of International Organization*, 1(4), pp. 379-396.

Carraro, C., Marchiori, C. and Oreffice, S., 2009. Endogenous Minimum Participation in International Environmental Treaties. *Environmental and Resource Economics*, 42(3), pp. 411-425.

Carraro, C. and Siniscalco, D., 1993. Strategies for the international protection of the environment. *Journal of Public Economics*, 52(3), pp. 309-328.

Courtois P. and Haeringer, G., 2012. Environmental cooperation: ratifying second-best agreements. *Public Choice*, 151, pp. 565?84.

d'Aspremont, C., Jacquemin, A., Gabszewicz, J.J. and Weymark, J.A., 1983. On the Stability of Collusive Price Leadership. *The Canadian Journal of Economics / Revue canadienne d'Economique*, 16(1) pp. 17-25.

Dietz, S., Marchiori, C. and Tavoni, A., 2017. Domestic policies and the formation of international environmental agreements. *Journal of Environmental Economics and Management* 81, 115-131.

Finus, M. and S. Maus, 2008. Modesty May Pay. *Journal of Public Economic Theory*, 10, pp. 801-826.

Finus, M. and P. Pintassilgo 2013.The role of uncertainty and learning for the success of international climate agreements. *Journal of Public Economic Theory*, 103, pp. 29-43.

Harstad, B., 2010. Strategic delegation and voting rules. *Journal of Public Economics*, 94, pp. 102?13.

Hoel, M. and Schneider, K., 1997. Incentives to participate in an international environmental agreement. *Environmental and Resource Economics*, 9(2), pp.153-170.

Hug, S., 2009. Models of multilateral negotiations and ratification, Working Paper.

Humphreys, M., 2007. Strategic ratification. *Public Choice*, 132(1-2), pp. 191-208.

Iida, K., 1993. When and How Do Domestic Constraints Matter?: Two-Level Games with Uncertainty. *Journal of Conflict Resolution*, 37(3), pp. 403-426.

Iida, K., 1996. Involuntary defection in two-level games. *Public Choice*, 89(3-4), pp. 283-303.

Kolstad, C.D., 2007. Systematic uncertainty in self-enforcing international environmental agreements. *Journal of Environmental Economics and Management*, 53(1), pp. 68-79.

Kolstad, C. and Ulph, A., 2008. Learning and international environmental agreements. *Climatic Change*, 89(1-2), pp. 125-141.

Kroll, S. and Shogren, J.F., 2008. Domestic politics and climate change: international public goods in two-level games. *Cambridge Review of International Affairs*, 21(4), pp. 563-583.

Kıbrıs, A., 2012. Uncertainty and ratification failure. *Public Choice*, 150(3-4), pp. 439-467.

Marchiori, C., Dietz, S., and Tavoni, A. (2017). Domestic politics and the formation of international environmental agreements. Journal of Environmental Economics and Management, 81, 115-131.

Milner, H. V. and Rosendorff, B.P., 1997. Democratic Politics and International Trade Negotiations: Elections and Divided Government As Constraints on Trade Liberalization. *Journal of Conflict Resolution*, 41(1), pp. 117-146.

Putnam, R.D., 1988. Diplomacy and domestic politics: the logic of two-level games. *International Organization*, 42(3), pp. 427-460.

Rutz, S., 2001. Minimum participation rules and the effectiveness of multilateral environmental agreements. Working Paper ETH Zurich.

Schelling, T. C. 1956. An Essay on Bargaining. *American Economic Review*, 46 (3), pp. 281-306.

Schelling, T.C., 1960. The strategy of conflict, Cambridge, MA: Harvard University Press.

Siqueira, K., 2003. International externalities, strategic interaction, and domestic politics. *Journal of Environmental Economics and Management*, 45, pp. 674?91.

Ulph, A. and Maddison, D., 1997. Uncertainty, learning and international environmental policy coordination. *Environmental and Resource Economics*, 9(4), pp. 451-466.

Ulph, A. and Ulph, D., 1996. Who Gains from Learning about Global Warming? In E. V. Ierland and K. Gorka, eds. Economics of atmospheric pollution. Heidelberg: Springer Verlag, pp. 31-67.

Wagner, U.J., 2001. The design of stable international environmental agreements: Economic theory and political economy. *Journal of Economic Surveys*, 15(3), pp. 377-411.

Weikard, H.-P., Wangler, L. and Freytag, A., 2015. Minimum participation rules with heterogeneous countries. *Environmental and Resource Economics 62*, pp. 711-727.

Chapter 3

Is Adversity a School of Wisdom? Experimental Evidence on Cooperative Protection Against Stochastic Losses

Co-authors: Andreas Nicklisch and Andreas Lange

3.1 Introduction

Protection against common stochastic losses is an apparent challenge for societies. Important contemporary examples within the environmental realm range from actions towards preventing forest fires, oil spills or nuclear accidents to preparing for extreme events triggered by climate change. Beyond this field, protecting public security against terror, aviation security, or international cooperation against pandemic diseases, for example, show similar features. The way how societies deal with such stochastic damages changes over time. Importantly, the experience of an actual damage event may trigger behavioral responses. At the individual level, the occurrence of a damage appears to increase protective actions (e.g., Meyer, 2012). At the societal level, Birkland (2006) interprets accidents, naturally occurring disasters or deliberately caused catastrophes as "focusing events" (see Kingdon, 1995) that may induce increased attention to a policy problem and thereby possibly trigger policy changes. Given that many environmental problems display stochastic occurrences of damage events, it is crucial to better understand both the behavioral drivers for individuals and groups when facing such stochastic damage as well as the behavioral reactions after experiencing such events.

Two qualitatively different channels can be distinguished through which actions may impact future damage events: first, they may impact the *size* of damages while potentially leaving the probability of an adverse event unaffected (e.g., preparing for earthquakes, adaptation for climate change). Second, they may change the *probability* that adverse events occur and thereby may fully prevent a damage event from happening (e.g., forest fire prevention, mitigation of climate change, aviation security). In this paper we investigate how the availability of these two channels affects voluntary cooperation on the protection against stochastic losses. We thereby concentrate on a voluntary cooperation setting as protective actions against probabilistic losses often require the cooperation of members of communities. We are particularly interested in the evolution of behavior over time, that is, how experiencing adverse events affects subsequent decisions. Although problems of repeated cooperation to reduce probabilistic losses are common place, there is surprisingly little known about how people actually behave when facing this type of challenges.

For these purposes, we provide experimental evidence within variants of a repeated n-person prisoner dilemma game with stochastic payoffs: subjects may (indefinitely) repeatedly choose to invest in protective actions which benefit the entire group. In the short run (one-shot), subjects have incentives to free-ride on the investments of others, while the (indefinitely) repeated interaction will allow for positive cooperation levels sustained in subgame-perfect equilibria. Specifically, we compare a setting where individual cooperation reduces the size of certain damages (*CertDam*) with settings in which cooperation either reduces the size (*DamRed*) of a damage that occurs with a given probability or reduces the probability of damages of fixed size (*ProbRed*). Expected payoffs conditional on the number of cooperators in the group are held constant across treatments.

Our experimental results show significant differences between cooperation rates in *Cert-Dam* and *DamRed* versus *ProbRed*: subjects are more likely to cooperate to reduce the probability of the all-or-nothing damage, rather than to marginally reduce the size of a certain or stochastic damage. These differences between treatments get more pronounced over time. When cooperation reduces the probability of an adverse event, cooperation remains rather stable over the series of interactions. In sharp contrast, cooperation rates decline over time when cooperation reduces the size of a certain damage or a stochastic damage that occurs with fixed probability.

In line with our motivating examples and the (German) proverb which inspired the title of our paper[1], we demonstrate that experiencing adverse events in treatments with stochastic damages is of particular importance for the dynamics of individual behavior: (i) non-cooperating players are more likely to switch to cooperation following a damage event. This tendency is particularly strong in *ProbRed*. (ii) The occurrence of damages makes it less likely for cooperating players to continue cooperation. In other words, the absence of the damage reinforces existing individual cooperation. Players therefore appear to assess their actions from an ex post perspective when deciding about future actions. As such, we demonstrate that our findings on cooperation rates and their dynamics deviate from predictions based on expected utility maximization in conventional game theoretic equilibrium concepts. Rather, the treatment differences and the dynamics of decisions are largely consistent with combinations of behavioral motives of anticipated regret (e.g., Loomes and Sugden, 1988; Zehlenberg, 1999; Filiz-Ozbay and Ozbay, 2007) and evolutionary learning dynamics which link back to notions of ex post regret (e.g., Selten and Chmura, 2008; Chmura et al., 2012).

With our findings, we thereby both identify differences in voluntary cooperation on damage reduction versus probability reduction, but also contribute to the understanding of cooperation decisions in a dynamic context. We show a differentiated behavioral response to damage events, even when their occurrence does not reveal any further information on the future likelihood of adverse events.

Our experiment relates to several different strands of theoretical and experimental litera-

[1]The German proverb reads "Aus Schaden wird man klug" which literally translates into "Failure makes smart".

ture. The incentive structure is similar to studies on policy instruments for dealing with non-point source pollution (e.g., Segerson, 1988; Miceli and Segerson, 2007; Barrett, 2011) where fines can only be put on ambient pollution levels. Here, fines are triggered based on the group rather than individual behavior.[2] Our setting also relates to recent experimental research on threshold public good games (e.g., Milinski et al., 2008; Tavoni et al., 2011): in our *ProbRed* treatment, damages are avoided if an ex ante unknown threshold of cooperating players is reached. Dannenberg et al. (2014) consider settings with commonly known horizons but unknown thresholds which differ from our study as we consider indefinitely repeated games in which cooperation could be sustained as an equilibrium.

Our paper also relates to the literature on "self-insurance" and "self-protection": following the seminal article by Ehrlich and Becker (1972), the function of protective and preventive actions as complements or substitutes for market insurance are analyzed at the individual level for purely private goods (Dionne and Eeckehoudt, 1985; Jullien et al., 1999; Briys and Schlesinger, 1990)[3] or related to some forms of externalities (Muermann and Kunreuther, 2008). Lohse et al. (2012) theoretically investigate a public good structure where actions either reduce the size or the probability of a loss, but do not explore how behavior in the two cases may differ.[4] Focusing completely on loss prevention, Keser and Montmarquette (2008) analyze individual contributions that reduce the risk of correlated public losses. They show that contributions decrease in initial loss probability and with ambiguity (in comparison to risk), while they increase with endowment. Likewise, Dickinson (1998) compares public good games with probabilistic and certain

[2]Similarly, incentives for cooperative behavior in groups have been discussed in the context of industrial organization and team production (e.g., Holmstrom, 1982; Rasmussen, 1987; Varian, 1990). This mostly theoretical literature considers typically the threat of group penalties to prevent shirking of group members in one-shot rather than repeated settings, whereas participants in our setting may choose to cooperate to avoid being potentially penalized by increased free-riding of other group members in the consecutive periods.

[3]For a setting of a single decision maker, Friesen (2012) shows by building on Becker's (1968) theory of crime that risk-averse participants are deterred more by an increase in fine than by an increase in the probability of being caught which leads to an identical expected fine. When translating the model to our setting, one would expect that cooperation is highest in the damage size reduction setting and lower in the probability reduction, exactly the opposite of our findings.

[4]Most of the papers use independent risks (uncorrelated realization of the loss), which makes sense when assuming an insurance market in the private good case and represents examples like individual risks like theft, rape and murder linked to public security or the individual benefits from cancer research. Muermann and Kunreuther (2008) have started to analyze partly correlated risks. In our setting, we are interested in fully correlated risks, which do better capture the incentive structure of our guiding examples within the realm of environmental problems.

gains from contributions and finds that risk decreases contributions. None of these paper provides a comparison of protective and preventive behavior in group settings nor considers the dynamics of behavior in repeated interactions. With our paper, we therefore enrich the existing literature not only by comparing the effectiveness of preventive vs. protective measures in voluntary interactions, but also by explicitly considering the determinants of the evolution of cooperative behavior over time in light of experiencing the damage events.

The remainder of the paper is structured as follows: section 3.2 describes the experimental setting: after describing the game in section 3.2.1, we derive predictions in section 3.2.2, before detailing the experimental design in section 3.2.3. Experimental results are presented in section 3.3 and a behavioral model consistent with the observed behavior is presented in section 3.4. Section 3.5 concludes.

3.2 Experimental Design and Predictions

3.2.1 Experimental Treatments

The starting point of our setting is a repeatedly played simultaneous move four-person prisoners' dilemma ($n = 4$). At the beginning of each period, each player is endowed with E tokens. At the end of each period, a damage of D tokens occurs with probability p and reduces the endowment of each player. Damages are fully correlated across the four players; that is, either all players or no player within a group incur the damage in a given period and damages are independent over time[5]. With their decisions, players may reduce either the size or the probability of the damage, depending on the treatment.

For this purpose, each player is asked before the damage realizes, whether she wants to cooperate or defect.[6] The action of individual i in period t is, therefore, the binary contribution choice $q_i^t \in \{0, 1\}$ with $q_i^t = 1$ being cooperative and $q_i^t = 0$ being defective. Cooperation costs the individual player c tokens. The sum of cooperators in a group and period is denoted by $Q^t = \sum_{j=1}^{n} q_j^t = q_i^t + Q_{-i}^t$. The potential damage, $D^{Treat}(Q^t)$, and

[5]For simplicity reasons in the experiment, we do not introduce the structure of a stock pollutant in this paper.

[6]In the experiment, we use neutral wording; the exact wording is "take/not take an action".

the probability of its occurrence, $p^{Treat}(Q^t)$, depend on the total cooperation level and differ between treatments ($Treat$). With this, the general payoff structure of individual i in period t for a certain treatment condition is given by

$$\pi_{i,t}(q_i^t, Q_{-i}^t, s^t) = E - cq_i^t - s^t D^{Treat}(Q^t) \tag{3.1}$$

where $s^t \in \{0, 1\}$ reflects the state of nature where the damage has ($s^t = 1$) or has not ($s^t = 0$) occurred, and c being the individual cost for cooperation.

In the experiment, we differentiate between three treatments which are calibrated to guarantee equivalence in expected damages, that is, $p^{Treat}(Q^t)D^{Treat}(Q^t)$ is equivalent for all treatments.

In the first treatment, hereafter denoted as *DamRed*, each player's cooperation leads to a reduction of the initial damage D_0 by the amount d, while the initial probability is kept constant at p_0. That is, we have $D^{DamRed}(Q^t) = D_0 - dQ^t$ and $p^{DamRed}(Q^t) \equiv p_0$. In the second treatment, hereafter denoted as *ProbRed*, cooperation leads to a reduction of the initial probability of the damage p_0 by the amount x for each cooperation decision ($p^{ProbRed}(Q^t) = p_0 - xQ^t$) while its level is fixed at $D^{ProbRed}(Q^t) \equiv D_0$. Equivalence of the expected payoffs is guaranteed by setting $dp_0 = xD_0$ which leads to expected damages in both treatments being given by $p_0(D_0 - dQ) = (p_0 - xQ)D_0$. In the third treatment, denoted as *CertDam*, expected damages occur with certainty: $D^{CertDam}(Q^t) = p_0D_0 - p_0dQ^t$ and $p^{CertDam}(Q^t) = 1$.

In order to guarantee the prisoners' dilemma structure, we assume $np_0d > c > p_0d$ and $nxD_0 > c > xD_0$. In other words, cooperation is socially beneficial in terms of expected payoffs, but does not pay off individually. Further, we assume that even full cooperation ($Q^t = n$) does not reduce the damage nor its probability to zero ($p_0 - nx > 0$, $D_0 - nd > 0$).

In our experiment, players at the end of each period get information about their own cooperation decision q_i^t, the resulting cost they incurred, and the total level of cooperation Q^t. They also get to know whether the damage event occurred or not and are informed about their individual payoff. With this information, players in *CertDam* and *DamRed* can calculate the payoff that they would have received if they had changed their own decision. This is different in *ProbRed*: for example after observing a damage event, a defecting player cannot know if the damage also would have occurred if she individually had cooperated. Conversely when no damage occurred, a cooperating player does not know if she was pivotal in preventing the damage event. In order to control for the impact of players' being informed about their marginal impact on the payoff, we introduce a fourth treatment condition *ProbRed*$^+$ which is identical with *ProbRed* in the mapping of cooperation into probability and damage, but gives players additional feedback after each period: players are informed whether the damage would have occurred if zero, one, two, three, or all four players had cooperated. Therefore, *ProbRed*$^+$ increases the subjects' awareness about their decision's marginal impact on the payoff. Table 3.1 summarizes the damage and probability functions as well as the resulting expected damages for all treatments.

Treatment	$D^{Treat}(Q^t)$	$p^{Treat}(Q^t)$	$D^{Treat}(Q^t)\ p^{Treat}(Q^t)$
CertDam	$p_0D_0 - p_0dQ^t$	1	$p_0(D_0 - dQ^t)$
DamRed	$D_0 - dQ^t$	p_0	$p_0(D_0 - dQ^t)$
ProbRed	D_0	$p_0 - xQ^t$	$(p_0 - xQ^t)D_0$
ProbRed$^+$	D_0	$p_0 - xQ^t$	$(p_0 - xQ^t)D_0$

Table 3.1: *Summary of damage size $D^{Treat}(Q^t)$ and damage probability $p^{Treat}(Q^t)$ for the respective treatments ProbRed, ProbRed$^+$, DamRed, and CertDam.*

In all treatment conditions our setting mimics infinite play. For this purpose, we apply the random stopping rule for supergames (e.g., Dal Bò and Fréchette, 2011). In our experiment, the number of supergames is not known to the players. At the beginning of each supergame, players are randomly re-matched into new groups. Each supergame consists of several periods of the game described above. A supergame has a publicly known termination probability δ after each period. That is, after each period, the supergame terminates with probability δ, and a new supergame starts in new randomly re-matched groups, whereas with probability $1 - \delta$ the supergame continues in the same group constellation. Playing in changing group compositions across supergames allows us to generate more observations per subject to better account for potential learning behavior. While players cannot predict the termination of the specific supergame, the random draws determining the lengths of the supergames are taken once and applied to all sessions and treatments.

Dal Bò and Fréchette (2011) provide evidence suggesting that the existence of a co-operative equilibrium may be a necessary (but not sufficient) condition for persistent cooperation or even cooperation levels which increase with experience. In Appendix B.1, we show that the minimum number of risk-neutral cooperating players in a cooperative subgame perfect Nash equilibrium is given by

$$Q \geq Q^{min} = \frac{c}{p_0 d(1-\delta)} + \frac{\delta}{1-\delta} \tag{3.2}$$

The proof rests on the assumption that $Q \leq n$ players follow a modified grim trigger strategy: they cooperate as long as at least $Q-1$ other players cooperate, otherwise they defect in all subsequent periods. The remaining $n-Q$ players always defect.[7]

As we want to give sustained cooperation a good chance, we choose the parameter in our experiment in a way that cooperative equilibria exist. Specifically, we set the parameters as follows: termination probability $\delta = 0.2$, initial damage probability $p_0 = 0.5$, probability reduction $x = 0.1$, initial damage size $D_0 = 20$, damage reduction $d = 4$, initial endowment $E = 25$ and cost $c = 5$. This allows for cooperative subgame perfect equilibria in which three or four risk neutral players cooperate ($Q \geq Q^{min} = 2.875$).

3.2.2 Predictions

It is obvious that the game has a subgame perfect equilibrium in which *all* players *always* defect: as in the one-shot prisoner's dilemma game, no player individually has an incentive to cooperate. The parameters were set to allow for cooperative equilibria in which $Q \geq Q^{min} = 2.875$ risk neutral players cooperate. Naturally, the equilibria for risk neutral players do not differ between treatments as all treatments are identical in the mapping of cooperation decisions into expected payoffs.[8] Differences may occur if

[7]This modified grim-trigger strategy calls for infinite punishment following a unilateral defection. It thereby introduces the highest costs possible for the deviation. As a consequence, the analysis of grim-trigger shows us the least restrictive condition for cooperative equilibria to exist. Naturally, the multiplicity of equilibria may motivate further discussions on equilibrium selection. While not being the focus of the paper, we note that the equilibrium which supports $Q = 4$ is not "renegotiation proof" as – following the defection of one player – the remaining three players collectively would not have an incentive to follow through with the punishment as it lowers their payoffs, while the cooperation of these three players can still be supported by the modified grim trigger strategies.

[8]Note, however, that the stochastic damage treatments could allow for additional strategies where players condition their actions or changes of actions on the occurrence of a damage event. However, there is no intuitive way to select between different possible equilibria.

subjects are risk-averse or risk-loving.

Intuitively, one may expect levels of cooperation to be higher in *DamRed* than in *ProbRed* for risk-averse subjects: while the expected utility of a player for $Q = 0$ is identical in the two treatments ($p = p_0$, $D = D_0$), it is larger in *DamRed* than in *ProbRed* and *ProbRed*$^+$ if $Q > 0$.[9] This suggests that the willingness to (collectively) cooperate among risk-averse players is higher in *DamRed* than in *ProbRed* and *ProbRed*$^+$. For risk-lovers, the opposite relationship would hold.

In order to study the stability of cooperation under different risk attitudes, we again concentrate on modified grim trigger strategies that have been introduced above. We model CARA risk-attitudes by $\sum_t \mathbb{E}[u_i(\pi_i^t)]$ where $u_i(\pi) = \pi^{1-\sigma}/(1-\sigma)$. Figure 3.1 depicts the minimal cooperation level Q^{min} needed in the respective treatments to make cooperation attractive for a subject of a given level of risk aversion σ.[10] We see that all of the curves collapse for risk-neutral players ($\sigma = 0$) for which we again obtain $Q^{min} = 2.875$. It can be seen that for risk-averse decision-makers ($\sigma > 0$) the threshold Q^{min} is lowest for *DamRed*, while for risk-lovers ($\sigma < 0$) *CertDam* mostly leads to the smallest Q^{min}.

For *DamRed*, the threshold Q^{min} is decreasing in σ. More risk averse players are thus willing to be part of a smaller subset of Q cooperating players, while very risk seeking players are not even willing to cooperate if everyone else cooperates. That is, more risk-averse players are more likely to cooperate. For *CertDam*, we observe that cooperation is rather insensitive to risk attitudes. For *ProbRed*, we obtain a U-shaped relationship between Q^{min} and σ in Figure 3.1. Intuitively, neither highly risk-averse nor highly risk-loving subjects are predicted to cooperate: if a subject is extremely risk-averse, she concentrates on the minimum payoff. As cooperation can not prevent the damage for sure, this minimum payoff is larger if the subject defects as then cooperation costs are saved. Conversely, an extremely risk-loving subject essentially only counts with the maximum payoff (i.e., the damage not occurring), and again has no incentives to spend the costs of cooperation. As such, only players with intermediate levels of risk aversion

[9] This can be seen from $(p_0 - xQ)u_i(E - D_0 - cq_i) + (1 - p_0 + xQ)u_i(E - cq_i) \geq p_0 u_i(E - D_0 + dQ - cq_i) + (1 - p_0)u_i(E - cq_i)$ which holds due to the concavity of $u_i(\cdot)$ for risk-averse players. As such that (collective) cooperation is more beneficial. This argument follows an analysis of individual decision making by Friesen (2012).

[10] The conditions that are used for the simulations are given in Appendix B.1.

may cooperate for any given threshold level Q^{min}. Note that for $Q^{min} = 2$ this set is empty, while for $Q^{min} = 3$ it is fully contained in the set of potentially cooperating players under *CertDam*. As such, we predict cooperation rates to be lower in *ProbRed* than in *CertDam* if players behave as expected utility maximizers. Cooperative equilibria with 2 players cooperating may only exist for *DamRed*.

Prediction 1 ***(Equilibrium Prediction)***

(a) *The likelihood to cooperate increases with players' degree of risk aversion in DamRed, it is relatively insensitive to risk aversion in CertDam. In ProbRed, only players with intermediate levels of risk aversion may choose to cooperate.*

(b) *Sustained cooperation of two players is most likely in DamRed. Cooperation of three or four players is most likely in CertDam.*

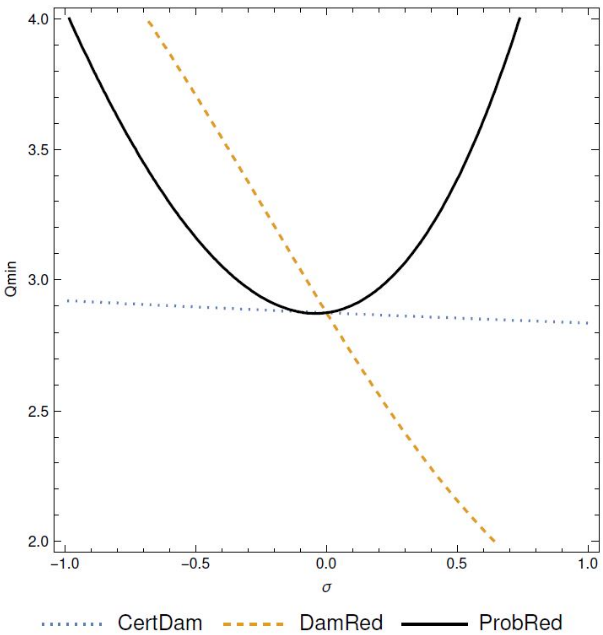

Figure 3.1: *Minimal cooperation level Q^{min} required to stabilize cooperation as a function of risk aversion σ for CRRA preferences ($u(\pi) = \pi^{1-\sigma}/(1-\sigma)$). Parameters as used in experiment ($\delta = 0.2$, $p_0 = 0.5$, $x = 0.1$, $D_0 = 20$, $d = 4$, $E = 25$, $c = 5$).*

The former discussion relied on subgame perfect equilibria where individuals' strategies condition their actions in each period only on group members' behavior in the previous periods. However, even if they additionally conditioned on the presence of a damage event, the conclusions for the minimal number of cooperating players would not change. Furthermore, no clear prediction based on subgame perfect equilibria can be made on how the occurrence of a damage affects future actions. However, while for one-shot or finite interactions, convincing evidence exists that standard (selfish) preferences as used above cannot fully describe individual behavior in dilemma situations, the indefinitely repeated game structure allows for cooperative equilibria. As such, it is an open question how well these predictions perform.

We therefore take an explorative approach when presenting our results in section 3.3 and first contrast them with the predictions based on subgame perfect Nash equilibria as derived above. In section 3.4, we then will present a behavioral model will prove better able to accommodate our results.

3.2.3 Experimental Procedure

In total, we ran 12 experimental sessions between January and March 2014 at the Experimental Laboratory of the School of Business, Economics and Social Sciences at the University of Hamburg. Three sessions were conducted for each of the treatment conditions that we described in Section 3.2.1. A total of 280 students from the University of Hamburg participated in the experiment, with a maximum of 24 and a minimum of 16 subjects per session. Median age was 24 years, 53% were female participants.

We applied the same sequence of periods and supergames across all sessions and treatments which we randomly determined by the computer prior to the first experimental session. Overall, all participants played seven supergames (participants did not know the total number of supergames beforehand), the supergames consisted of 5, 3, 7, 4, 7, 3 and 5 periods, respectively. We organized the rematching at the end of each supergame such that two new groups were randomly formed from a matching unit of 8 participants which remained constant for the entire duration of the session. This gave us 9 independent observations in *ProbRed*, *DamRed*, and *CertDam*, as well as 8 independent observations in *ProbRed*$^{+}$.

After the main experiment, we assessed participants' risk preferences following Eckel and Grossman (2008) and Dave et al. (2010) with an average payoff of 38 Cent (minimum

2 Cent, maximum 70 Cent), before adding some brief questions regarding the socio-demographic characteristics of our participants (e.g., gender, age, and years of study).

During the experiment, participants played for Taler, at the end of the experiment, the sum of the payoffs in all rounds were converted into Euros at an exchange rate of 1 Taler for 1 Euro-Cent and paid out privately. Subjects earned an average of 10.50 Euro in the repeated prisoners' dilemma part, with a maximum of 12.70 Euro and a minimum of 8.25 Euro. Each session lasted for about 60 minutes. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007), recruitment took place with hroot (Bock et al., 2014). The instructions (translated from German to English) can be found in the Appendix B.2; the decision screen (including instructions) for the risk assessment task is shown in Figure B.2 at the end of the Appendix B.2.

3.3 Results

We structure our discussion of the results by first considering average treatment differences, before explicitly exploring the individual adaptation dynamics after damage events.

3.3.1 Average Treatment Differences

	all periods	first periods	last periods
(1) *CertDam*	.26	.43	.20
(2) *DamRed*	.38	.47	.36
(3) *ProbRed*	.59	.66	.57
(4) *ProbRed*$^+$	.54	.60	.51
tests	(3),(4)$>^{**}$(2)	(3)$>^{**}$(2)	(3)$>^{**}$(2)
		(4)$>^{*}$(2)	(4)$>^{*}$(2)
	(3),(4)$>^{***}$(1)	(3),(4)$>^{***}$(1)	(3),(4)$>^{***}$(1)

Table 3.2: *Average cooperation rates by treatments over the entire experiment (left panel), over the first periods of all supergames (middle panel), and over the last periods of all supergames (right panel), tests refer to two-sided Wilcoxon Mann-Whitney rank sum tests,* *** *indicates significance at a* $p < 0.01$ *level,* ** *at a* $p < 0.05$ *level and* * *at a* $p < 0.1$ *level.*

Figure 3.2 shows the mean cooperation rates per period and treatment. Table 3.2 sum-

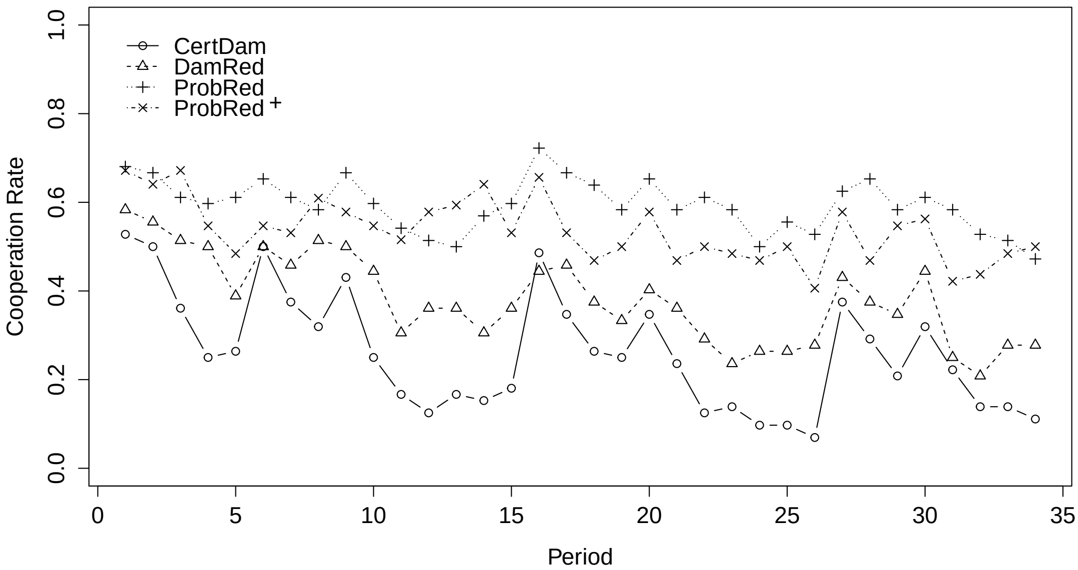

Figure 3.2: *Mean cooperation frequency per period by treatment.*

marizes the average cooperation rates across all periods as well as for the first and last periods of the supergames. It is immediately seen that cooperation rates in *ProbRed* and *ProbRed*$^+$ are substantially higher than in *DamRed* and *CertDam*. Overall, cooperation rates across all periods are 59% in *ProbRed*, 54% in *ProbRed*$^+$, 38% in *DamRed*, and 26% in *CertDam*. More specifically, cooperation rates in *ProbRed* and *ProbRed*$^+$ are significantly larger than in *CertDam* ($p < 0.01$)[11] and *DamRed* ($p < 0.05$). No significant difference exists between *ProbRed* and *ProbRed*$^+$. These results are largely robust to concentrating on the first or the last periods of supergames as is displayed in Table 3.2. We therefore formulate our first result:

Result 2 *Cooperation rates are larger when cooperation affects the probability of a damage event (ProbRed and ProbRed*$^+$*) rather than affecting the size of a stochastic damage (DamRed) or when it leads to a certain damage reduction (CertDam).*

Result 2 is not consistent with our predictions based on SPNE predictions as derived for expected utility maximizers. In fact, we find no significant impact of risk aversion on cooperation decisions in any of the treatments: Table 3.3 reports results from both

[11]Throughout the paper and unless specified otherwise, statistical significance is assessed by two-sided Wilcoxon Mann-Whitney rank sum tests relying on matching unit averages.

a regression analyzing decisions in the first period of the first supergame (left panel) as well as a random effect regression (errors are clustered at the matching group level) analyzing decisions in all periods and all supergames. Risk attitudes are measured by the lottery choice in the second part of the experiment. The variable *risk*, ranges from one to six, such that the lottery choice with larger numbers indicates more risk tolerance.[12] Estimated coefficients suggest that the behavior is not driven by the individual's risk aversion.

	dependent variable: q_i^t	
	only first period	*all periods*
DamRed	.251 (.206)	.11 (.148)
ProbRed	.256 (.19)	.278 (.169)
ProbRed$^+$	.272 (.189)	.191 (.137)
risk	.04 (.033)	−.022 (.017)
risk × *DamRed*	−.053 (.048)	.006 (.03)
risk × *ProbRed*	−.029 (.048)	.016 (.039)
risk × *ProbRed*$^+$	−.036 (.047)	.024 (.029)
constant	.386*** (.13)	.339*** (.092)
obs	280	9520
n	280	280
F-test/Wald-Chi^2-test	.92	68***

Table 3.3: *Left panel: linear regression of cooperation behavior in the first period, right panel: random effects regression of cooperation behavior in all periods of the experiment; coefficients are reported along standard errors in parenthesis (errors are clustered at the matching group level); *** indicates significance at a $p < 0.01$ level, ** at a $p < 0.05$ level and * at a $p < 0.1$ level. obs reports the number of observations while n reports the number of subjects; models' fitness are assessed by F-test and Wald-Chi^2-tests.*

The treatment differences reported in Result 2 qualitatively occur already in the very first period of the experiment: while 68% cooperate in *ProbRed*, 67% in *ProbRed*$^+$, only 58% cooperate in *DamRed* and 53% in *CertDam*. At the individual level (since each subject provides an independent observation in the first period of the first supergame), the differences between *CertDam* and *ProbRed* ($p = 0.06$) and *ProbRed*$^+$ ($p = 0.09$) are weakly significant based on two-sided Wilcoxon Mann-Whitney rank sum tests.[13]

The treatment differences are further strengthened over time as can be seen in Figure 3.3 which shows cooperation rates in the first period of the respective supergames. In

[12] Alternative specification which code risk attitudes as binary variable do not change any of the results.
[13] No significant differences occur when controlling for risk aversion (see Table 3.3).

contrast to our prediction 1(b), we find a negative trend of cooperation rates in the first periods of supergames in *DamRed* and *CertDam* (both $p = 0.05$, based on Cuzick's non-parametric test for trends), while the negative trend is not significant for the probability reduction treatments ($p = 0.19$ and $p = 0.13$, respectively).

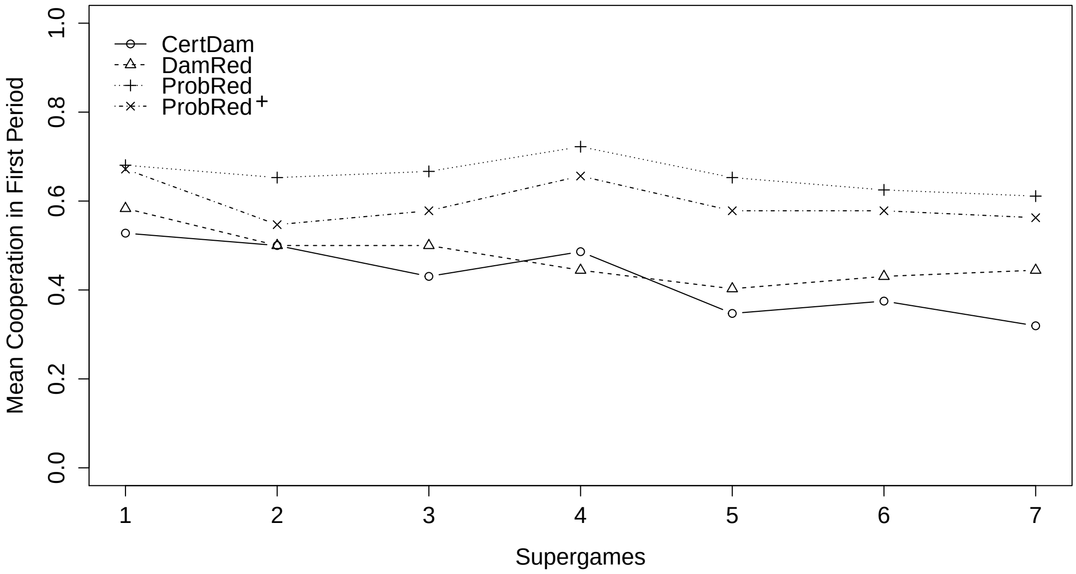

Figure 3.3: *Mean cooperation in the first period of all supergames across treatment conditions*

Table 3.4 reports further evidence for the cooperation trends based on a random-effects regression of the individual cooperation decision on the supergame (*supergame*, ranging form 1 to 7) and the period within a supergame (*period in supergame*, ranging from 1 to 7) as well as on dummies for the treatments and the corresponding interaction terms. We find negative time trends across supergames in *DamRed* and *CertDam*, and a significantly less negative trend in *ProbRed*$^{+}$, while there is no significant trend in *ProbRed*.[14] Negative time trends also occur within supergames in all treatments.[15] The downward trend within supergames is largest in *CertDam*, significantly smaller in both *DamRed* and *ProbRed* and weakest in *ProbRed*$^{+}$.

Result 3 *Cooperation rates follow different time trends: the downward trend is strongest*

[14] According to F-Tests, testing that *superg* $\times$ *treatment* + *supergame* is statistically different from zero for all treatments at $p < 0.05$ except *ProbRed* ($p = 0.146$).

[15] According to F-Tests, testing that *period in supergame* $\times$ *treatment* + *period in supergame* is statistically different from zero for all treatments at $p < 0.05$.

	dependent variable: q_i^t
DamRed	.06 (.071)
ProbRed	.147* (.078)
ProbRed$^+$	.099 (.08)
supergame	−.031*** (.006)
supergame × *DamRed*	−.004 (.01)
supergame × *ProbRed*	.02** (.009)
supergame × *ProbRed*$^+$	.01 (.008)
period in supergame	−.055*** (.007)
period in supergame × *DamRed*	.025*** (.009)
period in supergame × *ProbRed*	.033*** (.01)
period in supergame × *ProbRed*$^+$	.043*** (.008)
constant	.558*** (.057)
obs	9520
n	280
Wald-Chi^2-test	173***

Table 3.4: *Random-effects linear regression of time trends for individual cooperation decision q_i^t; coefficients are reported along standard errors in parenthesis (errors are clustered at the matching group level); *** indicates significance at a $p < 0.01$ level, ** at a $p < 0.05$ level and * at a $p < 0.1$ level. obs reports the number of observation while n report the number of subjects; model's fitness is assessed by a Wald-Chi^2-test.*

in CertDam, less strong in DamRed and least in ProbRed$^+$ and ProbRed, both within and across supergames.

Again, the slower average learning of defection in *ProbRed* and *ProbRed*$^+$ than in *DamRed* and *CertDam* is not in line with Prediction 1(b) which was derived under the assumption that individuals only condition their behavior on observed cooperation decisions by others.

3.3.2 Dynamics of Individual Behavior

To gain further insights into the different time trends, we now investigate determinants of behavioral adjustments at the individual level. Given Prediction 1, we expect no systematic time trend within supergames. However, empirical and anecdotal evidence (e.g., Meyer, 2012; Birkland, 2006) suggests that individuals may condition their choice on the realization of damage events.

In a first step, we consider the conditional frequencies of $q_i^{t+1} = 1$ given q_i^t and the occurrence of the damage s^t. Table 3.5 summarizes the frequencies by treatment conditions as well as the significant differences based on nonparametric Mann-Whitney tests.

	damage in t: $s^t = 1$		no damage in t: : $s^t = 0$	
	$q_i^t = 0$	$q_i^t = 1$	$q_i^t = 0$	$q_i^t = 1$
(1) *CertDam*	.13	.60	-	-
(2) *DamRed*	.16	.65	.17	.76
(3) *ProbRed*	.22	.79	.13	.90
(4) *ProbRed*$^+$	.23	.68	.31	.74
tests	(3),(4)$>^{**}$(1) (4)$>^{**}$(2)	(3)$>^{***}$(1),(2)	(4)$>^{**}$(2) (4)$>^{***}$(3)	(3)$>^{***}$(2),(3)

Table 3.5: *Mean q_i^{t+1} given q_i^t and the occurrence of the damage s^t; tests refer to two-sided Wilcoxon Mann-Whitney rank sum tests, *** indicates significance at a $p < 0.01$ level, ** at a $p < 0.05$ level and * at a $p < 0.1$ level.*

Overall, it seems that the effect of probability reduction on cooperation is two-fold: it leads to more stable cooperation of those players who already cooperate (their frequency to choose $q_i^{t+1} = 1$ is about 15% higher in *ProbRed*), and it induces non-cooperating players to cooperate after a damage event occurred (the frequency to choose $q_i^{t+1} = 1$ is 6-10% higher in *ProbRed* and *ProbRed*$^+$). That is to say, the "all-or-nothing" damage of *ProbRed* and *ProbRed*$^+$ prevents players from choosing defection and additionally leads more defecting players to switch to cooperation.[16]

For a detailed analysis of individual learning in our game, we estimate a series of Arellano-Bond panel regressions, for each treatment condition separately.[17] This allows us to analyze endogenous regressors (see Arellano and Bond, 1991): the dependent variable is q_i^{t+1} (i.e., the decision whether to cooperate or defect in the consecutive period). As explanatory variables, we use Q_{-i}^t (i.e., the number of cooperators except i in the current period), the occurrence of the damage in t (i.e., we compute a dummy variable s^t which is one if the damage occurred in t and zero otherwise; omitted in *CertDam*), q_i^t (i.e., the decision whether to cooperate or defect in the current period), and interaction terms

[16]Notice that there is also a surprising effect in *ProbRed*$^+$ for non-cooperators if the damage did not occur: here, the frequency of cooperation in $t+1$ is 14-18% higher than in the other treatment conditions. While we are lacking a clear explanation, this finding may be driven by the additional information that these players receive relative to *ProbRed*. We control for this effect in our following analysis.

[17]Arellano-Bond is typically applied to continuous rather than discrete dependent variables. However, we are not aware of a fully consistent method which can both incorporate the lagged contribution variable as well as control for the interdependencies at the individual and matching unit level. Our results are, however, robust to alternative specifications like random effects probit model, or OLS regressions with individually clustered errors.

$q_i^t \times s^t$, as well as $Q_{-i}^t \times s^t$. Furthermore, we control for the beginning of a new supergame (i.e., the dummy variable *newsupergame* is one in the first period of a new supergame and zero otherwise).

To access the additional information provided in *ProbRed*$^+$, we additionally introduce a variable measuring the number of cooperators exceeding the necessary number to avoid the realization of the damage. That is, the variable Δ*cooperator* computes the difference between the actual players cooperating and the cooperators required by nature for the absence of the damage. Δ*cooperator* is zero if the number of cooperators just coincides with the number required to avoid the damage, it is negative if too few players cooperate to prevent the damage and is positive if even a smaller number of cooperators were necessary to prevent the damage. Hence, we test whether players coordinate their cooperation onto the sufficient number of cooperators in the previous period. Estimations for coefficients along standard errors in parenthesis are reported in Table 3.6.

The estimation results in Table 3.6 confirm our previous findings in Table 3.5. They indicate that cooperation is highly path dependent in all treatment conditions: if a player cooperates in period t, it is very likely that she cooperates in period $t+1$ as well (significant positive marginal effect of q_i^t).

For all treatments, we also find evidence that subjects reciprocate on others' cooperation (significant positive coefficients for Q_{-i}^t). However, experiencing a damage event triggers also behavioral changes: non-cooperators are more likely to switch to cooperation following a damage event in both *ProbRed* and *DamRed* (significant positive coefficients for s^t). This effect seems to be dominated by the coordination of cooperation in *ProbRed*$^+$. We further find significant negative coefficients for the interaction $q_i^t \times s^t$: a damage event typically reduces the likelihood for cooperators to continue cooperation (or at least does lead to significantly smaller increases than found for defectors). In addition, the significant negative coefficient of Δ*cooperator* suggests that players condition their cooperativeness on the number of cooperators needed to prevent the damages in the previous period: if there are more (less) players than needed to avoid the damage, the likelihood to cooperate decreases (increases).

Result 4 *In all treatment conditions with stochastic payoffs, the non-occurrence of the*

	dependent variable: q_i^{t+1}				
	CertDam	*DamRed*	*ProbRed*	*ProbRed*$^+$	*ProbRed*$^+$
q_i^t	.219*** (.022)	.368*** (.03)	.29*** (.027)	.066** (.028)	.094*** (.029)
Q_{-i}^t	.083*** (.012)	.043** (.016)	.028** (.013)	.06*** (.016)	.096*** (.018)
s^t		.106*** (.03)	.113** (.036)	.052 (.043)	−.086 (.055)
$q_i^t \times s^t$		−.395*** (.038)	−.26*** (.032)	−.183*** (.042)	−.190*** (.042)
$Q_{-i}^t \times s^t$		.017 (.02)	.017 (.018)	.006 (.023)	−.001 (.023)
newsupergame	−.034* (.020)	−.005 (.020)	.011 (.016)	−.06*** (.022)	−.063*** (.021)
Δ*cooperator*					−.041*** (.011)
const	.130*** (.012)	.194*** (.023)	.360*** (.029)	.406*** (.033)	.412*** (.033)
obs	2232	2232	2232	1984	1984
n	72	72	72	64	64
Wald-Chi^2-test	191***	183***	9841***	50***	63***

Table 3.6: *Estimation results for an Arellano-Bond panel regressions with dependent variable q_i^{t+1}; coefficients are reported along standard errors in parenthesis; *** indicates significance at a $p < 0.01$ level, ** at a $p < 0.05$ level and * at a $p < 0.1$ level. Standard errors are clustered at the matching group level. obs reports the number of observation while n reports the number of subjects; models' fitness are assessed by Wald-Chi^2-tests.*

damage reinforces existing cooperation while the occurrence of a damage stimulates a strategy switch of players from defection to cooperation and from cooperation to defection.

Result 4 showcases the importance of experienced damage events for behavioral adjustments: players condition their behavior (partly) on the occurrence of the random event. Their current strategies are reinforced after experiencing the absence of the damage.

3.4 Explaining the Behavioral Dynamics

Our results both on average behavior as well as overall and individual time trends stand in stark contrast to the predictions derived for subgame-perfect Nash equilibria. The reciprocal behavior points towards a behavioral motivation that has also been identified in many other games. However, conditioning behavior solely on other group members' actions, falls short of explaining the dynamics of decisions. Instead, the occurrence of the damage event itself has predictive power for behavioral changes. This is surprising from the perspective of a forward looking individual as random draws (conditional on cooperation decisions) are independent across periods.

There is, however, substantial evidence that players often assess the success of their previously chosen action ex post and adapt the strategy accordingly, i.e. players apply *ex post rationality* (cf. Selten and Stoecker, 1986). Several approaches follow this general evolutionary idea that actions that (would) have been successful in the past will be reinforced and dissatisfying actions will be weakened: reinforcement learning (e.g., Roth and Erev, 1995; Erev and Roth, 1998), experience-weighted attraction learning (e.g., Camerer and Ho, 1999; Ho et al., 2008), and impulse balance learning (e.g., Selten and Chmura, 2008; Chmura et al., 2012), to state the most prominent examples. To show in which way ex post rationality can explain our treatment differences, we will concentrate on impulse balance learning.

Formally, there is an initial attraction $A_{i,0}(q)$ of player i to play action $q \in \{0, 1\}$. Selten and Chmura (2008) assume that the attraction of action q evolves according to

$$A_{i,t+1}(q) = A_{i,t}(q) + \max\{0, \pi_{i,t}(q, Q^t_{-i}, s^t(q)) - \pi_{i,t}(1-q, Q^t_{-i}, s^t(1-q))\}, \qquad (3.3)$$

where $s^t(q)$ $(s^t(1-q))$ denotes the state of the damage event if action q $(1-q)$ was chosen. That is, an action is reinforced if it would have or has been the better strategy. The probability of action q being played in period $t+1$ is simply its attraction relative to the sum of the attractions of both actions available to individual i:

$$P_{i,t+1}(q) = \frac{A_{i,t}(q)}{A_{i,t}(0) + A_{i,t}(1)}. \tag{3.4}$$

Note that the extent of reinforcement in (3.3) equals the payoff difference between both actions. In *CertDam* and *DamRed*, defection is a dominant strategy: the payoff difference to cooperation is 3 in *CertDam*, and 1 if a damage occurs or 5 if it does not (i.e., in expectations 3) in *DamRed*. As such, only defection is reinforced in *CertDam* and *DamRed* (on average by 3 per period). Impulse-balance learning would therefore explain cooperation to be phased out over time at a similar rate in *CertDam* and *DamRed*:[18]

$$\mathbb{E}[P_{i,t+1}(1)] = \frac{A_{i,0}(1)}{A_{i,0}(0) + 3t + A_{i,0}(1)} \rightarrow_{t\rightarrow\infty} 0. \tag{3.5}$$

In *ProbRed* and *ProbRed*$^+$, however, player i may be pivotal in triggering the damage event. This happens with 10% probability[19] and would lead to cooperation being the superior action (payoff difference $D_0 - c = 20 - 5 = 15$). With 90% probability, the player cannot affect the damage event in which case defection ex post would have been the better choice (payoff difference $c = 5$). If players behave according to the correct probability of having been pivotal, cooperation is therefore reinforced in 10% of the periods with a payoff difference of 15, while in the remaining 90% of the cases defection is reinforced by 5. Per period within a supergame, in expectation $A_{i,t}(0)$ grows by 4.5 and $A_{i,t}(1)$ by 1.5, such that the expected probability of cooperation after t periods is

$$\mathbb{E}[P_{i,t+1}(1)] = \frac{A_{i,0}(1) + 1.5t}{A_{i,0}(0) + A_{i,0}(1) + 6t} \rightarrow_{t\rightarrow\infty} 0.25. \tag{3.6}$$

[18]The alternative learning dynamics (e.g., Erev and Roth, 1995) lead to similar insights for *CertDam* and *DamRed* as Beggs (2005) shows that weakly dominated strategies (as cooperation in our case) are phased out over time. In *ProbRed* and *ProbRed*$^+$, however, the likelihood of cooperation would also be predicted to converge to zero as defection is still dominant in expected payoff terms.

[19]Imagine the damage occurs when a random draw between 0 and 1 is smaller than $p(Q)$. The impact of one more individual cooperating on $p(Q)$ is -0.1. The random draw determining the damage occurrence lies in this impact range with probability of 0.1. So the probability of the individual being pivotal in preventing the damage is 10%.

Thus, impulse-balance learning would explain that cooperation is not phased out in the long run. Instead the likelihood of cooperation converges towards 25%. Note if a damage event occurs, cooperating players have (ex post) obviously chosen the wrong action, while they may have been right if no damage event occurs. Cooperating players should thus be more likely to switch towards defection after a damage event than when no damage occurred. Conversely, defecting players may have been wrong in their choice if a damage occurs, while the absence of the damage event proves that their defection was right. Defectors are therefore predicted to be more likely to cooperate following a damage event than when no damage occurred, which is exactly what we identified in our results. The described behavioral dynamics also explains that cooperators in *ProbRed* and *ProbRed*$^+$ regret their action when a damage has occurred and are more likely to switch towards defection. It does not explain, however, the same behavioral change in *DamRed* where again cooperators are more likely to switch towards defection following a damage event. Overall, however, impulse balance learning appears to be able to explain many of our results: the adaptation of individual behavior crucially depends on the occurrence of the damage, but only indirectly on the strategies of other players (as they influence the occurrence of the damage).

While the described model of ex post rationality can explain *changes* in behavior over time, it is silent about the initial strategies (i.e. the initial attraction levels $A_i^0(q)$). The same logic of looking back to identify ex post regret, however, can also be applied ex ante, that is, when players anticipate regret (e.g., Loomes and Sugden, 1982; Zehlenberg, 1999; Filiz-Ozbay and Ozbay, 2007). Again, defection is a dominant strategy in *CertDam* and in *DamRed*, while players in the probability reduction treatments must anticipate to regret having defected with 10% probability and a payoff difference of 15. As such, we may posit that individuals who apply ex post logic, will also anticipate such potential regret when making their first period choice. Such a model would explain why the frequency of cooperation even initially tends to to be larger in *ProbRed* and *ProbRed*$^+$ than in *CertDam* and in *DamRed*.

Naturally, individual behavior is probably best described by a combination of different behavioral drivers, i.e. a combination of ex post rationality and forward looking behavior as used in subgame perfect equilibria. However, we view this section as highlighting that ex post rationality can help in explaining individual cooperation decision in stochastic contexts.

3.5 Conclusion

This paper investigates determinants of cooperation in repeated social dilemmas with stochastic damages. Such stochastic damages are linked to many environmental problems reaching from attempts to prevent of forest fires or oil spills to climate policy and hurricane prevention, but also to other challenges like public security protection against terror, or international health cooperation against pandemic diseases. We study the evolution of cooperation when the entire group benefits from individual cooperation while individual players have incentives to free-ride and may cooperate only due to (indefinitely) repeated interactions. With stochastic damages, players may take actions which either reduce the size of damages or reduce the probability that such adverse events occur.

Our results show that cooperation on probability reduction leads to significantly higher cooperation rates than cooperation on damage reduction. Specifically, the cooperation rates are sustained for probability reduction, whereas they decline for damage reduction as well as in a setting where damages are certain. The difference between the two settings can be explained by a learning dynamics which reinforces the ex post optimal action. Moreover, in line with our introductory discussion of natural disasters or accidents serving as "focussing events" (Kingdon, 1995; Birkland, 2006), we find that experiencing adverse events indeed lead to behavioral changes as players tend to revise their current strategies. However, adversity is not necessarily a school of wisdom: while non-cooperating players are more likely to switch towards cooperation following an adverse event, formerly cooperating players may rather switch towards defection.

Overall, our results may provide some optimistic view on the prospects of voluntary cooperation in dilemma situations: differently from situations where cooperation leads to (continuous) changes in the size of damages (or payoffs), more sustained cooperation can be expected if it may lead to a discrete payoff change as an adverse event may be prevented with some probability. Cautiously interpreting the results from our lab experiment in terms of our introductory examples, our findings suggest that shifting the public attention from activities which are likely to reduce the occurrence of extreme negative events (mitigation activities) to measures which reduce their impact (e.g., adaptation) may lead to declining chances for successful voluntary cooperation. More generally, our results may also guide the search for successful group incentives schemes when applied to specific policy contexts, for example, to non-point source pollution. As such, it is worthwhile to further investigate the robustness of our results in different natural settings.

3.6 References

Arellano, M., and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies* 58, 277-297.

Barrett, S. (2011). Avoiding disastrous climate change is possible but not inevitable, *PNAS* 108, 11733-11734.

Becker, G.S. (1968). Crime and punishment, an economic approach. *Journal of Political Economy* 76, 169-217.

Beggs, A.W. (2005). On the convergence of reinforcement learning. *Journal of Economic Theory* 122, 1-36.

Birkland, T.A. (2006). Lessons of disaster: poliocy change after catastrophic events. Georgetown University Press, Washington DC.

Bock, O., A. Nicklisch and I. Baetge (2014). hroot: Hamburg registration and organization online tool. *European Economic Review* 71, 117-120.

Briys, E. and H. Schlesinger (1990). Risk aversion and the propensities for self-insurance and self-protection. *Southern Economic Journal* 57, 458-467.

Camerer, C.F. and T.H. Ho (1999). Experience-weighted attraction learning in normal form games. *Econometrica* 67, 837-874.

Chmura, T., S. Goerg and R. Selten (2012). Learning in experimental 2x2 games. *Games and Economic Behavior* 76, 44-73.

Dal Bò, P.D. and G. Fréchette (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *The American Economic Review* 101, 411-429.

Dannenberg, A., A. Löschel, G. Paolacci, C. Reif, and A. Tavoni (2014). On the provision of public goods with probabilistic and ambiguous thresholds. *Environmental and Resource Economics.*

Dave, C., C.C. Eckel, C.A. Johnson and C. Rojas (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty* 41, 219-243.

Dickinson, D.L. (1998). The voluntary contribution mechanism with uncertainty group payoffs. *Journal of Economic Behavior and Organization* 35, 517-533.

Dionne, G. and L. Eeckhoudt (1985). Self-insurance, self-protection and increased risk aversion. *Economics Letters* 17, 39-42.

Eckel, C.C. and P.J. Grossman (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior and Organization* 68, 1-17.

Ehrlich, I. and G.S. Becker (1972). Market insurance, self-insurance, and self-protection. *Journal of Political Economy* 80. 623-648.

Erev, I. and A.E. Roth (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review* 88, 848-881.

Filiz-Ozbay, E. and E.Y. Ozbay (2007). Auctions with anticipated regret: Theory and experiment. *The American Economic Review*, 1407-1418.

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

Friesen, L. (2012). Certainty of punishment versus severity of punishment: An experimental investigation. *Southern Economic Journal* 79, 399-421.

Ho, T.H., C.F. Camerer and J.-K Chong (2008). Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory* 133, 177-198.

Holmström, B. (1982). Moral hazards in teams. *The Bell Journal of Economics* 13, 324-340.

Jullien, B., B. Salanie and F. Salanié (1999). Should more risk-averse agents exert more effort? *The Geneva Papers on Risk and Insurance Theory* 24, 19-28.

Lohse, T., J.R. Robledo and U. Schmidt (2012). Self-insurance and self-protection as public goods. *Journal of Risk and Insurance* 79, 57-76.

Loomes, G. and R. Sugden (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 805-824.

Keser, C. and C. Montmarquette (2008). Voluntary contributions to reduce expected public losses. *Journal of Economic Behavior and Organization* 66, 477-491.

Kingdon, J.W. (1995). Agendas, alternatives, and public policities. 2nd edition. Harper Collings, New York.

Meyer, R.J. (2012). Failing to learn from experience about catastrophes: The case of hurricane preparedness. *Journal of Risk and Uncertainty* 45, 25-50.

Miceli, T.J. and K. Segerson (2007). Punishing the innocent along with the guilty: The economics of individual versus group punishment. *Journal of Legal Studies* 36, 81-106.

Milinski, M., R.D. Sommerfeld, H.-J. Krambeck, F.A. Reed and J. Marotzke (2008). The collective-risk social dilemma and the prevention of simulated dangerous climate change. *PNAS* 105, 2291-2294.

Muermann, A. and H. Kunreuther (2008). Self-protection and insurance with interdependencies. *Journal of Risk and Uncertainty* 36, 103-123.

Rasmussen, E. (1987). Moral hazard in risk averse teams. *Rand Journal of Economics* 18, 428-435.

Roth, A. and I. Erev (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* 8, 164-212.

Segerson, K. (1988). Uncertainty and incentives for non-point pollution control. *Journal of Environmental Economics and Management* 15, 87-98.

Selten, R. and T. Chmura (2008). Stationary concepts for experimental 2x2-games. *The American Economic Review* 98, 938-966.

Selten, R. and R. Stöcker (1986). End behavior in sequences of finite prisoner's dilemma supergames. *Journal of Economic Behavior and Organization* 7, 47-70.

Tavoni, A., A. Dannenberg, G. Kallis and A. Löschel (2011). Inequality, communication and the avoidance of disastrous climate change in a public goods game. *PNAS* 108, 11825-11829.

Varian, H.R. (1990). Monitoring agents with other agents. *Journal of Institutional and Theoretical Economics* 46, 153-174.

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of behavioral decision making*, 12(2), 93-106.

Chapter 4

For “better” or “worse”: a new life-satisfaction measure able to elicit preference rankings

Co-author: Grischa Perino

4.1 Introduction

The definition of social welfare has occupied philosophers and social scientists for centuries if not millennia. On the measurement side, differences in schools of thought are compounded with limited availability of data and challenges in eliciting the specific object of interest. Purely monetary measures such as the highly influential but strongly disputed Gross Domestic Product compete with other “objectively” measurable indicators (Human Development Index, Jones and Klenow 2016, Sen 1985) and subjective measures of wellbeing such as happiness or life satisfaction (Frey and Stutzer 2002, Benjamin et al. 2014b).

A prominent example that objective and subjective welfare measures might produce vastly different assessments over time is the Easterlin Paradox. While income and subjective well-being tend to be positively correlated in the short run, average life satisfaction

(LS) has been almost constant in industrialized countries over the past decades despite a substantial increase in per capita GDP (Easterlin 1974, Easterlin et al. 2010).[1] Several explanations for this discrepancy between measures have been proposed. A leading one is that individuals get used to, i.e. adapt, to new (especially higher) levels of income [2]The literature so far has not addressed the question of whether adaptation of reported levels of LS reflect changes in preferences or are (merely) a re-scaling of the physiological and psychological sensitivity to external stimuli. While preference rankings over states of the world plausibly include a very large number of strictly separate ranks, our bodies and minds face limitations in both range and step size when attributing satisfaction levels to states of the world (Rayo and Becker 2007a,b). These limitations extend – in an even stronger form – to the answer scales used in surveys eliciting levels of LS using a discrete scale from "completely dissatisfied" to "completely satisfied", spanning three to eleven categories.[3] We label this measure of LS the level measure (LM). There is evidence that responses correspond well both to neurological measures and other indicators of well-being such as frequency of genuine smiles, spousal fights, self-assessments of mental health and suicide (Berridge and Kringelbach 2011, Camerer et al. 2005, Daly and Wilson 2009, Daly et al. 2013, Fehr and Rangel 2011, Kristoffersen 2017). It hence plausibly captures how intensely satisfaction is perceived as a sensation. But does it also capture preference rankings over states of the world? We will argue, that it doesn't.

Differences in reported levels of LS across individuals or time have been used to identify the socio-economic drivers of LS[4] and strong prescriptions have been proposed on the normative premise that LS should be maximized. Take the response to the Easterlin Paradox. It has led some to conclude that fostering economic growth should no longer be among governments' objectives but that on the contrary, income should be more heavily taxed due to the externality a higher income imposes on other citizen (Layard 2006). The claim being that rising income levels don't increase satisfaction at the aggregate

[1]The Easterlin Paradox has been challenged by Stevenson and Wolfers (2008, 2013).

[2]See e.g. Frederick and Loewenstein (1999), Diener et al. (2006), Loewenstein and Ubel (2008), diTella et al. (2010) and Vendrik (2013).Clark et al. (2016) recently presented evidence from a large-scale sample that there is no adaptation to poverty while De Neve et al. (2017) show that the subjective well-being measures are more sensitive to reductions in average income than to increases.

[3]Surveys including such a question are e.g. World Value Survey, German SOEP, UK Understanding Society, Eurobarometer, Latinobarometer, US General Social Survey, Happy Planet Index, World Happiness Index and the OECD Better Life Index.

[4]Di Tella et al. 2003, Di Tella and MacCulloch 2006, Dolan et al. 2008, Ferrer-i-Carbonell 2005, Frey and Stutzer 2002, Frey et al. 2004, Luttmer 2005, Oswald 1997, Stevenson and Wolfrers 2013.

societal level but keep citizens captured in the "hedonic treadmill" (Diener et al. 2006).

If one regards the objective of policymaking to be the maximization of perceived satisfaction levels in the population, then the above might be effective. This position is held by representatives of what Fleurbaey (2009) calls "hedonic welfarism". However, these policies might fail at maximizing utility based on preference rankings in the tradition of liberal welfare economics, which is concerned with "obtaining what one wants" instead of "being satisfied" (Fleurbaey 2009). As Frederick and Loewenstein (1999) point out, the two concepts are likely to differ in many situations. When using satisfaction data to inform policy making it is thus important to know how these two concepts differ and how they can best be measured.

We show that changes in LS are unreliable indicators of individuals' ex-post preference rankings over changes in states of the world.[5] We identify conceptual reasons for them to differ and provide evidence from representative surveys indicating that basic requirements for identifying preference rankings from panel data are not met by reported levels of LS. This challenges interpretations of LS data as valid indicators of preference rankings [6] However, a reliable subjective measure of preference rankings is desirable (Fleurbaey 2009).[7] In contrast to decision utility obtained from revealed preferences, LS is able to also evaluate changes that lie outside the control of the people affected and therefore captures a much broader set of policy relevant aspects such as the distribution of income and environmental quality.[8]

We therefore introduce the ranking measure (RM), asking subjects to directly compare the present to a specific past, i.e. two situations they have actually experienced.[9] The

[5] Welfare economics traditionally is concerned with decision utility, i.e. ex-ante preferences, this difference in interpretation is necessary due to the elicitation method of stated as compared to revealed preferences and that LS is not elicited on hypothetical but rather on actually experienced situations.

[6] As formulated e.g. by Clark et al. 2008, Daly and Wilson 2009, Decancq et al. 2015, Frey et al. 2004, Krueger and Schkade 2008, Layard et al. 2008, Liberini et al. 2017, MacKerron 2012, Oswald and Wu 2010 and Kristoffersen 2017). We focus on life satisfaction as a cognitive measure of subjective well-being instead of happiness or others that are more affective since preferences also are a cognitive concept.

[7] "[N]ew questionnaires that would enable the respondents to express their ordinal preferences more directly than through the prism of a satisfaction level, may provide valuable information about people's preferences and values relative to the various dimensions of life..." (p. 1064)

[8] Köszegi and Rabin (2008) argue that both the revealed preference approach and subjective measures of well-being should be combined when attempting to "identify preferences, mistakes, and well-being" (p. 1821).

[9] Focusing on experienced situations distinguishes the subjective well-being approaches from stated preference approaches, which ask questions about hypothetical situations.

RM avoids the conceptual shortcomings of the LM in measuring preference rankings. It imposes no limits on the number of (successive) improvements or deteriorations reported, it compares two situations while keeping both preferences and reporting functions fixed, it has a natural common scale (better, indifferent, worse) and it does not need to cardinalize an ordinal concept when collecting the data and to then make it ordinal again when analyzing it.

Bond and Lang (2014) and Schroeder and Yitzhaki (2017) provide strong evidence that measuring a continuous variable such as LS using only a few categories makes it impossible to unambiguously rank mean satisfaction of groups of respondents without imposing strong assumptions on the underlying distribution of LS and on a commonly shared reporting function. Indeed, Bond and Lang (2014) show that standard results of the LS literature can be reversed by using plausible alternative assumption about the distribution of LS. Our critique differs from theirs in that we focus on intra-personal rather than inter-personal (or inter-group) comparisons of survey responses and we propose a solution to the problem of the ranking challenge. As is the case for Bond and Lang (2014), the basic argument of our critique (and solution) can be extended to other continuous variables that are measured with a small number of categories (such as Likert scales) and where identification relies on the ability to rank responses. Such variables are abound in economics, psychology and the social sciences in general.

Comparing the preference rankings implied by LM and RM for three representative surveys with a total of well over a hundred thousand observations confirms that changes in LM are only weakly correlated with RM. Moreover, deviations between the two measures are systematically correlated with socio-demographic variables indicating that at least one of them is a biased representations of preference rankings. The RM correlates more strongly and intuitively with changes in socio-economic aspects, while the LM has an inherent tendency towards adaptation both conceptually and empirically, especially if there is a trend in the evolution of income. We extend previous work on the relationship between subjective well-being measures and preference rankings in several dimensions: First, we focus on the comparability of satisfaction measures within subjects across time, evaluating the validity of changes over time. Other studies have analyzed the validity of LS at one point in time, comparing it with measures of revealed or hypothetical choices and anticipated subjective well-being (e.g. Benjamin et al. 2012, 2014a, 2014b, Perez-Truglia 2015). The inter-temporal comparability is most important when using LS data from representative panel surveys and hence for much of the empirical work on LS (Fleurbaey and Schwandt 2016). It also seems the more natural direction of comparison, when trying to identify preference rankings, which by definition are intra-personal comparisons.

We combine conceptual reasoning with empirical evidence drawing on three representative surveys, one of which involves a randomized controlled trial.

Second, we propose a new elicitation method, the ranking measure, fixing several of the conceptual issues arising when comparing the LM across time to elicit preference rankings over states of the world.

Last, we provide evidence that the adaptation results characteristic of the LM might not primarily be driven by changes in preferences but rather represent adaptation in the reporting function.[10]

The remainder of this paper is organized as follows: The next section (4.2) relates the LM of LS to requirements for eliciting preference rankings over states of the world when using panel surveys. Section 4.3 presents evidence from a field experiment with a representative sample testing for some of the key assumptions necessary for the LM to represent preference rankings. The RM is introduced in Section 4.4 and Section 4.5 compares the two measures based on three representative datasets. Section 4.6 concludes.

4.2 Preference rankings and the level measure of life satisfaction

In what follows, we derive necessary conditions for the LM to be able to capture individual preference rankings.[11] Assume that individual i has rational and continuous,

[10]Note that we do not claim that perceived satisfaction does not adapt but merely that it does not represent a change in the underlying preference ranking.

[11]Note that we don't normatively argue that preferences should be the objective of policies nor that they are the only possible interpretation of LS. We positively identify the conditions under which the LM is able to represent the underlying preferences of respondents.

potentially time-dependent, preferences $\succeq_{i,t}$ over all relevant pairs of situations or states of the world X. A situation $x_{i,t} \in X$ is a vector including individual i's characteristics like number of children, size of the house, education but also aggregates such as GDP, its distribution, inflation, pollution or combinations of both such as own position in the income distribution of a peer group etc. at a particular point in time t.

4.2.1 Reporting functions and measurement of life satisfaction

For a LS measure to capture preferences, a higher reported value has to imply that the respondent strictly prefers the situation to which she assigns a higher reported level of LS. In contrast to utility functions, the reverse might not hold, as the LM features only a small number of categories. Formally, $LM_{i,t}(x_{i,t}) = f_{i,t}(x_{i,t}, \succeq_{i,t})$ is the reported satisfaction level of individual i for situation x at time t and $f_{i,t}$ a monotonically increasing reporting function mapping preferences $\succeq_{i,t}$ onto answer categories in the survey. With this notation, we can therefore distinguish between changes in the preference order and changes in how a given preference order is reported in a survey that might occur across individuals or time. Layard et al. (2008) and Fleurbaey and Schwandt (2016) e.g. use a reporting function but allow variation only across individuals but not across time.

LM questions slightly differ across surveys, but they all ask for the current level of LS and the answer scale provides a fixed number of categories represented by labels and numbers, e.g. the British Household Panel Survey (BHPS) asks "[...] how dissatisfied or satisfied are you with your life overall?" and the answer categories range from 1 "Not satisfied at all" to 7 "Completely satisfied".

4.2.2 Within-subject variation of reporting functions

Many authors use cross-sectional data for LS analysis and thus use between-subject variation. However, interpersonal comparison of what essentially is an ordinal concept is highly controversial (see also Bond and Lang 2014). Researchers therefore often use panel data, if available, focusing on within-subject comparisons over time. This is mostly done using fixed effects estimators, analyzing the variation around within person means. For detecting preference rankings of two situations, first differences, i.e. $dLM(x_{i,t}, x_{i,t-1}) =$

$LM_{i,t}(x_{i,t}) - LM_{i,t-1}(x_{i,t-1}) = f_{i,t}(x_{i,t}, \succeq_{i,t}) - f_{i,t-1}(x_{i,t-1}, \succeq_{i,t-1})$, or rather $sign(dLM)$, is more suitable. The latter, however, involves not only two states of the world, but potentially two different preference relations and two reporting functions.[12] To infer the underlying preferences over situations $x_{i,t}$ and $x_{i,t-1}$ from changes in the observed LM, both the preference relation and the reporting function need to remain unchanged over the two periods.

In principle, both might change from one period to the next. For the moment, let's assume that preferences remains unchanged for at least two consecutive measurements and focus on adjustments of the reporting function. Let's also assume that respondents try to report their preferences, i.e. that monotonicity of $f_{i,t}$ at any point in time is given.

Two main reasons for adjusting the reporting function are: first, the answer scale has no intuitive meaning. The condition for reporting to be in the fifth out of seven satisfaction categories is ambiguous.[13] Respondents need to come up with an ad hoc reporting function making it hard to remember the function until the next wave of the survey. They might hence use a different one each time. As long as there is no systematic change distorting the representation of the underlying preferences this simply introduces noise to the data and in principle can be fixed by using large data sets (Bertrand and Mullianathan 2001, Frey and Stutzer 2002).

The second and more problematic reason for changes in the reporting function is that the scales of the preference concept and of the survey question don't match. While the answer scale has only a few categories, e.g. 7 in the BHPS, a complete mapping of preference relations over all possible states of the world requires as many categories as there are states that can be ranked in a strict sense. For most respondents the latter will be a much higher number.

Respondents therefore face a trade-off when choosing a reporting function. They might either choose a steep one to report small to medium changes relevant in the present or they opt for a flat one to capture the "big picture". However, both can be achieved

[12] Fixed effects models, by relying on the within person mean, implicitly assume that the reporting function and preferences are constant over the entire period covered by the data.

[13] Conti and Pudney (2011) report evidence that minor changes such as displaying labels for only three instead of all seven categories in the BHPS substantially affects reported satisfaction levels, especially for women.

when using a steep reporting function but adjusting it to recent experiences. Such an adjustment is common for sensory perceptions[14] and in line with neurological limits to perceive sensations. A moving reference base for neurologically perceived LS allows an individual to capture the direction of changes in the short run even with a limited set of discrete levels of perception available (Rayo and Becker 2007a,b). With such an adjustment, the adaptation of the reporting function is not random but systematically depends on past and expected changes, which introduces biases that cannot be rectified by large numbers of observations (Bertrand and Mullianathan 2001).

In the literature, adaptation of LS to new situations is often interpreted as a "getting used to", i.e. a change in $\succeq_{i,t}$ (e.g. DiTella et al. 2010). However, in general one cannot tell whether the observed pattern is due to a preference change or a shift in the reporting function.[15] Thus, unless respondents remember previous answers and use the same reporting function that is independent of any variables of interest across all waves of the survey, LM does not allow to reliably infer preference rankings from changes in its level even if participants attempt to report their preferences.

There is compelling evidence that self-assessments in surveys, including quality-of-life measures, are subject to reference biases, social norms and therefore vary over time even for a given individual (Groot 2000, Gil and Mora 2011, Bond and Lang 2014). Bond and Lang (2014) show that becoming disabled induces an adjustment of the reporting function relative to the control group. Reporting functions hence not only differ between individuals but also change over time and in response to changes in key socio-economic variables. This casts serious doubts on attempts to identify the impact of changes in such variables on preference rankings based on the LM.

[14]When judging sensory magnitudes with an objective cardinal scale, humans build an ad-hoc reference frame that depends on the range of stimuli they are exposed to when reporting them on a categorical scale. "Sensory magnitudes are selected for this review of biases in judgement because the stimuli can be measured on a physical scale. Judgements of the quality of life [...] lack a precise measure of the stimuli. Thus the biases are more difficult to specify exactly." (Poulton 1979, p. 777). See also Oswald (2008).

[15]The possibility that reporting functions change over time has been ignored not only in much of the economics literature on LS but also in psychology. See e.g. Karney and Coombs (2000) where current subjective assessments of marital satisfaction are taken to be the accurate measure when compared with recollections ten years later although the recollection question asked neither implies that participants should use the past utility nor the past reporting function. Hence, in line with the Reassess treatment reported in Section 3, there is no reason to expect that recollection answers match former current assessments. One might therefore question, whether the study really identifies a memory bias or merely an adjustment in either utility or reporting function.

4.2.3 Reporting life satisfaction when its drivers follow a trend

The identification problem is stronger when many improvements (or deteriorations) are experienced or expected in a row, because then the limited answer scale is particularly restricting. While in dimensions such as income improvements might be expected to happen repeatedly, in others such as marital status this seems less plausible. A respondent might therefore choose a flatter reporting function or a different adaptation rule for the income dimension than for the marriage dimension such that the relative impact of both dimensions on LS cannot be compared to infer preferences. Hence, estimates of marginal rates of substitution are distorted (Benjamin et al. 2014a). This is in line with observations that income coefficients in LS regressions are often found to be relatively small (see e.g. Kopmann and Rehdanz 2013). As (differential) adjustments in reporting functions cannot be distinguished from changes in preferences in LM data, one might draw misleading conclusions on people's preferences.

The Easterlin Paradox for example might not imply that people are indifferent to higher levels of (average) income in the long run. LS data might simply not reveal the preference for increasing income due to an adapting reporting function. While the problem of preference misrepresentation has been mentioned in the literature before, solutions are still rare.[16]

We want to stress that our analysis only applies to the preference relation interpretation of LS. The LM might be a good representation of the mental state, i.e. the intensity of sensation in a given situation. Intensity of sensation might thus well have adapted to higher levels of income. But this does not mean that respondents are indifferent to increasing income. Our aim is to show why these two interpretations of LS will systematically differ. We abstain from making a normative judgment as to which of the concepts should be policy relevant.

[16]See e.g. Benjamin et al. (2014a), Bertrand and Mullianathan (2001), Fleurbaey (2009), Frederick and Loewenstein (1999), Frey and Stutzer (2002), Loewenstein and Ubel (2008).

4.3 Experimental evidence on the level measure

In this section we test how well participants remember past answers or the reporting function which are preconditions for eliciting preference rankings from within-subject comparisons of LM data. Since estimation methods rely on participants reporting higher levels of LS in situations they prefer, it is crucial that participants remember not only how satisfied they were in previous rounds of the survey but also the answer categories chosen. Some white noise can be taken care of by using large numbers of observations but if memories are biased – or non-existent – elicitation of preference rankings is greatly hampered.

4.3.1 The experiment

We commissioned a survey of a representative sample of the German population. The survey was conducted in two waves about three months apart by the market research company Lightspeed between June and October 2015. In wave 1 we asked a total of 2,300 participants for their age and gender, the LM (LM1) and whether they would be willing to participate in a later wave on a similar topic. This creates a situation similar to those faced by participants in large annual panels like the BHPS or GESIS. All participants agreed to be re-contacted. A total of 1,600 participants also participated in the second wave equally and randomly distributed over four treatments.

The treatments differed in the type and order of LS questions in wave 2 (Table 4.1). In treatments RecallExAnte and RecallExPost participants were asked to recall the answer to LM1 given three months ago (LM1recall) and state whether they felt sure or unsure about the accuracy of their answer. The option “I cannot remember at all” was also available. The two treatments differed only with respect to the ordering of questions. In RecallExAnte LM1recall was asked first and hence participants could use it to anchor their response to the level measure (LM2) and the ranking measure (RM, see next section). This was not possible in RecallExPost where LM1recall was elicited last. Participants were prevented from changing their responses to previous questions.

In treatment Reassess participants had to give a retrospective assessment of LS for the point in time they had completed wave 1 of the survey (LM1reassess). In contrast to LM1recall there is no objectively correct answer to LM1reassess. LM1reassess might differ from LM1 because of imperfect recollection of the situation three months ago or because either the reporting function or preferences might have changed. Answering LM1reassess requires participants to be able and willing to apply the current reporting function to the situation three month ago. A precondition is that they perceive their recollection of that situation to be sufficient to do that. All questions as well as the summary statistics can be found in Appendix C.1.

Treatment	Wave 1	Wave 2			
Control	LM1		LM2	RM	
RecallExAnte	LM1	LM1recall	LM2	RM	
RecallExPost	LM1		LM2	RM	LM1recall
Reassess	LM1	LM1reassess	LM2	RM	

Table 4.1: Order of questions asked in Lightspeed survey. 400 participants per treatment.

4.3.2 Recollection ability and precision

Treatments RecallExAnte and RecallExPost check how well participants remember answers to the LM over three months. Interpreting changes in the LM over time as preference rankings requires that participants use their previous answer as a reference point and correctly report improvements or deteriorations accordingly. This requires that they either remember the answer given in the previous wave of the survey or that they correctly remember both the situation they were in and the reporting function used in order to reconstruct their previous response.

Over both treatments 43.3 percent of participants stated that they could not at all recollect their answer to LM1 or gave no number (see Table 4.2). Out of the participants that answered LM1 and took part in the wave 2 only 20 percent correctly recalled their answer three months later.[17] In RecallExAnte the conditional rate of accurate answers was higher for those participants that stated a lower confidence (32.8 vs. 37.8 percent).[18] Thus, confidence here is a poor predictor of ability. The frequency of accurate answers is

[17] Based on a uniform distribution, the success rate of a random draw would have been 14.3 percent.

[18] Over both treatments combined the accuracy rate of the confident is 37.2 percent vs. 34.7 percent of the less confident.

better than a random guess but the amount of noise introduced by imperfect recollection is substantial.

Next, we test whether the noise dominates the signal, i.e. the information on the underlying preferences the respondent tries to transmit.

Treatment	LM1 (#)	LM1recall Confidence level	LM1recall Total (#)	LM1recall Accurate (#)
RecallExAnte	397	Precise	67	22
		Guess	188	71
		Cannot recall/no answer	119/23	–
RecallExPost	399	Precise	27	13
		Guess	169	53
		Cannot recall/no answer	192/11	–

Table 4.2: Self-assessed and real accuracy of recollection of LM1. LM1recall and accuracy conditional on answering LM1.

The direction of the observed change in LS (sign(LM2 – LM1)) only measures the preference ranking accurately if the recollection error (LM1recall – LM1) does not change the "signal" on their preferences (LM2 – LM1recall) that participants are assumed trying to send. In RecallExPost[19], the recollection error results in a misrepresentation of the preference order for 102 out of 196 participants (52%) that answered all three questions. Thus, the preference rankings implied by (LM2 – LM1) and (LM2 - LM1recall) differ. Using the reference point that is commonly observable in surveys (LM1) results in a different preference ranking than using the one that participants have in their minds (LM1recall). A further 203 participants that answered LM1 and LM2 started to be unable to recall LM1. For the latter the consistency rate cannot be determined but is likely to be lower than for those that did recall LM1.

For the latter the consistency rate cannot be determined but is likely to be lower than for those that did recall LM1.

Result 1: *Within subject variation of LM is an unreliable measure of preferences. About half (50.9 percent) of respondents stated to be unable to remember the answer given to a LM question three months ago, although a 'low confidence' option was available. For*

[19]We analyze RecallExPost since in RecallExAnte the answer to LMrecall might influence the later answer to LM2 and thus distort the observed change in LS. On the other hand, we thereby have to accept that LMrecall might be influenced by the earlier answer to LM2.

those that stated to remember, the preference ranking implied by the standard measure (LM2 - LM1) differed from that based on the recollection (LM2 - LM1recall) in about half of all cases (52 percent).

Consistency rates in standard surveys will be lower due to the four times longer gap between consecutive waves. Hence, even when abstracting from the possibility that reporting functions and preferences can change over time, within-subject comparisons of the LM seem ill fitted to produce reliable information on preference rankings.

4.3.3 Reassessment ability

Allowing participants to reassess the situation three months ago rather than asking them to recall their prior assessment, increases response rates significantly. Instead of 64 and 49 percent (of those answering LM1) in treatments RecallExAnte and RecallExPost, respectively, 98.25 percent (393 out of 400) answered the LM1reassess question in treatment Reassess including four participants that did not answer LM1. The difference to LM1 is comparable to the answers given to LM1recall in terms of mean deviation and number of exact matches. The variance of (LM1 - LM1reassess) is significantly larger than that of (LM1 - LM1recall) ($p < 0.001$, variance ratio test). However, there is no reason to expect participants (even those with perfect memories) giving the same answers to LM1reassess and LM1 since both the reporting function and preferences might have changed. The precision of reassessments is unobservable.

Result 2: *Almost all participants (98.25 percent) provide reassessments of the satisfaction three months ago based on current preferences and reporting function.*

Respondents struggle to remember the category selected in a LM question three month ago but have no trouble in reporting a reassessment of the past situation using a new ad hoc reporting function. This implies that many cannot remember the previously used reporting function as otherwise they could reconstruct (with error) their past answer. The experimental test indicates that within subject changes in the LM are a poor indicator of preference rankings even with only three months in between.

4.4 The ranking measure

Asking individuals directly to rank the current situation against the situation from the previous period prevents an uncontrollable change in reporting function from one wave to the next and the memory issues detected in the previous section. An example of such a question from the BHPS:

"Would you say that you are more satisfied with life, less satisfied or feel about the same as you did a year ago?" where the answer options were "More satisfied", "Less satisfied", "About the same" and "Don't know".

Responses to an RM question can be described by $RM_{i,t}(x_{i,t}, x_{i,t-1}) = r_{i,t}(x_{i,t}, x_{i,t-1}, \succeq_{i,t}$). The individual compares today's situation $x_{i,t}$ with last period's situation $x_{i,t-1}$ using the current preference relation $\succeq_{i,t}$. The preference ranking is transformed into an answer on the answer scale according to the reporting function $r_{i,t}(.)$, i.e. it gives a retrospective evaluation of the experienced change in life.

4.4.1 When does the ranking measure capture preferences?

Under which conditions is the RM reliably able to capture preference rankings? First, if there is no bias in memory about last period's situation, the preference ranking of the two situations is weakly correct. The reporting function might be coarser than actual preferences, such that a very small improvement might still be reported as indifference. The RM requires individuals to remember the past situation (without bias). However, people might be reluctant reporting that things have become worse or glorify the past, therefore biasing the RM up or down. While the RM is not immune against memory biases[20], it requires memories of one's life only while the LM on top of that critically relies on remembering an ad hoc reporting function or the answer to a specific survey questions for an entire year.

Secondly, participants need to intend to report their preference rankings and not another concept. Thus, $r_{i,t}(.)$ needs to (weakly) preserve the preference ranking. This condition is

[20] On memory biases see Hoffrage et al. (2000), Karney and Coombs (2000), Levine and Safer (2002) and Morewedge et al. (2005). Note that studies of memory bias are not immune against misspecification either.

equivalent to the monotonicity requirement of $f_{i,t}(.)$ for the LM. The BHPS question did not clearly ask to report a preference ranking instead of a change in sensation intensity and can be improved in that respect. Therefore we commissioned the following question in the GESIS panel and used the same phrasing in the Lightspeed survey:

"In what follows we would like to know how you evaluate the development of your life over the past 12 months. In your opinion, has your life overall improved or deteriorated over the past 12 months? Please evaluate the changes from today's perspective."

4.4.2 Advantages of the ranking measure

In contrast to the LM (see section 2.2), two situations are assessed and reported using the same preference relation and reporting function. Moreover, it is possible to observe arbitrarily long sequences of improvements or deteriorations. The RM also uses more intuitive categories. It seems easier to agree on what "improvement" means than on what is a "4 out of 7". This increases comparability across individuals. While there might still be heterogeneity in cut-offs between categories across individuals and time, this does not impede to rank two states of the world based on a "better" or "worse" response. With the LM heterogeneity in cut-offs between categories can reverse rankings (Bond and Lang 2014).

Direct measurements of changes have been found to outperform computed changes based on repeatedly reported levels in subjective assessments of health (Gunasekara et al. 2012) and preference reversals (Bateman et al., 2007).

The RM thus eliminates the uncontrollable change in utility and reporting functions, reduces the memory issues, is more intuitive, avoids the bounded scale, is ordinal and promises therefore to be better suited to measure preference rankings as compared to the LM.

4.5 Comparison of measures: empirical evidence

Above we argued that the LM is a highly imperfect way to elicit preference rankings and suggested the RM as an, at least conceptually, more reliable alternative. Next we compare the two measures empirically.

To allow a direct comparison we reduce the LM to the information relevant for a preference ranking, i.e. whether an individual evaluates the present situation as "better" or "worse" than or is "indifferent" to the situation a year ago. This is achieved by using the sign of the change in the level measure $signdLM_{i,t} = sign(dLM_{i,t})$ with $dLM_{i,t} = LM_{i,t} - LM_{i,t-1}$.

We label the difference between RM and signdLM $deviation = RM - signdLM$. It captures when the preferences elicited by the two measures differ. If $deviation = 0$, both measures report the same preference ranking. If, however, $|deviation| = 2$ they give opposite rankings. If $|deviation| = 1$, they weakly disagree.

4.5.1 Data

We draw on data from BHPS (University of Essex 2010), GESIS (GESIS 2017) and Lightspeed surveys which contain the LM and the RM. The RM question in the GESIS panel and the complete Lightspeed survey were commissioned by us. Data descriptions and summary statistics can be found in Appendix C.1.

The BHPS panel is representative for Great Britain and starting from 2001 also for the whole of the UK. The two LS measures are available for the years 1996 to 2008 (with 2001 missing) adding up to 117,244 observations. As far as we are aware, the data of the RM in the BHPS has never been used so far in any publication.

The GESIS panel is representative for Germany and features a more precise RM question for eliciting preferences (as discussed in the previous section) than the BHPS. While the LM was included in the GESIS panel starting 2014, the RM was added on our behalf in 2015. There are 9,553 valid observations for the years 2015 to 2017 in total for which both LM and RM are available.

The Lightspeed dataset (see section 4.3) is representative of Germany and contains 794 observations from the *Control* and *RecallExPost* treatments where LM2 and RM are not affected by the experimental setup.

	BHPS	GESIS	Lightspeed
Country	UK	Germany	Germany
Years covered	1996 - 2000 2002 - 2008	2014 - 2017	2015
Observations	117,244	9,553	794
Mean signdLM	-.15*** (.002)	.006 (.008)	.006 (.028)
Mean RM	.11*** (.002)	.279*** (.010)	.144*** (.022)
Mean deviation	.046*** (.002)	.273*** (.010)	.137*** (.033)
signdLM = 0 (%)	46.4	32.1	37.2
RM = 0 (%)	57.9	39.2	59.4
$signdLM < 0(\%)$	27.6	33.6	31.1
$RM < 0(\%)$	15.6	16.4	13.1
Weak disagreement (%)	47.1	42.73	48.7
Agreement (%)	45.5	41.1	41.8
Corr(signdLM,RM)	.21	.18	.17
t-test signdLM = RM (p-value)	.000	.000	.000
Marginal homogeneity (Stuart Maxwell, p-value)	.000	.000	.000

Table 4.3: Comparison of RM and signdLM. Standard error in brackets. Lightspeed based on treatments Control and RecallExPost. Difference from zero is reported at * $p < 0.05$, **$p < 0.01$, ***$p < 0.001$ significance levels.

4.5.2 Consistency between the two life satisfaction measures

Table 4.3 reports the descriptive data and test results for signdLM, RM and deviation. Across all data sets, comparing the means, we find that signdLM tends more towards the negative than RM, confirmed by a significant positive mean deviation. Still, the share of participants reporting no noteworthy change in LS is higher for the RM than the LM which is consistent with substantially more noise in LM answers. Across all datasets, the preference rankings elicited from the two measures at least weakly disagree in more than half of all cases confirming results from section 4.3. Correlation of both measures is around .2 and tests of asymptotic symmetry and marginal homogeneity are rejected. The two measures are thus significantly different in terms of means and distributions.

Result 3: *RM and signdLM are significantly different in mean and distribution, only* *weakly correlated and the elicited rankings of states of the world disagree in more than half of all cases.*

Some of the inconsistency between signdLM and RM is due to noise in signdLM that is caused by inaccurate but unbiased recollection of the LM reported in the previous wave. Treatment RecallExAnte in the Lighspeed dataset allows to quantify this by comparing consistency rates between signdLM and RM (44.4%) and between LM2 – LM1recall and RM (68.2%), thus 43% (= 1 – (100 – 68.2) / (100 - 44.4)) of inconsistency between signdLM and RM is due to inaccurate recollection. In the first case participants had to remember their answer for three months, in the second only for a few seconds. Systematic adjustments in the reporting function cannot be eliminated by this procedure.

4.5.3 Drivers of life satisfaction measures

Next we test whether the difference between the measures is systematic. Tables 4.4 and 4.5 contain results for the BHPS and Table 4.6 for the GESIS data set. The tables report regressions of deviation, signdLM and RM on changes in explanatory variables ($dX_{i,t}$) such as income, employment status and marital status, on time-invariant personal variables (X_i) such as gender, on year dummies (X_t) and allowing for the possibility of unobserved personal fixed effects (u_i):

$$signdLM_{i,t} = sign(LM_{i,t} - LM_{i,t-1}) = \beta_1 dX_{i,t} + \beta_2 X_i + \beta_3 X_t + u_i + e_{i,t}$$

$$RM_{i,t} = \hat{\beta}_1 dX_{i,t} + \hat{\beta}_2 X_i + \hat{\beta}_3 X_t + \hat{u}_i + \hat{e}_{i,t}$$

$$deviation = RM_{i,t} - signdLM_{i,t} = \tilde{\beta}_1 dX_{i,t} + \tilde{\beta}_2 X_i + \tilde{\beta}_3 X_t + \tilde{u}_i + \tilde{e}_{i,t}$$

For BHPS, two specifications are presented differing in the way changes in income are measured: self-assessed changes in financial situation (improved, stayed the same, deteriorated) are used in Table 4.4 and changes in reported household income in Table 4.5.

For GESIS, results based on the change in financial situation are reported in Table 4.6.[21] We have checked for multicollinearity by means of the variance inflation factor, which for all variables in all specifications is between 1 and 2 except for age and age2, which by definition are correlated with each other.

We report the results of pooled OLS regressions. According to the Breusch-Pagan test, unobserved personal fixed effects are present for the RM and a Hausman test suggests using a fixed effects regression which we report in addition to pooled OLS. The results are quite similar. The fixed effects specification takes into account personal fixed effects in changes over time. Person fixed effects in levels are already accounted for in the pooled OLS regressions since we use first differences (changes over time) as dependent variable and for most independent variables. Since the dependent variables have either three or five categories, an ordered logistic regression is reported in the appendix as robustness check (Table C.10). Specifications using the natural logarithm of income are also reported in the appendix (Table C.12). This does not improve the significance of the income variable but also doesn't change the other coefficients.

Regressions (1) in Tables 4.4 and 4.5 confirm that the deviation between the preference relations elicited by LM and RM is systematically correlated with key socio-demographic variables. Age, gender, changes in the financial situation, changes in employment status and changes in marital status are significant drivers of deviation in the BHPS. Hence, at least one of the LS measures provides biased estimates of preference rankings over these attributes.[22] In regresseions (2) and (3), the R^2 is consistently about ten times higher in the RM specification than for signdLM, consistent with the RM being less noisy. Note that the only difference between regressions (2) and (3) is the dependent variable which both take one of three values (-1, 0, 1). The direction of direct impacts of changes in socio-economic variables on the two measures of LS is the same and intuitive – with the exception of age were both the linear and the quadratic terms have opposing signs.[23] However, the size of coefficients for RM is roughly twice that for LM which explains the systematic differences in deviation.

Result 4: *Changes in socio-economic variables have intuitive effects on both LS mea-*

[21]The version with income is presented in the appendix (Table C.9).
[22]This also holds for the Lightspeed dataset (see Table C.4).
[23]Results excluding the lags are reported in Tables C.5 and C.6.

sures but their impact is consistently and significantly stronger for RM.

4.5.4 Adaptation

When comparing the coefficients of changes and lagged changes in socio-economic variables in Tables C.5 and C.6, a clear pattern emerges. An improvement in financial situation, finding a job, leaving the labor market (not due to unemployment but e.g. into retirement) and getting married all have positive immediate impacts on both measures, a positive impact on next year's RM but a negative impact on next year's LM. The same pattern but with reversed signs holds for becoming unemployed.

The LM therefore has a clear tendency towards findings that people get used to whatever happens to them. The key question is whether this adaptation reflects changes in preferences ('once you got to know it, more money isn't that great and unemployment not that bad after all compared to other things'), which has been a dominant interpretation in the adaptation literature, or merely an adjustment of the reporting function (that might reflect changes in satisfaction sensation but not preference rankings).

The RM exhibits the opposite pattern. Some of the benefits of an improvement in the financial situation and some of the downsides of unemployment only seem to occur to (or hit) people with some delay. By construction of the RM, this cannot be attributed to changes in the reporting function. At least for some of the variables plausible explanations for the reinforcing impact of lags exist. After losing a job, people might first be hopeful that they find a new one soon. One year later, this hope might have died. Note that both changes in the financial situation and whether a new job has been found in the meantime are controlled for.

Result 5: *Lagged changes in socio-economic variables exhibit the standard adaptation pattern for the LM but the opposite pattern for RM.*

For the GESIS data in Table 4.6, this adaptation pattern is only observed for the change in financial situation, because socio-economic drivers are generally far from being significant.[24] The results for GESIS using income instead of financial situation can be found

[24] This seems to be due to the fact that there are only three years of observations. When the BHPS is only analyzed for three periods (instead of ten), the picture gets very similar (see appendix Tables C.7

in Table C.9.

4.5.5 Trends in socio-economic variables and changes in reported LS

Is it possible to identify what is driving this difference in the impact of lagged changes in life circumstances? One of the motivations for adjusting the reporting function of the LM is to preserve the ability to report future changes on a bound answer scale. This is especially relevant if there is a clear trend in the variable to be assessed. Changes in the financial situation are the only candidate among the variables available here that can exhibit a trend over several periods.

Table 4.7 provides some additional insights into the adaptation pattern that is observed for the LM but not the RM. It reports how deviation is affected by trends in the financial situation. For this purpose, changes in financial situation have been categorized into improvements, no change and deteriorations that enter both directly and lagged.[25] Moreover, the dummy variables 'positive trend' and 'negative trend' have been created. They equal one if the financial situation has improved (deteriorated) for two successive periods. Coefficients for both are highly significant and lend support to the hypothesis that reporting functions of the LM are adjusted more if there is an experienced trend in the assessed variable.[26]

Result 7: *The deviation between preference rankings elicited by signdLM and RM increases in absolute terms if there is a trend in a socio-economic driver of LS.*

Do adjustments of the reporting function only happen when participants have reached the bounds of the answer scale? The main regression results from Table 4.4 remain when excluding observations where LM is either 1 or 7 in the BHPS (Table C.13) or the lag of LM is either 1 or 7 (Table C.14), i.e. when participants have reached the end

and C.8).

[25]De Neve et al. (2017) study the differential impact of positive and negative growth on the level measure and find that levels of subjective well-being are about twice as sensitive to negative as compared to positive growth. While we find a substantially larger coefficient for improvements than for deteriorations, this does not contract their results. Most importantly, the deviation measure captures the difference between signLM and RM and hence does not speak on the change in the LM directly. Moreover, we measure changes in financial satisfaction at the individual level while De Neve et al. (2017) consider macroeconomic events.

[26]Results are similar but not significant when using leads instead of lags for creating the trends, which can be interpreted as proxies for expected trends (see Table C.11).

Table 4.4: BHPS: Regressions on key socio-economic variables (financial situation). Variable description in Appendix C.1. Results from pooled OLS and fixed effects regressions including year fixed effects.

	(1)	(2)	(3)	(4)
	deviation OLS	signdLM OLS	signRM OLS	signRM FE
female	0.0171**	0.00332	0.0205***	0
	(0.00615)	(0.00336)	(0.00567)	(.)
age	-0.0209***	0.00466***	-0.0162***	-0.0292
	(0.00109)	(0.000649)	(0.000971)	(0.0326)
age2	0.000157***	-0.0000456***	0.000111***	-0.00000131
	(0.0000106)	(0.00000642)	(0.00000942)	(0.0000231)
change_fin_situation	0.105***	0.0844***	0.189***	0.163***
	(0.00474)	(0.00399)	(0.00386)	(0.00349)
L.change_fin_situation	0.105***	-0.0593***	0.0458***	0.0179***
	(0.00469)	(0.00400)	(0.00357)	(0.00350)
stayed unemployed	-0.120***	-0.00902	-0.129***	-0.0907**
	(0.0332)	(0.0245)	(0.0277)	(0.0326)
stayed out of labour market	-0.0428***	0.00382	-0.0390***	0.0359**
	(0.00883)	(0.00522)	(0.00792)	(0.0121)
got employed	0.0548**	0.0439**	0.0987***	0.125***
	(0.0175)	(0.0148)	(0.0136)	(0.0139)
got unemployed	-0.0922***	-0.0896***	-0.182***	-0.114***
	(0.0250)	(0.0212)	(0.0199)	(0.0186)
exited labor market	0.0726***	0.0432**	0.116***	0.138***
	(0.0166)	(0.0141)	(0.0132)	(0.0127)
lag got employed	0.0198	0.00470	0.0245*	0.0415***
	(0.0151)	(0.0129)	(0.0115)	(0.0115)
lag got unemployed	-0.154***	0.0501*	-0.104***	-0.0635**
	(0.0272)	(0.0233)	(0.0207)	(0.0199)
lag exited labor market	0.109***	-0.0403**	0.0688***	0.0398**
	(0.0169)	(0.0142)	(0.0126)	(0.0126)
got married	0.190***	0.0535**	0.244***	0.113***
	(0.0230)	(0.0194)	(0.0173)	(0.0180)
marriage ended	-0.141***	-0.0213	-0.163***	-0.252***
	(0.0310)	(0.0284)	(0.0277)	(0.0233)
stayed married	-0.00202	-0.00939*	-0.0114	-0.174***
	(0.00705)	(0.00400)	(0.00645)	(0.0136)
lag got married	0.192***	-0.0328	0.159***	0.150***
	(0.0234)	(0.0197)	(0.0180)	(0.0177)
lag marriage ended	-0.0469	0.0895***	0.0426	-0.0251
	(0.0307)	(0.0264)	(0.0228)	(0.0219)
Constant	0.233***	-0.117***	0.116***	1.099
	(0.0306)	(0.0197)	(0.0271)	(1.368)
Year dummies	Yes	Yes	Yes	Yes
Observations	83533	83533	83533	83533
R^2	0.052	0.012	0.109	0.052
Adjusted R^2	0.051	0.011	0.109	-0.220
AIC	207001.0	183091.7	151818.8	114298.3
BIC	207243.7	183334.4	152061.5	114531.6

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.5: BHPS: Regressions on key socio-economic variables (income). Variable description in Appendix C.1. Results from pooled OLS and fixed effects regressions including year fixed effects (not reported).

	(1) deviation OLS	(2) signdLM OLS	(3) signRM OLS	(4) signRM FE
female	0.0146*	0.00300	0.0176**	0
	(0.00633)	(0.00336)	(0.00591)	(.)
age	-0.0252***	0.00407***	-0.0211***	-0.0126
	(0.00110)	(0.000642)	(0.000989)	(0.0332)
age2	0.000191***	-0.0000402***	0.000151***	0.0000175
	(0.0000108)	(0.00000637)	(0.00000961)	(0.0000233)
d_hh_inc_month	0.00177	0.00159	0.00336*	0.00151
	(0.00192)	(0.00181)	(0.00142)	(0.00137)
L.d_hh_inc_month	0.00262	-0.00237	0.000245	-0.00131
	(0.00194)	(0.00163)	(0.00144)	(0.00137)
stayed unemployed	-0.183***	-0.0230	-0.206***	-0.166***
	(0.0333)	(0.0247)	(0.0285)	(0.0328)
stayed out of labour market	-0.0822***	-0.00102	-0.0833***	-0.00271
	(0.00895)	(0.00515)	(0.00811)	(0.0122)
got employed	0.0440*	0.0695***	0.114***	0.140***
	(0.0174)	(0.0148)	(0.0139)	(0.0141)
got unemployed	-0.165***	-0.125***	-0.290***	-0.210***
	(0.0250)	(0.0213)	(0.0206)	(0.0187)
exited labor market	0.0203	0.0158	0.0361**	0.0700***
	(0.0166)	(0.0141)	(0.0133)	(0.0128)
lag got employed	0.0418**	0.00164	0.0434***	0.0538***
	(0.0153)	(0.0129)	(0.0120)	(0.0116)
lag got unemployed	-0.186***	0.0755**	-0.111***	-0.0610**
	(0.0273)	(0.0233)	(0.0212)	(0.0201)
lag exited labor market	0.0799***	-0.0327*	0.0472***	0.0291*
	(0.0169)	(0.0141)	(0.0127)	(0.0127)
got married	0.200***	0.0526**	0.252***	0.117***
	(0.0231)	(0.0195)	(0.0176)	(0.0182)
marriage ended	-0.185***	-0.0398	-0.225***	-0.306***
	(0.0311)	(0.0285)	(0.0282)	(0.0234)
stayed married	-0.00838	-0.00995*	-0.0183**	-0.181***
	(0.00723)	(0.00399)	(0.00670)	(0.0138)
lag got married	0.200***	-0.0366	0.163***	0.152***
	(0.0235)	(0.0197)	(0.0182)	(0.0179)
lag marriage ended	-0.0867**	0.0967***	0.0100	-0.0452*
	(0.0311)	(0.0263)	(0.0233)	(0.0220)
Constant	0.809***	-0.0493**	0.760***	0.766
	(0.0265)	(0.0166)	(0.0234)	(1.390)
Year dummies	Yes	Yes	Yes	Yes
Observations	84112	84112	84112	84112
R^2	0.036	0.005	0.064	0.020
Adjusted R^2	0.035	0.004	0.064	-0.259
AIC	209865.3	184980.1	157105.5	118102.5
BIC	210108.1	185222.9	157348.3	118336.0

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.6: GESIS: Regressions on key socio-economic variables. Variable description in Appendix C.1. Results from pooled OLS and fixed effects regressions including year fixed effects (not reported).

	(1) deviation OLS	(2) signdLM OLS	(3) signRM OLS	(4) signRM FE
female	-0.00509	-0.00523	-0.0103	0
	(0.0247)	(0.0191)	(0.0195)	(.)
age	-0.0252**	0.00475	-0.0205**	0.0460
	(0.00821)	(0.00648)	(0.00630)	(0.0704)
age2	0.000167*	-0.0000501	0.000117	-0.000501
	(0.0000847)	(0.0000664)	(0.0000650)	(0.000664)
change_fin_situation	0.232***	0.119***	0.351***	0.304***
	(0.0163)	(0.0144)	(0.0123)	(0.0194)
L.change_fin_situation	0.152***	-0.0937***	0.0584***	0.0247
	(0.0161)	(0.0140)	(0.0112)	(0.0180)
stayed unemployed	0.165	-0.274	-0.109	-0.0599
	(0.212)	(0.161)	(0.131)	(0.289)
stayed out of labour market	-0.0748	0.0112	-0.0635	0.0263
	(0.0425)	(0.0340)	(0.0336)	(0.129)
got employed	-0.0336	-0.0159	-0.0496	-0.0710
	(0.0738)	(0.0603)	(0.0503)	(0.104)
got unemployed	-0.0671	0.0510	-0.0161	-0.00631
	(0.138)	(0.116)	(0.0861)	(0.161)
exited labor market	0.0498	0.0653	0.115*	0.0966
	(0.0673)	(0.0548)	(0.0470)	(0.0773)
lag got employed	0.0536	0.0225	0.0761*	0.0969
	(0.0621)	(0.0526)	(0.0388)	(0.0685)
lag got unemployed	0.0812	-0.0521	0.0292	-0.146
	(0.177)	(0.137)	(0.105)	(0.164)
lag exited labor market	0.0815	0.0251	0.107*	0.0878
	(0.0665)	(0.0559)	(0.0470)	(0.0757)
got married	-0.00950	0.0972	0.0877	-0.0392
	(0.113)	(0.100)	(0.0748)	(0.135)
marriage ended	-0.0725	0.0743	0.00175	-0.128
	(0.132)	(0.103)	(0.0881)	(0.159)
stayed married	0.0364	-0.0332	0.00315	-0.345
	(0.0294)	(0.0228)	(0.0238)	(0.181)
lag got married	-0.0174	0.0785	0.0611	0.0527
	(0.0952)	(0.0836)	(0.0694)	(0.101)
lag marriage ended	-0.0332	0.0957	0.0625	0.0652
	(0.104)	(0.0914)	(0.0745)	(0.115)
Year dummies	yes	yes	yes	yes
Constant	-0.108	-0.195	-0.303	-1.509
	(0.201)	(0.163)	(0.159)	(1.841)
Observations	4864	4864	4864	4864
R^2	0.107	0.022	0.257	0.127
Adjusted R^2	0.103	0.018	0.254	-0.998
AIC	13106.7	11680.3	9219.6	4058.3
BIC	13236.5	11810.0	9349.4	4175.2

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.7: Deviation between measures on direction of change and trends in financial situation (BHPS). Description of variables given in Appendix C.1. Results from pooled OLS regression including year fixed effects.

	(1)	
	deviation OLS	
female	0.0182**	(0.00594)
age	-0.0206***	(0.00105)
age2	0.000155***	(0.0000104)
financial deterioration	-0.0438***	(0.00907)
Lag financial deterioration	-0.102***	(0.00968)
financial improvement	0.134***	(0.00957)
Lag financial improvement	0.0783***	(0.00903)
negative trend	-0.0462**	(0.0157)
positive trend	0.0442**	(0.0146)
stayed unemployed	-0.116***	(0.0320)
stayed out of labour market	-0.0409***	(0.00851)
got employed	0.0489**	(0.0169)
got unemployed	-0.0986***	(0.0227)
exited labor market	0.0668***	(0.0159)
lag got employed	0.0209	(0.0147)
lag got unemployed	-0.150***	(0.0253)
lag exited labor market	0.111***	(0.0162)
got married	0.189***	(0.0227)
marriage ended	-0.150***	(0.0297)
stayed married	-0.00147	(0.00678)
lag got married	0.192***	(0.0229)
lag marriage ended	-0.0453	(0.0284)
Constant	0.624***	(0.0265)
Year dummies	Yes	
Observations	83533	
R^2	0.053	
Adjusted R^2	0.052	
AIC	206936.4	
BIC	207216.4	

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

of the answer scale and thus have no further opportunity of indicating improvement or deterioration respectively without adjustments in the reporting function. This indicates that adaptation of reporting functions starts before the bounds of the answer scale are reached.

Combined, our results provide supporting evidence for systematic adaptation of the reporting function and hence systematic distortion in the representation of preference rankings when using the LM.[27]

4.6 Conclusion and Discussion

Several conceptual concerns raise doubt that the established level-measure of life satisfaction is suitable to elicit reliable preference rankings over states of the world. The adjustment of preferences and, more importantly, of the way they are mapped onto the discrete and bound answer scale via ad hoc reporting functions reduce the comparability of answers given by survey panel members in subsequent years. Empirical evidence from three representative surveys covering two countries, almost two decades and well over a hundred thousand observations give credence to these concerns. An alternative elicitation method, the ranking measure of life satisfaction is introduced and compared with the level measure.

The key empirical findings are that survey participants are largely unable to remember answers given to life satisfaction questions three months ago but are able to reassess the situation they experienced using a new reporting function. The noise introduced by imprecise recollection of previous answers reverts the preference ranking elicited by the level measure in more than half of all cases. The ranking measure raises consistency rates by more than 50 percent compared to the level measure in a controlled experiment. In three representative panel surveys the change in the level measure and the ranking measure are only weakly correlated and significantly differ in their means and distributions. These differences are systematically correlated with key socio-economic variables such as age, changes in the financial situation, employment and marital status.

[27] For evidence that the LM is unreliable in eliciting estimates of the marginal rate of substitution between different attributes that are complementary to those presented here, see Benjamin et al. (2014a).

The level measure but not the ranking measure systematically features adaptation to changes in life's circumstances and especially so when there is a trend in the underlying socio-economic variable.

Combining the conceptual concerns and the empirical evidence and previous findings in the literature (e.g. Benjamin et al. 2014a, Bond and Lang 2014) suggests that the level measure of life satisfaction produces biased estimates of preference rankings. In particular the widely reported adaptation to income changes and other aspects of life seems to be mainly driven by adjustments in the reporting function rather than by changes in the underlying preferences. This challenges at least some interpretations of key results from the life satisfaction literature such as the Easterlin Paradox. Levels of life satisfaction have remained mostly flat in developed countries despite substantial increases in average income over the last decades. This might be an adequate description of the satisfaction sensation in the population, but the results presented here call into question that this implies an indifference towards increases in per capita income in the long run. Depending on which concept policy makers decide to target, recommendations will be very different. According to ex-post preferences, increases in income do matter, also in the long run, whereas for long-run satisfaction intensities this seems not to be the case.

The reference to the Easterlin Paradox raises another issue. The level measure allows to compare satisfaction levels at very different points in time and to plot charts in terms of cardinal levels. However, given that the level measure is a poor proxy for preference rankings even for two subsequent years, it is important to exercise great care in interpreting these charts, i.e. not to take them as representations of preferences. The ranking measure does not allow drawing the same types of charts. This draws attention to the fact that the ordinal concept of preference rankings requires a direct comparison of two states of the world. Hence, if states five or ten years apart are to be compared, corresponding questions need to be included in the survey. However, the issue of imperfect and potentially biased recollection of past states of the world might then be a more serious issue. The level measure obscures problems with comparability (even for subsequent periods) by using an absolute (cardinal) scale for a relative (ordinal) concept. While such comparisons might be enlightening for some research questions and policy issues, their use and interpretation requires careful judgment by the analyst as not to be taken as indicators of preference rankings. The latter seem to be more reliably captured by the ranking measure of life satisfaction.

Knowing how people evaluate changes they have experienced is important for policy evaluation since this is both an essential input for many conventional welfare measures and, more practically, it is valuable information for policymakers in order to evaluate the acceptance of policy changes. In our view, the ranking measure is well suited to provide such insights.

Future research could compare the level and the ranking measure of life satisfaction in more detail shedding light on the differences in drivers, developments over time, and especially marginal rates of substitution for public goods (e.g. as in Anderson et al. 2016, Frey et al. 2004, Kopmann and Rehdanz 2013, Levinson 2012). This will further identify where the two measures and interpretations of life satisfaction differ and where a clear political decision for one of the target concepts needs to be made.

While we have focused attention on elicitation of preferences using life satisfaction data, many of the points raised and the potential to address them by using a direct measure of changes rather than within or between subject comparisons of levels, can also be transferred to many other variables elicited using Likert scales in (panel) surveys.

4.7 References

Anderson, M. L., Lu, F., Zhang, Y., Yang, J., and Qin, P. (2016). Superstitions, street traffic, and subjective well-being. *Journal of Public Economics*, 142, 1-10.

Bateman, I., Day, B., Loomes, G., and Sugden, R. (2007). Can ranking techniques elicit robust values? *Journal of Risk and Uncertainty*, 34(1), 49-66.

Benjamin, D. J., Heffetz, O., Kimball, M.S., and Rees-Jones, A. (2012). What do you think would make you happier? What do you think you would choose? *American Economic Review*, 102(5), 2083-2110.

Benjamin, D. J., Heffetz, O., Kimball, M. S., and Rees-Jones, A. (2014a). Can marginal rates of substitution be inferred from happiness data? Evidence from residency choices. *American Economic Review*, 104(11), 3498-3528.

Benjamin, D. J., Kimball, M. S., Heffetz, O., and Szembrot, N. (2014b). Beyond happiness and satisfaction: Toward well-being indices based on stated preference. *American Economic Review*, 104(9), 2698-2735.

Berridge, K. C., and Kringelbach, M. L. (2011). Building a neuroscience of pleasure and well-being. *Psychology of Well-Being: Theory, Research and Practice*, 1(1), 3.

Bertrand, M., and Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *American Economic Review*, 91(2), 67-72.

Camerer, C., Loewenstein, G., and Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, 43(1), 9-64.

Bond, T. N., and Lang, K. (2014). The sad truth about happiness scales (No. w19950). National Bureau of Economic Research.

Clark, A. E., D'Ambrosio, C., and Ghislandi, S. (2016). Adaptation to poverty in long-run panel data. *Review of Economics and Statistics*, 98(3), 591-600.

Clark, A. E., Frijters, P., and Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, 46(1), 95-144.

Conti, G., and Pudney, S. (2011). Survey design and the analysis of satisfaction. *Review of Economics and Statistics*, 93(3), 1087-1093.

Daly, M. C., and Wilson, D. J. (2009). Happiness, unhappiness, and suicide: An empirical assessment. *Journal of the European Economic Association*, 7(2-3), 539-549.

Daly, M. C., Wilson, D. J., and Johnson, N. J. (2013). Relative status and well-being: evidence from US suicide deaths. *Review of Economics and Statistics*, 95(5), 1480-1500.

Diener, E., Lucas, R. E., and Scollon, C. N. (2006). Beyond the hedonic treadmill: revising the adaptation theory of well-being. *American Psychologist*, 61(4), 305.

Di Tella, R., Haisken-De New, J., and MacCulloch, R. (2010). Happiness adaptation to income and to status in an individual panel. *Journal of Economic Behavior and Organization*, 76(3), 834-852.

Di Tella, R., and MacCulloch, R. (2006). Some uses of happiness data in economics. *Journal of Economic Perspectives*, 20(1), 25-46.

Di Tella, R., MacCulloch, R. J., and Oswald, A. J. (2003). The macroeconomics of happiness. *Review of Economics and Statistics*, 85(4), 809-827.

Decancq, K., Fleurbaey, M., Schokkaert, E. (2015). Happiness, equivalent incomes, and respect for individual preferences, *Economica* 82: 1082-1106.

De Neve, J., Ward, G., De Keulenaer, F., Van Landeghem, B., Kavetsos, G., and Norton, M. I. (2017). The Asymmetric Experience of Positive and Negative Economic Growth: Global Evidence Using Subjective Well-Being Data. *Review of Economics and Statistics*, DOI: 10.1162/REST_a_00697

Dolan, P., Peasgood, T., and White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology*, 29(1), 94-122.

Easterlin, R. (1974). Does economic growth improve the human lot? Some empirical evidence. In P. David and M. Reder (Eds.), *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz* (pp. 98-125). New York and London: Academic Press.

Easterlin, R. A., McVey, L. A., Switek, M., Sawangfa, O., and Zweig, J. S. (2010). The happiness–income paradox revisited. *Proceedings of the National Academy of Sciences*, 107(52), 22463-22468.

Fehr, E., and Rangel, A. (2011). Neuroeconomic foundations of economic choice — recent advances. *Journal of Economic Perspectives*, 25(4), 3-30.

Ferrer-i-Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of Public Economics*, 89(5), 997-1019.

Fleurbaey, M. (2009). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature*, 47(4), 1029-1075.

Fleurbaey, M. and Schwandt, H. (2016). Do People Seek to Maximize Their Subjective Well-Being – And Fail?. Paper presented at the Annual Meeting of the American Economic Association, Chicago, 6th Jan. 2017, https://www.aeaweb.org/conference/2017/preliminary/paper/zYFd2Rdd [17th August 2017].

Frederick, S., and Loewenstein, G. (1999). Hedonic Adaptation. in Kahneman, Diener, and Schwarz (eds.) *Well-being: The foundations of hedonic psychology*, New York: Russell Sage Foundation, 302-29.

Frey, B. S., Luechinger, S., and Stutzer, A. (2004). Valuing public goods: The life satisfaction approach. *CESifo Working Paper*, (No. 1158).

Frey, B. S., and Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature*, 40(2), 402-435.

GESIS (2017). GESIS Panel - Standard Edition. GESIS Data Archive, Cologne. ZA5665 Data file Version 19.0.0, doi:10.4232/1.12743

Gil, J., and Mora, T. (2011). The determinants of misreporting weight and height: the role of social norms. *Economics & Human Biology*, 9(1), 78-91.

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*, 19(3), 403-420.

Gunasekara, F. I., Carter, K., and Blakely, T. (2012). Comparing self-rated health and self-assessed change in health in a longitudinal survey: Which is more valid?. *Social Science and Medicine*, 74(7), 1117-1124.

Hoffrage, U., Hertwig, R., and Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 566.

Jones, Charles I. and Peter J. Klenow. (2016)."Beyond GDP? Welfare across Countries and Time." *American Economic Review*, 106(9): 2426-57.

Karney, B. R., and Coombs, R. H. (2000). Memory bias in long-term close relationships: Consistency or improvement?. *Personality and Social Psychology Bulletin*, 26(8), 959-970.

Kopmann, A., and Rehdanz, K. (2013). A human well-being approach for assessing the value of natural land areas. *Ecological Economics*, 93, 20-33.

Köszegi, B., and Rabin, M. (2008). Choices, situations, and happiness. *Journal of Public Economics*, 92(8), 1821-1832.

Kristoffersen, I. (2017). The metrics of subjective wellbeing data: An empirical evaluation of the ordinal and cardinal comparability of life satisfaction scores. *Social Indicators Research*, 130(2), 845-865.

Krueger, A. B., and Schkade, D. A. (2008). The reliability of subjective well-being measures. *Journal of Public Economics*, 92(8), 1833-1845.

Layard, R. (2006). Happiness and public policy: A challenge to the profession. *Economic Journal*, 116(510), C24-C33.

Layard R., G. Mayraz, S. Nickell (2008). The marginal utility of income, *Journal of Public Economics*, 92, 1846 -1857.

Levine, L. J., and Safer, M. A. (2002). Sources of bias in memory for emotions. *Current Directions in Psychological Science*, 11(5), 169-173.

Levinson, A. (2012). Valuing public goods using happiness data: The case of air quality. *Journal of Public Economics*, 96(9), 869-880.

Liberini, F., Redoano, M., and Proto, E. (2017). Happy voters. *Journal of Public Economics*, 146, 41-57.

Loewenstein, G., and Ubel, P. A. (2008). Hedonic Adaptation and the Role of decision and Experience Utility in Public Policy. *Journal of Public Economics*, 92(8–9): 1795–1810.

Luttmer, Erzo F. P. (2005). Neighbors as Negatives: Relative Earnings and Well-Being. Quarterly Journal of Economics, 120(3), 963-1002.

MacKerron, G. (2012). Happiness economics from 35 000 feet. *Journal of Economic Surveys*, 26(4), 705-735.

Morewedge, C. K., Gilbert, D. T., and Wilson, T. D. (2005). The least likely of times: How remembering the past biases forecasts of the future. *Psychological Science*, 16(8), 626-630.

Oswald, A. J. (1997). Happiness and economic performance. *Economic Journal*, 107(445), 1815-1831.

Oswald, A. J. (2008). On the curvature of the reporting function from objective reality to subjective feelings. *Economics Letters*, 100(3), 369-372.

Oswald, A. J., and Wu, S. (2010). Objective confirmation of subjective measures of human well-being: Evidence from the USA, *Science*, 327(5965), 576-579.

Perez-Truglia, R. (2015). A Samuelsonian validation test for happiness data. *Journal of Economic Psychology*, 49, 74-83.

Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, 86(4), 777-803.

Rayo, L., and Becker, G. S. (2007a). Evolutionary efficiency and happiness. *Journal of Political Economy*, 115(2), 302-337.

Rayo, L., and Becker, G. S. (2007b). Habits, peers, and happiness: An evolutionary perspective. *American Economic Review*, 97(2), 487-491.

Schroeder, C., and Yitzhaki, S. (2017). Revisiting the evidence for cardinal treatment of ordinal variables. *European Economic Review*, 92, 337-358.

Sen, A. (1985). *Commodities and Capabilities.* North-Holland, Amsterdam.

Stevenson, B., and Wolfers, J. (2008). Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox. *Brookings Papers on Economic Activity*, 39(1), 1-102.

Stevenson, B., and Wolfers, J. (2013). Subjective well-being and income: Is there any evidence of satiation? *American Economic Review*, 103(3), 598-604.

University of Essex. Institute for Social and Economic Research (2010). British Household Panel Survey: Waves 1-18, 1991-2009. [data collection]. 7th Edition. UK Data Service. SN: 5151, http://doi.org/10.5255/UKDA-SN-5151-1

Vendrik, M. C. (2013). Adaptation, anticipation and social interaction in happiness: An integrated error-correction approach. *Journal of Public Economics*, 105, 131-149.

Chapter 5

Valuation of non-market goods: the life satisfaction approach under methodological examination

5.1 Introduction

In perfectly competitive markets, individual decisions lead to quantities that equate marginal cost and marginal benefits and thus to efficiency. In cases of market failure, such as the presence of a public good, however, market outcomes are inefficient. Efficiency then hinges on policy intervention to measure and equate social marginal cost and social marginal benefits of the respective goods. To this end, different valuation methods have been developed over the last decades (see e.g. Freeman III 2003). The resulting estimates of marginal cost and marginal benefits can be used to identify the socially optimal provision of a public good using in a cost benefit-analysis (see e.g. Atkinson and Mourato 2006), on optimal regulation (Palmquist and Smith 2002) such as a Pigouvian tax (Pigou 1932, Mas-Colell et al. 2006) or to determine damages in litigation (Palmquist and Smith 2002, Carson et al. 2003).

In this paper, I make a methodological contribution to the life satisfaction approach (LSA) of valuation (see e.g. Welsch and Ferreira 2014). The LSA uses subjective reports of life satisfaction (LS) to determine the marginal rate of substitution between income and the non-market good of interest. It thus attempts to captures the amount of income that people are willing to give up in order to gain a certain amount of the non-market good while keeping life satisfaction constant and can thus be interpreted as marginal willingness to pay (MWTP).[1] The LSA has been used to value a diverse set of non-market goods: environmental goods and services, social interactions, health and many more. Welsch and Kühling (2009) provide an early methodological discussion and Frey et al. (2010) an early review. More recent reviews and discussions are provided by Welsch and Ferreira (2014) and Fleming and Ambrey (2017). The LSA is the youngest valuation method in the valuation toolbox. It differs from *stated preference methods* that use statements of individuals about hypothetical choices. It is also different from *revealed preference methods* which observe behavior on markets for goods that are either substitutes or complements to the non-market good in question in order to indirectly infer its value.

Building on the findings in Chapter 4, I suggest using a different measure of life satisfaction in the LSA that is better suited to capture preference rankings than the one currently used. As Chapter 4 demonstrates, the currently used level measure (LM) of life satisfaction has an inherent tendency towards adaptation due to the bounded scale on which it is elicited and the lack of objective units in which satisfaction is measured[2]. This induces adaptation of the reporting function over time and leads to systematic biases in the representation of preference rankings, possibly even to preference reversals. This, in consequence, also biases the inferred marginal rate of substitution in the LSA. The ranking measure (RM) of life satisfaction, instead, directly asks survey respondents for changes in their satisfaction and thus for a preference ranking of the current against a previous situation. It thus uses the same reporting function for two consecutive periods and avoids the ensuing biased representation of preference rankings.

Previous studies on the LSA and on LS more generally have argued that the LM is a

[1]The term non-market good describes goods that are not traded on markets due to market failure and are thus not efficiently provided through markets.

[2]The LM is typically elicited by asking respondents to state their level of satisfaction on a categorical scale, e.g. ranging from 1 “completely unsatisfied” to 7 “completely satisfied”.

valid and reliable measure of utility based on internal and external consistency, validity, reliability, and stability over time (e.g. Diener et al. (1999 and 2013), Welsch and Kühling (2009)). This, however, only shows that the LM correlates in a predictable and sensible way with other variables. It avoids a clear definition of what is meant by utility. In the economics literature, utility typically is a representation of the ordinal concept of preference rankings. The LM would thus need to capture preference rankings if utility in the traditional economic sense is meant. As Bond and Lang (2014) have shown, one has to rely on strong assumptions when assuming that preference rankings are represented through LS data such that preference reversals are to be expected. The results of Chapter 4 show that adaptation to new situations is only observed for the LM but not for the RM. This suggests that the LM biases the representation of preference rankings in spite of the validity and reliability results.[3]

Valuation studies that use the LSA with the LM, employ diverse econometric models. When using pooled OLS or ordered models, which is chosen in the majority of LSA valuation studies, this assumes interpersonal and intertemporal comparability of satisfaction levels since both between- and within-person variation in the LM is used. For the preference ranking interpretation of the LM, interpersonal comparison of satisfaction levels, however, makes no sense, since preferences are ordinal and individual-specific. Interpersonal comparison can be avoided by using fixed effects estimation methods which use only within person variation. This requires panel data which are not often available which is why it is only used in a minority of studies.[4] The problem of unobservable adaptation in the reporting function in intertemporal comparison however remains even with fixed effects models. This can be resolved by the RM. One structural reason why the RM can keep the reporting function constant in intertemporal comparison is that it directly asks for changes in satisfaction and thus uses variation in the structure of first differences instead of computing differences to within-person means as in the fixed effects model. In order to properly compare the valuation results of the LM and the RM, a first difference model of the LM needs to be estimated. In my analysis I therefore report valuation results from pooled OLS, fixed effects, first differences for the LM and

[3]This does not exclude that the LM is a good representation for other concepts of well-being, e.g. it might be a good measure for sensation intensity. As the work of Rayo and Becker (2007a, 2007b) indicates, sensation intensity is likely to be different from preference rankings in many situations.

[4]Fixed effects models have been used e.g. by Carroll et al. 2009, Finkelstein et al. 2013, Levinson 2012, Luechinger 2009, Powdthavee 2008, Rehdanz and Maddison 2005, Krekel et al. 2016.

regressions of the RM on first differences of explanatory variables.

This paper develops the theoretical foundation for using the RM for valuation of non-market goods in the LSA. Then, the valuation results of the LM and the RM are empirically compared using the British Household Panel Survey (BHPS, University of Essex 2010), the only dataset that contains the RM for a sequence of twelve years.[5] A sequence of empirical models allows to isolate the effects of person fixed effects in the LM, the effects of reducing variation from time-demeaned variables to first differences and finally the measurement effect when switching from LM to RM.

The findings are that when moving from pooled OLS, to fixed effects, to first differences with the LM to the RM, the inferred MWTP is reduced to an increasingly realistic level. This is demonstrated for social interaction with friends and neighbours[6] and for health status as non-market goods. The income variables are monthly household income or monthly personal income. The coefficients of yearly income are not significant for first differences and the RM and thus also the inferred MWTP is not significant. Other coefficients in the regression however are robust to using different income variables. The valuation results thus depend on the income variable chosen.

The results indicate that it is worthwhile discussing the importance of person fixed effects in existing results that use cross sectional data. They also imply that the lower and more realistic size of valuation results with the RM than with the LM (fixed effects model) is not only due to the lower variation used in the RM regression but also the different measure of life satisfaction used (when comparing the LM fixed effects, the LM first differences and the RM results).

The remainder of the paper is structured as follows: Section 5.2 presents the theoretical foundation of extending the LSA to being used with the RM. Section 5.3 presents the empirical comparison of the LM and the RM in a valuation study. Section 5.4 concludes.

[5]Surprisingly, to the best of my knowledge, the RM has never been used before, except in Chapter 4.
[6]Following Powdthavee (2008).

5.2 The Life Satisfaction Approach to Valuation

5.2.1 The level measure of life satisfaction

The LSA assumes that the reported level of satisfaction ($LM_{i,t}$) by individual i in period t is a function of the amount of the non-market good ($G_{i,t}$) that is available, of income ($Y_{i,t}$), of other socio-economic control variables ($X_{i,t}$) and preferences of i over different situations ($\succsim_{i,t}$) which can possibly change over time. Thus $f_{i,t}(.)$ is a reporting function relating preferences for a given situation to reported satisfaction:

$$LM_{i,t} \ = \ f_{i,t} \ (G_{i,t}, \ Y_{i,t}, \ X_{i,t}, \ \succsim_{i,t}). \tag{5.1}$$

The LM of life satisfaction is typically elicited in surveys, where participants are asked to state their level of satisfaction on a categorical scale. The categories are either number coded or labelled or both. The BHPS, which will be used in the next section, asks the question "[...] how dissatisfied or satisfied are you with your life overall?" and the 7 answer categories range from 1 "Not satisfied at all" to 7 "Completely satisfied" (University of Essex 2010).

The valuation of the non-market good then uses the partial effects of income and of the non-market good on LM. The implicit marginal willingness to pay for the non-market good is derived as follows:

$$MWTP_G = \frac{dY}{dG}|_{dLM=0} = \frac{\partial LM/\partial G}{\partial LM/\partial Y}. \tag{5.2}$$

It represents the marginal rate of substitution between income and the non-market good that keeps LM constant. This is the ratio of a change in Y and a change in G that offset each other such that there is no change in LM. It thus (approximately) measures how much income a person is willing to give up in order to increase the availability of the non-market good by one unit.[7]

In order to compute the $MWTP_G$ from equation 5.2, the function of equation 5.1 is

[7] Non-marginal changes of G can be valued using the compensating surplus (CS) or equivalent surplus (ES). This is discussed e.g. in Ambrey and Fleming (2011) and Welsch and Ferreira (2014). This will lead to different values per unit of the non-market good in case LM is not a linear function of G and Y.

estimated using for example the following linear specification:

$$LM_{i,t} = a + b_G\ G_{i,t} + b_Y\ Y_{i,t} + b_X\ X_{i,t} + v_{i,t} \qquad (LM). \tag{5.3}$$

The error term $v_{i,t}$ captures both the person fixed effect e_i that includes all person specific effects that have not been controlled for by the control variables, and an ideosyncratic error $\epsilon_{i,t}$: $v_{i,t} = e_i + \epsilon_{i,t}$.

For the simple linear specification in equation 5.3, the $MWTP_G$ is simply the ratio of the coefficients of G and Y: $MWTP_G = \frac{b_G}{b_Y}$.

$MWTP_G$ is computed by keeping $dLM = 0$. The question then is, what exactly is to be held constant here, i.e. which concept the LM is supposed to measure. The literature on the LSA so far has assumed that the LM is a good proxy for "utility" or "individual well-being"[8] based on validity and reliability results that show that the LM varies in a statistically significant, predictable and intuitive way with other variables such as smiling, suicide, being married, having a job and so on[9]. This, however, shows only that there is a relation but not that interpersonal and intertemporal comparison of the LM reliably capture preferences. This however would be necessary in order to get a correct $MWTP_G$ that keeps respondents indifferent to the simultaneous change in G and Y such that it can be interpreted as willingness to pay for the change in G.

If equation 5.3 is estimated with pooled OLS (or ordered models), this assumes intertemporal and interpersonal comparability of LM. Intertemporal comparability of LM, however, is hampered by unobservable adaptation of the reporting function (Köke and Perion 2017). The reason is that a natural way of fitting the open scaled and ordinal concept of preferences on the bounded and categorical scale of the questionnaire, is to adapt the reporting function to changing situations (Rayo and Becker 2007a,b). Interpersonal comparability is not defined in the concept of preference rankings which is purely ordinal and person specific and should thus be avoided when inferring $MWTP_G$.

In order to avoid interpersonal comparison of LM, fixed effects models or first differenced models can be used. In order to add intertemporal comparability, the ranking measure

[8] See e.g. Levinson 2012, Ambrey and Fleming 2011, Ferreira and Moro 2010, Welsch and Ferreira 2014

[9] See e.g. Frey et al. 2009, Berridge and Kringelbach 2011, Camerer et al. 2005, Daly and Wilson 2009, Daly et al. 2013, Fehr and Rangel 2011, Diener et al. 2013

will be used in the next section.

The fixed effects model uses transformed variables that contain differences to within-person mean:

$$\widetilde{LM}_{i,t} = a - b_G \ \widetilde{G}_{i,t} + b_Y \ \widetilde{Y}_{i,t} + b_X \ \widetilde{X}_{i,t} + \epsilon_{i,t}, \qquad (LM\_FE) \tag{5.4}$$

where the $\widetilde{\ }$ indicates that the difference from the observed value to the within-person mean over all t for the respective variable was taken. This eliminates all time-invariant factors, e.g. the variable gender would typically drop out as well as the person fixed effect in the error term such that only the ideosyncratic error remains.

The LSA relies on a set of further assumptions and faces a set of further issues which have been extensively discussed in the literature.[10] Even though they are not directly addressed by using the RM instead of the LM, they are still important for understanding and correctly interpreting the results of the LSA. For instance, the measurement of the independent variables needs to be without bias and too much noise. Especially income is difficult to elicit properly in a survey since respondents might not know what their actual income is and because there are many different definitions of income, such that there is a danger of noise in the data but also of biases due to incentives to over- or understate true income. This has been discussed to be problematic especially since income coefficients have very often been found to be surprisingly small, leading to unrealistically large MWTP (e.g. Ferreira and Moro 2010, Bertram and Rehdanz 2015, Welsch and Ferreira 2014). While this might be due to measurement issues, it might as well be due to adaptation of the reporting function when income keeps increasing, thus distorting the relative coefficients of variables with and without (expected) trends (Chapter 4). A more objective measure of income, e.g. by matching LM data to tax data would help to alleviate the measurement issues of the income variable. In this paper I report results with different income variables in order to test robustness at least with the available data.

The model specification as in equations 5.1 and 5.3 neglects possible compensation mechanisms in the market that would (partly) offset the effect of changes in G on LS. For

[10] See e.g. Levinson 2012, Ambrey and Fleming 2011, Fleming and Ambrey 2017, Ferreira and Moro 2010, Frey et al. 2009, Welsch and Ferreira 2014, Smith 2008

example lower rents or higher wages might offset part of the negative effects of air pollution. Thus when estimating the MWTP for reducing air pollution, we would only capture the residual effect of the reduction in air pollution on LM. In order to capture the total WTP for pollution changes we would thus have to control for rent, income and other compensating mechanisms. For a discussion of this issue and also its relation to the hedonic pricing method of valuation see e.g. Ferreira and Moro (2010), Welsch and Ferreira (2014), Rehdanz and Maddison (2008). It is thus advisable for the interpretation of the MWTP estimates to explicitly state which compensation mechanisms have been controlled for in the analysis and have thus been accounted for.

The estimation techniques need to take into account possible person-fixed effects not only to avoid interpersonal comparison of preferences but also to capture interpersonal differences in unobserved independent variables.

Taking into account the ordinal nature of the dependent variable would require estimating ordered models. Those however cannot control for person fixed effects. However the results are often similar in ordered and pooled OLS models such that using FE models is more advisable (e.g. Ferrer-i-Carbonell and Frijters (2004)).

Exogeneity of the independent variables is important. They thus need to influence LM but not to be influenced themselves by LM. Especially for income this has been doubted and IV approaches have been used, e.g. using professional group average income instead of personal income (e.g. Levinson 2012, Luttmer 2005).

5.2.2 The ranking measure of life satisfaction

While the level measure of life satisfaction captures the current level of satisfaction and infers preferences from comparing those levels over time or individuals, the ranking measure directly captures the change in satisfaction over a specified period. It thus directly measures the preference ranking. The question asked in the BHPS is “Would you say that you are more satisfied with life, less satisfied or feel about the same as you did a year ago?”. The four answer options are “More satisfied”, “Less satisfied”, “About the same” and “Don’t know”. Analogously to equation 5.1 for the LM, the RM can be described as the following function:

$$RM_{i,t,t-1} = r_{i,t}\ (dG_{i,t,t-1},\ dY_{i,t,t-1},\ dX_{i,t,t-1},\ \succsim_{i,t}) \tag{5.5}$$

The reported change in satisfaction $RM_{i,t,t-1}$ by individual i from period $t-1$ to period t is a function of changes in the non-market good, income and other socio-economic control variables over the same time span ($dG_{i,t,t-1}$, $dY_{i,t,t-1}$, $dX_{i,t,t-1}$) and preferences as evaluated in period t ($\succsim_{i,t}$). Thus the reporting function $r_{i,t}(.)$ relates preferences over two consecutive situations to reported change in satisfaction.

Again, we aim at deriving the implicit willingness to pay for changes in G as in equation 5.2. The only difference is that instead of setting $dLM = 0$, now $RM = 0$ is required:

$$MWTP_G = \frac{dY}{dG}|_{RM=0} \tag{5.6}$$

The empirical model with the RM can be specified in equivalence to equation 5.3:

$$RM_{i,t,t-1} = a + b_G\ dG_{i,t,t-1} + b_Y\ dY_{i,t,t-1} + b_X\ dX_{i,t,t-1} + \epsilon_{i,t} \qquad (RM) \tag{5.7}$$

The change in satisfaction as measured by the RM is explained by changes in G, Y and X. As error term only the ideosyncratic error $\epsilon_{i,t}$ remains. The person specific fixed effect disappears since only changes over time are considered.[11]

In this simple linear case, the $MWTP_G$ is again simply the ratio of the coefficients of dG and dY: $MWTP_G = \frac{b_G}{b_Y}$.

The basic difference between LM and RM is that the RM is measuring first differences in satisfaction levels. It is thus equivalent in structure to taking the first difference in the level measure of satisfaction (dLM). Even more equivalent would be using the sign of the first differences of the LM (signdLM), since the RM also only captures positive, negative or neutral changes but not the size of changes [12]:

$$dLM_{i,t,t-1} = a + b_G\ dG_{i,t,t-1} + b_Y\ dY_{i,t,t-1} + b_X\ dX_{i,t,t-1} + \epsilon_{i,t} \qquad (LM\_FD) \tag{5.8}$$

[11] There might still be person fixed effects in *differences* instead of in *levels*.

[12] The relation of the RM to the LM is discussed more extensively in Chapter 4.

$$signdLM_{i,t,t-1} = a + b_G\, dG_{i,t,t-1} + b_Y\, dY_{i,t,t-1} + b_X\, dX_{i,t,t-1} + \epsilon_{i,t} \quad (LM\_signFD) \quad (5.9)$$

Thus using pooled OLS or FE models with the LM is not directly comparable to using the RM, since they use dimensions of variation not captured with the RM. OLS uses within and between subject variation and FE uses differences to within-person means instead of differences to the previous period in FD. Only in the case of datasets with two periods the FE and FD results would be the same.

This can be used to discuss differences in how well LM and RM are suited to measure $MWTP_G$. If the regressions with *signdLM* significantly differ from those with the RM, this must be due to different representations of preferences through both reporting functions $f(.)$ and $r(.)$ and thus be a measurement issue. Differences in results between LM_OLS, LM_FE and LM_FD stem from using different types of variation in the data.

The RM is able to better capture preference rankings than the LM since it makes sure that the reporting function is the same for two situations that are being compared. The "price" at which the more direct measurement of preference rankings has to be bought is that the RM only uses within person variation from one period to the next. It is thus an empirical question of how much the better measure will result in more clear-cut results.

5.3 Empirical Analysis

In this section, I empirically compare the valuation results of the LM and the RM of life satisfaction in the LSA. As data set, I use the BHPS (British Household Panel Survey) which provides a unique opportunity since it contains both the LM and the RM in a panel of twelve years.

My empirical analysis follows the study by Powdthavee (2008), who derives the value for the frequency of social interaction with neighbours and friends or family as the non-market good. The study was chosen as a starting point as it uses the BHPS and as it can easily be replicated because no additional location specific data needs to be merged with the BHPS as would be the case when valuing e.g. pollution or the availability of public goods or bads in general.

Using pooled OLS, fixed effects and ordered probit models, Powdthavee (2008) finds implicit values in the range of 37,000 – 85,000 1997 GBP per year for having the highest frequency ("on most days") of seeing neighbors or friends as compared to the base category of "less than once a month".[13] This seems quite large, even when considering that with four categories, this implies 9,000 to 21,000 GBP per year for improving by one category.

In addition to the value of social interaction I also infer the value of health status. In contrast to social interaction, health is to a much lesser degree subject to individuals' choices. Health is thus more exogenous and less prone to be compensated by other mechanisms than life satisfaction. One would thus expect to find that it is "easier" to measure a value for health than for social interaction.

5.3.1 Data

The data is drawn from the BHPS (University of Essex 2010). I use all available waves that contain the RM, which are those from 1997 to 2009 with 2001 missing.[14]

The *LM* of life satisfaction is measured on a scale from 1 to 7 (from "very unsatisfied" to "very satisfied"). The *RM* of life satisfaction is measured as "More satisfied", "Less satisfied" and "About the same". The frequency of social interaction (meeting friends and relatives outside respondent's own household and talking to neighbours) is measured in five categories ("never", "less often than once a month", "once or twice a month", "once or twice a week" and "on most days") where the lowest two categories are merged into one. *Health status* is measured in five categories, "very poor", "poor", "fair", "good" and "excellent". The number of *days in hospital* is used as an additional measure of health.

Income is a composite variable that has been composed of different income categories in the BHPS. Different income variables will be used in this paper. Personal and household

[13] His results can be found in Appendix D.
[14] This corresponds to waves F-J and L-Q. Powdthavee (2008) used the years 1997 to 2003.

income will be used, in real and nominal terms, with and without log- transformation. *Real household income per year and capita* means that income was aggregated at the household level, divided by the household size and corrected for inflation to 1997 GBP. It thus captures real household income per capita in 1997 GBP. Some control variables such as age, gender, marital status, employment status, education and region are included. The summary statistics are reported in in Table D.1 in Appendix D.

From the analysis were excluded those observations which had *Real household income per year and capita* larger than 100,000 GBP and less than 0 GBP (reducing the sample by 5,475 observations). Equivalently, those were excluded that had age changes over two consecutive periods of 0 or more than 1 (excluding 69 observations).[15]

Table 5.1: Frequency of social interaction with friends/family and neighbours

	Neighbours				
	Less than once a month	Once or twice a month	Once or twice a week	On most days	Total
Friends/Family					
Less than once a month	790	717	1,303	905	3,715
Once or twice a month	1,629	3,271	6,410	3,696	15,006
Once or twice a week	4,735	8,838	26,670	18,028	58,271
On most days	5,824	7,514	20,559	32,691	66,588
Total	12,978	20,340	54,942	55,320	143,580

BHPS 1997-2009

Table 5.2: Frequency of Health status and Days in hospital

	Days in hospital							
	0	0-5	5-10	10-50	50-100	100-200	200-365	Total
Health status								
very poor	1,859	456	298	578	93	42	7	3,333
poor	8,569	1,607	734	874	74	19	6	11,883
fair	27,386	2,758	839	636	55	15	2	31,691
good	60,066	3,306	615	344	16	4	0	64,351
excellent	30,989	1,120	150	62	1	0	0	32,322
Total	128,869	9,247	2,636	2,494	239	80	15	143,580

BHPS 1997-2009

Table 5.3: Frequency of signdLM and RM

	RM			
signdLM	worse	indifferent	improved	Total
worse	8,822	16,687	5,252	30,761
indifferent	5,562	32,037	14,337	51,936
improved	2,918	16,020	10,050	28,988
Total	17,302	64,744	29,639	111,685

BHPS 1997-2009

The conditional frequencies of social interaction with *friends/family* and *neighbours* can

[15] Powdthavee (2008) only considered individuals of working age between 16 an 65. Since it does not make much difference for the valuation results I included the full range of age from 16 to 100.

Table 5.4: Frequencies of Real household income per year and capita by LM

	LM							
	1	2	4	4	5	6	7	Total
Real hh inc. p.y/p.c.								
0-10	1,830	2,366	6,288	13,904	25,473	26,954	15,019	91,834
10-20	364	738	2,323	5,815	14,800	17,619	5,589	47,248
20-30	28	120	360	899	3,023	3,566	855	8,851
30-40	9	35	97	240	743	1,000	206	2,330
40-50	5	9	28	65	219	288	61	675
50-60	0	8	6	20	68	112	20	234
60-70	0	4	4	13	35	47	11	114
70-80	0	0	4	3	11	22	4	44
80-90	1	1	1	5	9	17	3	37
90-100	1	1	7	16	41	37	16	119
Total	2,238	3,282	9,118	20,980	44,422	49,662	21,784	151,486

BHPS 1997-2009

be found in Table 5.1, their correlation is 0.16. The conditional frequencies of *health status* and *days in hospital* (the latter bundled in categories) can be found in Table 5.2, their correlation is -0.2. The conditional frequencies of *signdLM* and *RM* can be found in Table 5.3, their correlation is 0.21. Table 5.4 shows the relation of *LM* and *Real household income per year and capita* (the latter bundled in categories), their correlation is 0.06. Further frequencies and histograms can be found in Table D.2, Figure D.1 and Figure D.2 in the Appendix.

5.3.2 Results

Table 5.5 presents the regression results of pooled OLS (1), fixed effect (2), first differences (3) and signed first differences (4) for the level measure as well as the RM regressed on first differences of the independent variables (5) (corresponding to equations 5.3, 5.4, 5.8, 5.9 and 5.7 from the theory section). The $MWTP_G$ are reported for G being social interaction with friends or neighbours, or health status or days in hospital.[16] Since the $MWTP_G$ is a ratio of two coefficients, I apply a test for non-linear combinations of estimates (based on the delta method) in order to determine its significance.[17] Income is chosen to be *Personal (nominal) income per month*. An overview of the valuation results when using different income variables are reported in Tables 5.6, 5.7 and 5.8.

[16]Different from Powdthavee (2008), I included the non-market goods as continuous instead of factor variables. This does not loose much information since the steps between the different categories were rather similar in his analysis. Also the interpretation seems more natural to think about the value of a change from one category to the next and not as compared to the extreme base category.

[17]This possibly was not done in the Powdthavee paper where the MWTP is reported to be significant as soon as the social interaction coefficient and the income coefficient are significant without explanation as to how significance is determined.

In all regressions year dummies are included. Regions are either accounted for by region dummies ((1) and (2)) or by a dummy that is one in case of a region change since the previous period ((3) to (5)). A change in region only took place in 1,647 observations (1,212 observations in the RM regression), thus in roughly 1% of observations.

Pooled OLS and Fixed Effects

When comparing regressions (1) and (2) in Table 5.5, the MWTP is lower in (2) for all G except for *Days in hospital* when considering only within person variation instead of within and between variation in (1). For both social interactions with friends and neighbours, the MWTP for going up by one category decreases from roughly 4,000 to 2,000 GBP per month and person. The MWTP for going up one category of health decreases from 18,000 GBP in the pooled OLS to 13,000 GBP in the FE model, again by month and person. The MWTP for an extra day in hospital changes from -160 GBP to -220 GBP per month and person (the willingness to accept a compensation for staying at a hospital increases thus). This is driven by a reduction by one third of the coefficients of social interaction, a reduction by one half of health status and less than one half of income. All of this leads to a reduction in the MWTP. For *Days in hospital*, the coefficient remains unchanged such that the reduced income coefficient leads even to a larger (absolute) MWTP. The single effects of income, health and social interaction are thus generally smaller in the FE estimation as compared to pooled OLS. Not controlling for person fixed effects has a considerable effect on the coefficients and the MWTP.[18]

The Breusch and Pagan test for random effects excludes the use of pooled OLS (test that Var(u)=0 is rejected with chi2=72071.12, p = 0.000). The F-test on personal fixed effects comes to the same conclusion (F test that all $u_i = 0$: F(25373, 118155) = 5.27 , p = 0.0000). The Hausman test favours the use of the FE model over the RE model (chi2(50) = 2755.72, p = 0.0000). The results of the FE model are thus to be favoured over pooled OLS and RE.

As in the Powdthavee paper, an ordered probit model was estimated in order to account for the ordinal character of the dependent variable. However, the results are not much

[18]Thus per year, the MWTP for one category of social interaction is 18,000 GBP with FE and 36,000 GBP with pooled OLS. This is even more than what Powdthavee found (on average over categories between 9,000 and 21,000 GBP per category).

different from pooled OLS. Even more importantly, a consistent fixed effects estimator is not available for ordered dependent variable models (Welsch and Kühling 2009, Cameron and Trivedi 1998), which would be necessary here. Also, the assumption of cardinality (using OLS instead of ordered models) was typically found not to change much about the relative size (in the MWTP) and significance of the coefficients (Ferrer-i-Carbonell and Frijters 2004). I will therefore focus on OLS regressions.

Result 1: *Person fixed effects are present such that the use of a fixed effects model is to be favored over a pooled OLS regression. The Hausman test favors the FE model over a RE model. The MWTP estimates in the FE model are lower than in the pooled OLS. This is driven by lower coefficients of both social interaction, health and income, and a smaller difference for the income coefficient. For days in hospital the coefficient remains similar.*

When interpreting the estimates of cross-sectional studies for political decision-making we thus need to be cautious. Person fixed effects can play a considerable role. If they are neglected, policy implications will be misguided. In our example, if implementing a policy to foster social interaction or income increases, the effect on life satisfaction would be smaller for people affected by the change than cross-section estimates would predict. The MWTP for social interaction however would be larger in cross sectional studies in this example, because the income coefficient is relatively underestimated compared to the coefficient on social interaction.

First differences and Ranking Measure

When further comparing the results with dLM, signdLM and RM in regressions (3), (4) and (5), the trend of reduced MWTP is continued for social interaction and health status. The results for days in hospital again are less clear. Both social interaction variables have a per month MWTP of roughly 1,300 GBP for dLM, 1,000 GBP for signdLM and 500 GBP for RM. Per year, this is then 15,600, 12,000 and 6,000 GBP. This seems much more realistic. While the income coefficient is rather similar in all three regressions, the coefficients of all G decrease.

For health status the MWTP is 8,500 (dLM), 6,500 (signdLM) and 4,000 (RM) GBP per month and person. Also this gets closer to a realistic number, even though it is still

Table 5.5: Valuation results with the level measure and the ranking measure of life satisfaction (BHPS 1997-2009)

	(1) OLS LM		(2) FE LM		(3) OLS dLM		(4) OLS sign(dLM)		(5) OLS RM	
Men	-0.0519***	(0.0122)	0	(.)						
Age	-0.0291***	(0.00218)	-0.0815**	(0.0315)						
Age squarred	0.000333***	(0.0000228)	-0.0000216	(0.0000270)						
Seeing friends and family	0.0940***	(0.00618)	0.0319***	(0.00429)						
Talking to neighbours	0.0990***	(0.00527)	0.0351***	(0.00385)						
Personal income per month / 1000	0.0236***	(0.00408)	0.0157***	(0.00366)						
Health status	0.434***	(0.00588)	0.207***	(0.00403)						
Days in hospital	-0.00390***	(0.000816)	-0.00347***	(0.000484)						
Education: O- or A-level	-0.0923***	(0.0152)	0.0627	(0.0379)						
Education: university level	-0.120***	(0.0178)	0.0350	(0.0464)						
Own house	0.109***	(0.0142)	0.0556***	(0.0120)						
Number of own children in household	-0.0786***	(0.00772)	-0.0123	(0.00682)						
Number of people in household	-0.00756	(0.00523)	-0.0224***	(0.00476)						
married	0.363***	(0.0205)	0.164***	(0.0215)						
living as couple	0.299***	(0.0194)	0.187***	(0.0180)						
widowed	-0.00101	(0.0339)	-0.205***	(0.0333)						
divorced	-0.156***	(0.0329)	-0.0780**	(0.0298)						
separated	-0.343***	(0.0413)	-0.257***	(0.0324)						
job: unemployed	-0.422***	(0.0273)	-0.296***	(0.0181)						
job: self-employed	0.0436*	(0.0194)	-0.00564	(0.0173)						
job: retired	0.123***	(0.0227)	0.0368*	(0.0166)						
job: student	0.120***	(0.0203)	0.0955***	(0.0195)						
job: disabled	-0.508***	(0.0351)	-0.377***	(0.0216)						
job: look after home	-0.0538*	(0.0220)	-0.0441**	(0.0153)						
D.Seeing friends and family					0.0246***	(0.00541)	0.0140***	(0.00327)	0.00791***	(0.00228)
D.Talking to neighbours					0.0235***	(0.00529)	0.0135***	(0.00309)	0.00830***	(0.00227)
D.Personal income per month / 1000					0.0179***	(0.00387)	0.0136***	(0.00271)	0.0157***	(0.00249)
D.Health status					0.151***	(0.00554)	0.0881***	(0.00318)	0.0632***	(0.00233)
D.Days in hospital					-0.00156	(0.000821)	-0.000434	(0.000426)	-0.00122***	(0.000303)
D.Education: O- or A-level					0.0882	(0.0569)	0.0267	(0.0354)	0.310***	(0.0304)
D.Education: university level					0.0560	(0.0729)	-0.0164	(0.0474)	0.466***	(0.0418)
D.Own house					0.00217	(0.0188)	0.00277	(0.0115)	-0.000360	(0.00883)
D.Number of own children in household					0.00824	(0.0137)	-0.00183	(0.00865)	0.160***	(0.00827)
D.Number of people in household					-0.0117	(0.00805)	-0.00915	(0.00480)	-0.0147***	(0.00414)
D.married					0.139***	(0.0367)	0.0926***	(0.0230)	0.452***	(0.0201)
D.living as couple					0.137***	(0.0295)	0.0863***	(0.0181)	0.235***	(0.0161)
D.widowed					-0.130	(0.0670)	-0.0260	(0.0364)	-0.0325	(0.0284)
D.divorced					-0.0758	(0.0519)	-0.0241	(0.0303)	0.0873**	(0.0266)
D.separated					-0.153**	(0.0551)	-0.0715*	(0.0315)	0.152***	(0.0272)
D.job: unemployed					-0.274***	(0.0269)	-0.142***	(0.0150)	-0.165***	(0.0120)
D.job: self-employed					0.0274	(0.0241)	0.0258	(0.0154)	0.0340**	(0.0125)
D.job: retired					-0.0109	(0.0265)	0.0159	(0.0154)	-0.0364**	(0.0115)
D.job: student					0.0581*	(0.0288)	0.0297	(0.0171)	-0.0803***	(0.0155)
D.job: disabled					-0.286***	(0.0393)	-0.125***	(0.0198)	-0.158***	(0.0152)
D.job: look after home					-0.0381	(0.0226)	-0.000846	(0.0134)	-0.0457***	(0.0101)
Region control / Change_region	Yes		Yes		0.130***	(0.0327)	0.0884***	(0.0201)	0.229***	(0.0212)
Year conrol	Yes		Yes		Yes		Yes		Yes	
Constant	3.350***	(0.0736)	7.530***	(1.261)	0.0602***	(0.0123)	0.0314***	(0.00761)	0.112***	(0.00653)
Observations	143580		143580		102661		102661		104272	
R^2	0.187		0.041		0.020		0.017		0.032	
Adjusted R^2	0.186		-0.165		0.020		0.017		0.032	
MWTP_freq_friends	3.980***	(0.726)	2.025***	(0.540)	1.371**	(0.417)	1.034***	(0.312)	0.503**	(0.165)
MWTP_freq_neighbours	4.193***	(0.751)	2.229***	(0.570)	1.310**	(0.409)	0.994**	(0.302)	0.528**	(0.167)
MWTP_healthstatus	18.36***	(3.190)	13.19***	(3.075)	8.435***	(1.852)	6.495***	(1.322)	4.013***	(0.652)
MWTP_days_in_hospital	-0.165***	(0.0449)	-0.220***	(0.0597)	-0.0871	(0.0495)	-0.0320	(0.0320)	-0.0775***	(0.0228)

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

far beyond what the average panel participant earns (median monthly income is roughly 1,000 GBP, the mean is roughly 1,300 GBP).

MWTP for days in hospital is not significant for dLM and signdLM, and is around -80 GBP per month for one additional day in hospital per year. First, it is important to remember that this number is calculated while keepint health status constant. And calculated per year, this is then 960 GBP, the average individual would need to be paid per extra day in hospital in order to be as satisfied as without the extra day in hospital.

It thus seems that not only accounting for person fixed effects but also the RM as a different measure help to reduce the MWTP from unrealistically high numbers to something that gets into the range of actual income budgets. It does so since the relative effect of income relative to that of the non-market good is increased. This is consistent with the idea that trending variables lead to adaptation of the reporting function of LM, thus keeping the income coefficient relatively small compared to coefficients of other non-trending variables.

Result 2: *When comparing the RM and the LM in the same structure of variation (first differences), the RM finds more realistic (as compared to available income) and at least as significant MWTP than dLM or signdLM.*

Tables 5.6, 5.7 and 5.8 report the results on MWTP for different income variables. They contain the b_Y, b_G, $MWTP_G$ and the number of observations for the regressions (1) to (5) with exactly the same control variables as in Table 5.5.

What is striking is the robustness of the coefficients b_G to different income specifications both in terms of size and significance. Also the trend of b_G decreasing from regression (1) to (5) is very robust to changes in income variables, except for *Days in hospital* which in some cases increases again from (4) to (5). Differences in MWTP are thus driven by the income coefficients. For both household and personal nominal income, the monthly measures are much more significant than yearly measures. When adjusting household income per year to per capita or needs adjusted measures, this does not help in finding significant income coefficients and thereby significant MWTP. Log personal income per year leads to a significant income coefficient and MWTP only for the RM. The MWTP calculated from log-income needs to be multiplied with the income of interest. Another

Table 5.6: Valuation results for different household income variables (BHPS 1997-2009)

	(1) LM OLS		(2) LM FE		(3) dLM OLS		(4) sign(dLM) OLS		(5) RM OLS	
Household income per month	0.0303***	(0.00256)	0.00873***	(0.00216)	0.00781**	(0.00266)	0.00655***	(0.00178)	0.00798***	(0.00142)
Seeing friends and family	0.0945***	(0.00616)	0.0317***	(0.00429)	0.0245***	(0.00541)	0.0140***	(0.00327)	0.00784***	(0.00228)
Talking to neighbours	0.101***	(0.00527)	0.0350***	(0.00385)	0.0234***	(0.00529)	0.0134***	(0.00309)	0.00825***	(0.00228)
Health status	0.431***	(0.00587)	0.207***	(0.00403)	0.151***	(0.00555)	0.0881***	(0.00318)	0.0632***	(0.00233)
Days in hospital	-0.00396***	(0.000816)	-0.00346***	(0.000484)	-0.00155	(0.000821)	-0.000426	(0.000426)	-0.00121***	(0.000303)
MWTP_freq_friends	3.113***	(0.330)	3.631***	(1.022)	3.132*	(1.267)	2.134**	(0.761)	0.982**	(0.335)
MWTP_freq_neighbours	3.314***	(0.324)	4.004***	(1.082)	2.995*	(1.221)	2.054**	(0.734)	1.033**	(0.340)
MWTP_healthstatus	14.22***	(1.224)	23.74***	(5.903)	19.34**	(6.620)	13.47***	(3.700)	7.910***	(1.436)
MWTP_days_in_hospital	-0.130***	(0.0292)	-0.396***	(0.113)	-0.198	(0.125)	-0.0651	(0.0676)	-0.152**	(0.0468)
Observations	143580		143580		102661		102661		104272	
Household income per year	0.00266***	(0.000235)	0.000360	(0.000205)	0.000102	(0.000297)	0.000231	(0.000180)	0.000372*	(0.000150)
Seeing friends and family	0.0945***	(0.00616)	0.0316***	(0.00429)	0.0245***	(0.00541)	0.0139***	(0.00327)	0.00781***	(0.00228)
Talking to neighbours	0.100***	(0.00526)	0.0349***	(0.00385)	0.0233***	(0.00529)	0.0134***	(0.00309)	0.00817***	(0.00228)
Health status	0.431***	(0.00587)	0.207***	(0.00403)	0.151***	(0.00555)	0.0882***	(0.00318)	0.0632***	(0.00233)
Days in hospital	-0.00396***	(0.000816)	-0.00346***	(0.000484)	-0.00156	(0.000820)	-0.000430	(0.000426)	-0.00121***	(0.000303)
MWTP_freq_friends	35.47***	(3.865)	87.75	(51.22)	239.5	(698.4)	60.28	(48.98)	21.00*	(10.45)
MWTP_freq_neighbours	37.64***	(3.806)	96.74	(55.97)	228.4	(666.0)	57.81	(46.96)	21.98*	(10.85)
MWTP_healthstatus	161.9***	(14.53)	575.6	(327.4)	1480.9	(4306.6)	381.1	(296.2)	170.0*	(68.68)
MWTP_days_in_hospital	-1.488***	(0.334)	-9.604	(5.621)	-15.26	(45.10)	-1.858	(2.347)	-3.263*	(1.549)
Observations	143580		143580		102661		102661		104272	
Household income per y. and capita	0.00629***	(0.000636)	0.00163**	(0.000530)	0.000544	(0.000744)	0.000633	(0.000462)	0.000594	(0.000390)
Seeing friends and family	0.0940***	(0.00616)	0.0317***	(0.00429)	0.0245***	(0.00541)	0.0140***	(0.00327)	0.00782***	(0.00228)
Talking to neighbours	0.100***	(0.00526)	0.0349***	(0.00385)	0.0233***	(0.00529)	0.0134***	(0.00309)	0.00819***	(0.00228)
Health status	0.432***	(0.00588)	0.207***	(0.00403)	0.151***	(0.00555)	0.0882***	(0.00318)	0.0632***	(0.00233)
Days in hospital	-0.00395***	(0.000817)	-0.00346***	(0.000484)	-0.00156	(0.000820)	-0.000430	(0.000426)	-0.00122***	(0.000302)
MWTP_freq_friends	14.95***	(1.793)	19.39**	(6.804)	44.95	(62.26)	22.03	(16.90)	13.17	(9.475)
MWTP_freq_neighbours	15.92***	(1.795)	21.38**	(7.312)	42.90	(59.43)	21.15	(16.24)	13.79	(9.870)
MWTP_healthstatus	68.67***	(7.060)	127.1**	(41.30)	277.9	(380.2)	139.2	(101.7)	106.5	(70.12)
MWTP_days_in_hospital	-0.628***	(0.145)	-2.119**	(0.749)	-2.862	(4.196)	-0.679	(0.839)	-2.048	(1.441)
Observations	143580		143580		102661		102661		104272	
Needs adjusted household income	0.00511***	(0.000450)	0.00113**	(0.000384)	0.000387	(0.000538)	0.000504	(0.000333)	0.000643*	(0.000280)
Seeing friends and family	0.0945***	(0.00616)	0.0317***	(0.00429)	0.0244***	(0.00541)	0.0140***	(0.00327)	0.00782***	(0.00228)
Talking to neighbours	0.100***	(0.00526)	0.0349***	(0.00385)	0.0233***	(0.00529)	0.0134***	(0.00309)	0.00819***	(0.00228)
Health status	0.431***	(0.00588)	0.207***	(0.00403)	0.151***	(0.00555)	0.0882***	(0.00318)	0.0632***	(0.00233)
Days in hospital	-0.00397***	(0.000817)	-0.00346***	(0.000484)	-0.00156	(0.000820)	-0.000429	(0.000426)	-0.00121***	(0.000303)
MWTP_freq_friends	18.48***	(2.010)	27.96**	(10.20)	63.22	(89.13)	27.69	(19.45)	12.16	(6.375)
MWTP_freq_neighbours	19.65***	(1.982)	30.81**	(10.98)	60.32	(85.04)	26.57	(18.67)	12.74	(6.606)
MWTP_healthstatus	84.39***	(7.568)	183.2**	(62.29)	390.9	(544.6)	175.0	(115.9)	98.35*	(42.93)
MWTP_days_in_hospital	-0.776***	(0.175)	-3.056**	(1.122)	-4.025	(5.993)	-0.852	(1.020)	-1.889*	(0.949)
Observations	143580		143580		102661		102661		104272	
Real household income p.y p.c	0.00706***	(0.000713)	0.00170**	(0.000596)	0.000631	(0.000830)	0.000724	(0.000516)	0.000469	(0.000436)
Seeing friends and family	0.0940***	(0.00616)	0.0316***	(0.00429)	0.0244***	(0.00541)	0.0140***	(0.00327)	0.00782***	(0.00228)
Talking to neighbours	0.100***	(0.00526)	0.0349***	(0.00385)	0.0233***	(0.00529)	0.0134***	(0.00309)	0.00819***	(0.00228)
Health status	0.432***	(0.00588)	0.207***	(0.00403)	0.151***	(0.00555)	0.0882***	(0.00318)	0.0632***	(0.00233)
Days in hospital	-0.00395***	(0.000817)	-0.00346***	(0.000484)	-0.00156	(0.000820)	-0.000430	(0.000426)	-0.00122***	(0.000302)
MWTP_freq_friends	13.31***	(1.595)	18.61**	(6.977)	38.74	(51.66)	19.27	(14.47)	16.67	(16.23)
MWTP_freq_neighbours	14.18***	(1.596)	20.52**	(7.523)	36.98	(49.32)	18.49	(13.91)	17.45	(16.95)
MWTP_healthstatus	61.17***	(6.279)	122.0**	(42.79)	239.5	(315.2)	121.8	(86.91)	134.8	(125.3)
MWTP_days_in_hospital	-0.560***	(0.129)	-2.034**	(0.767)	-2.467	(3.496)	-0.594	(0.728)	-2.595	(2.497)
Observations	143580		143580		102661		102661		104272	

Income is measured in 1000 GBP. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.7: Valuation results for different personal income variables (BHPS 1997-2009)

	(1) LM OLS		(2) LM FE		(3) dLM OLS		(4) sign(dLM) OLS		(5) RM OLS	
Personal income per month	0.0236***	(0.00408)	0.0157***	(0.00366)	0.0179***	(0.00387)	0.0136***	(0.00271)	0.0157***	(0.00249)
Seeing friends and family	0.0940***	(0.00618)	0.0319***	(0.00429)	0.0246***	(0.00541)	0.0140***	(0.00327)	0.00791***	(0.00228)
Talking to neighbours	0.0990***	(0.00527)	0.0351***	(0.00385)	0.0235***	(0.00529)	0.0135***	(0.00309)	0.00830***	(0.00227)
Health status	0.434***	(0.00588)	0.207***	(0.00403)	0.151***	(0.00554)	0.0881***	(0.00318)	0.0632***	(0.00233)
Days in hospital	-0.00390***	(0.000816)	-0.00347***	(0.000484)	-0.00156	(0.000821)	-0.000434	(0.000426)	-0.00122***	(0.000303)
MWTP_freq_friends	3.980***	(0.726)	2.025***	(0.540)	1.371**	(0.417)	1.034***	(0.312)	0.503**	(0.165)
MWTP_freq_neighbours	4.193***	(0.751)	2.229***	(0.570)	1.310**	(0.409)	0.994**	(0.302)	0.528**	(0.167)
MWTP_healthstatus	18.36***	(3.190)	13.19***	(3.075)	8.435***	(1.852)	6.495***	(1.322)	4.013***	(0.652)
MWTP_days_in_hospital	-0.165***	(0.0449)	-0.220***	(0.0597)	-0.0871	(0.0495)	-0.0320	(0.0320)	-0.0775***	(0.0228)
Observations	143580		143580		102661		102661		104272	
Personal income per year	0.00144***	(0.000381)	0.000438	(0.000344)	-0.000113	(0.000485)	0.0000283	(0.000306)	0.00106***	(0.000261)
Seeing friends and family	0.0938***	(0.00618)	0.0317***	(0.00430)	0.0244***	(0.00541)	0.0140***	(0.00327)	0.00790***	(0.00228)
Talking to neighbours	0.0986***	(0.00527)	0.0349***	(0.00385)	0.0233***	(0.00529)	0.0134***	(0.00309)	0.00818***	(0.00228)
Health status	0.434***	(0.00588)	0.207***	(0.00403)	0.151***	(0.00555)	0.0882***	(0.00318)	0.0633***	(0.00233)
Days in hospital	-0.00388***	(0.000816)	-0.00346***	(0.000484)	-0.00156	(0.000820)	-0.000434	(0.000426)	-0.00121***	(0.000303)
MWTP_freq_friends	65.39***	(17.73)	72.29	(57.37)	-216.3	(929.8)	493.8	(5343.9)	7.431**	(2.804)
MWTP_freq_neighbours	68.71***	(18.48)	79.62	(62.96)	-206.3	(886.5)	473.0	(5121.3)	7.696**	(2.876)
MWTP_healthstatus	302.4***	(80.58)	473.5	(371.4)	-1337.3	(5738.5)	3118.7	(33752.0)	59.53***	(14.76)
MWTP_days_in_hospital	-2.706**	(0.918)	-7.900	(6.291)	13.81	(59.71)	-15.33	(166.6)	-1.141**	(0.399)
Observations	143580		143580		102661		102661		104272	
Real personal income per year	0.00157***	(0.000428)	0.000401	(0.000386)	-0.000134	(0.000540)	0.0000256	(0.000341)	0.00113***	(0.000292)
Seeing friends and family	0.0938***	(0.00618)	0.0317***	(0.00430)	0.0244***	(0.00541)	0.0140***	(0.00327)	0.00789***	(0.00228)
Talking to neighbours	0.0986***	(0.00527)	0.0349***	(0.00385)	0.0233***	(0.00529)	0.0134***	(0.00309)	0.00818***	(0.00228)
Health status	0.434***	(0.00588)	0.207***	(0.00403)	0.151***	(0.00555)	0.0882***	(0.00318)	0.0632***	(0.00233)
Days in hospital	-0.00388***	(0.000816)	-0.00346***	(0.000484)	-0.00156	(0.000820)	-0.000434	(0.000426)	-0.00121***	(0.000303)
MWTP_freq_friends	59.81***	(16.62)	78.87	(76.52)	-182.7	(738.6)	545.9	(7279.9)	6.960**	(2.682)
MWTP_freq_neighbours	62.85***	(17.33)	86.90	(84.13)	-174.3	(704.1)	522.8	(6976.0)	7.216**	(2.754)
MWTP_healthstatus	276.6***	(75.65)	516.9	(497.9)	-1130.0	(4557.1)	3447.6	(45981.9)	55.79***	(14.50)
MWTP_days_in_hospital	-2.475**	(0.853)	-8.624	(8.389)	11.67	(47.47)	-16.95	(226.7)	-1.070**	(0.383)
Observations	143580		143580		102661		102661		104272	

Income is measured in 1000 GBP. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.8: Valuation results for different logartihmic income variables (BHPS 1997-2009)

	(1) LM OLS		(2) LM FE		(3) dLM OLS		(4) sign(dLM) OLS		(5) RM OLS	
Log household income p.y. and p.c.	0.0808***	(0.00750)	0.0245***	(0.00592)	0.0141	(0.00834)	0.0108*	(0.00502)	0.00815*	(0.00380)
Seeing friends and family	0.0942***	(0.00616)	0.0317***	(0.00429)	0.0244***	(0.00541)	0.0139***	(0.00327)	0.00781***	(0.00228)
Talking to neighbours	0.0995***	(0.00526)	0.0350***	(0.00385)	0.0234***	(0.00529)	0.0134***	(0.00309)	0.00822***	(0.00228)
Health status	0.432***	(0.00587)	0.207***	(0.00403)	0.151***	(0.00555)	0.0882***	(0.00318)	0.0632***	(0.00233)
Days in hospital	-0.00400***	(0.000818)	-0.00347***	(0.000484)	-0.00155	(0.000821)	-0.000428	(0.000426)	-0.00121***	(0.000303)
MWTP_freq_friends	1.166***	(0.132)	1.291***	(0.357)	1.731	(1.088)	1.289	(0.671)	0.958	(0.527)
MWTP_freq_neighbours	1.231***	(0.131)	1.427***	(0.377)	1.658	(1.049)	1.242	(0.646)	1.008	(0.547)
MWTP_healthstatus	5.344***	(0.504)	8.458***	(2.049)	10.71	(6.340)	8.151*	(3.795)	7.750*	(3.619)
MWTP_days_in_hospital	-0.0495***	(0.0112)	-0.141***	(0.0394)	-0.110	(0.0872)	-0.0396	(0.0436)	-0.149	(0.0785)
Observations	143580		143580		102661		102661		104272	
Log real household income p.y. and p.c.	0.0808***	(0.00750)	0.0245***	(0.00592)	0.0141	(0.00834)	0.0108*	(0.00502)	0.00815*	(0.00380)
Seeing friends and family	0.0942***	(0.00616)	0.0317***	(0.00429)	0.0244***	(0.00541)	0.0139***	(0.00327)	0.00781***	(0.00228)
Talking to neighbours	0.0995***	(0.00526)	0.0350***	(0.00385)	0.0234***	(0.00529)	0.0134***	(0.00309)	0.00822***	(0.00228)
Health status	0.432***	(0.00587)	0.207***	(0.00403)	0.151***	(0.00555)	0.0882***	(0.00318)	0.0632***	(0.00233)
Days in hospital	-0.00400***	(0.000818)	-0.00347***	(0.000484)	-0.00155	(0.000821)	-0.000428	(0.000426)	-0.00121***	(0.000303)
MWTP_freq_friends	1.166***	(0.132)	1.291***	(0.357)	1.731	(1.088)	1.289	(0.671)	0.958	(0.527)
MWTP_freq_neighbours	1.231***	(0.131)	1.427***	(0.377)	1.658	(1.049)	1.242	(0.646)	1.008	(0.547)
MWTP_healthstatus	5.344***	(0.504)	8.458***	(2.049)	10.71	(6.340)	8.151*	(3.795)	7.750*	(3.619)
MWTP_days_in_hospital	-0.0495***	(0.0112)	-0.141***	(0.0394)	-0.110	(0.0872)	-0.0396	(0.0436)	-0.149	(0.0785)
Observations	143580		143580		102661		102661		104272	
Log real personal income p.y.	-0.00768	(0.00460)	-0.00654	(0.00374)	-0.00142	(0.00516)	-0.00186	(0.00322)	0.0202***	(0.00266)
Seeing friends and family	0.0924***	(0.00625)	0.0310***	(0.00434)	0.0239***	(0.00549)	0.0134***	(0.00333)	0.00781***	(0.00232)
Talking to neighbours	0.0987***	(0.00540)	0.0349***	(0.00392)	0.0251***	(0.00544)	0.0146***	(0.00319)	0.00940***	(0.00233)
Health status	0.434***	(0.00596)	0.206***	(0.00409)	0.148***	(0.00564)	0.0864***	(0.00324)	0.0645***	(0.00237)
Days in hospital	-0.00359***	(0.000818)	-0.00330***	(0.000485)	-0.00138	(0.000828)	-0.000381	(0.000431)	-0.00122***	(0.000306)
MWTP_freq_friends	-12.03	(7.294)	-4.747	(2.803)	-16.84	(61.28)	-7.182	(12.55)	0.387**	(0.126)
MWTP_freq_neighbours	-12.85	(7.759)	-5.331	(3.120)	-17.63	(64.07)	-7.810	(13.57)	0.466***	(0.131)
MWTP_healthstatus	-56.58	(33.93)	-31.48	(18.03)	-103.9	(377.1)	-46.33	(80.01)	3.196***	(0.437)
MWTP_days_in_hospital	0.467	(0.300)	0.504	(0.298)	0.969	(3.569)	0.205	(0.423)	-0.0606***	(0.0172)
Observations	138796		138796		98443		98443		100008	
Log personal income per year	-0.00768	(0.00460)	-0.00654	(0.00374)	-0.00142	(0.00516)	-0.00186	(0.00322)	0.0202***	(0.00266)
Seeing friends and family	0.0924***	(0.00625)	0.0310***	(0.00434)	0.0239***	(0.00549)	0.0134***	(0.00333)	0.00781***	(0.00232)
Talking to neighbours	0.0987***	(0.00540)	0.0349***	(0.00392)	0.0251***	(0.00544)	0.0146***	(0.00319)	0.00940***	(0.00233)
Health status	0.434***	(0.00596)	0.206***	(0.00409)	0.148***	(0.00564)	0.0864***	(0.00324)	0.0645***	(0.00237)
Days in hospital	-0.00359***	(0.000818)	-0.00330***	(0.000485)	-0.00138	(0.000828)	-0.000381	(0.000431)	-0.00122***	(0.000306)
MWTP_freq_friends	-12.03	(7.294)	-4.747	(2.803)	-16.84	(61.28)	-7.182	(12.55)	0.387**	(0.126)
MWTP_freq_neighbours	-12.85	(7.759)	-5.331	(3.120)	-17.63	(64.07)	-7.810	(13.57)	0.466***	(0.131)
MWTP_healthstatus	-56.58	(33.93)	-31.48	(18.03)	-103.9	(377.1)	-46.33	(80.01)	3.196***	(0.437)
MWTP_days_in_hospital	0.467	(0.300)	0.504	(0.298)	0.969	(3.569)	0.205	(0.423)	-0.0606***	(0.0172)
Observations	138796		138796		98443		98443		100008	

Income is measured in 1000 GBP. For log-income variables, the MWTP needs to be multiplied with the income level of interest. Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

finding is that real and nominal income leads to very similar results.

Result 3: *Significance of MWTP hinges on significant income coefficients. All other coefficients are robust to using differnt income variables. Monthly income is more often significant than yearly income (for both personal and household income). There is no noticeable differences between nominal and real income variables. Log income has a significiant coefficient only with RM in regression (5).*

5.4 Discussion and conclusion

In this paper, I have considered two problems of using the LM for the LSA: person fixed effects and adaptation of the reporting function. While theoretically, the RM can remedy both of them, it also drastically reduces the variation used. The RM is thus better suited to measure individual preference rankings over situations and does not suffer from differing or changing reporting functions. However, only using individual preference rankings over pairs of situations reduces the variation used and thus makes it more difficult to infer the MWTP. When choosing the life satisfaction measure, we thus face the trade-off between a more accurate measure paired with a reduction in variation used (RM) and a less accurate measure which uses more variation and therefore can produce potentially biased estimates (LM). The additional variation in the LM is thus not necessarily helpful if the LM does not reliably capture changes in the latent variable but rather random or systematic changes of the reporting function. Empirically, the results show that the RM allows to infer much more realistic MWTP estimates but it is less often significant.

Future research will have to show if the conceptual potential of the RM can be better exploited when valuing other, more suitable non-market goods that cannot be influenced by the individual or which are more objectively measured, e.g. by using location specific data to value local amenities (as done e.g. for air pollution by Levinson (2012)). One should also use more suitable RM questions which ask for the evaluation of changes and not for the change in valuation. Possibly also the control variables would need to include rents in addition to income in order to control for possible compensation mechanisms and to identify the whole effect of the non-market good and not only the residual one that has not been compensated for through markets (see. e.g. Welsch and Ferreira 2014).

Future studies applying the LSA will also have to be more clear about the definition of the well-being concept they use as a measuring rod, since as I have discussed, different well-being concepts will lead to different MWTP values depending on whether adaptation is interpreted as adaptation in well-being or as adaptation in the reporting function.

The LSA with both the LM and the RM will need further investigation in the future in order to further establish the LSA as a valuable alternative or complement to stated and revealed preference methods to valuation of non-market goods.

5.5 References

Ambrey, C. L., and Fleming, C. M. (2011). Valuing scenic amenity using life satisfaction data. *Ecological Economics* , 72, 106-115.

Atkinson, G., and Mourato, S. (2006). *Cost-benefit analysis and the environment: recent developments.* OECD, Paris.

Berridge, K. C., and Kringelbach, M. L. (2011). Building a neuroscience of pleasure and well-being. *Psychology of Well-Being: Theory, Research and Practice*, 1(1), 3.

Bertram, C. and Rehdanz, K., (2015). The role of urban green space for human well-being, *Ecological Economics*, 120, issue C, p. 139-152.

Bond, T. N., and Lang, K. (2014). The sad truth about happiness scales (No. w19950). *National Bureau of Economic Research.*

Brereton, F., Clinch, J. P., and Ferreira, S. (2008). Happiness, geography and the environment. *Ecological Economics*, 65(2), 386-396.

Camerer, C., Loewenstein, G., and Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, 43(1), 9-64.

Cameron, A.C. and Trivedi, P.K. (1998) *Regression Analysis of Count Data.* New York: Cambridge University Press.

Carroll, N., Frijters, P., and Shields, M. A. (2009). Quantifying the costs of drought: new evidence from life satisfaction data. *Journal of Population Economics*, 22(2), 445-461.

Carson, R. T., Mitchell, R. C., Hanemann, M., Kopp, R. J., Presser, S., and Ruud, P. A. (2003). Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill. *Environmental and resource economics*, 25(3), 257-286.

Daly, M. C., and Wilson, D. J. (2009). Happiness, unhappiness, and suicide: An empirical assessment. *Journal of the European Economic Association*, 7(2-3), 539-549.

Daly, M. C., Wilson, D. J., and Johnson, N. J. (2013). Relative status and well-being: evidence from US suicide deaths. *Review of Economics and Statistics*, 95(5), 1480-1500.

Diener, E., Inglehart, R., and Tay, L. (2013). Theory and validity of life satisfaction scales. *Social Indicators Research*, 112(3), 497-527.

Diener, E., Suh, E.M., Lucas, R.E. and Smith, H.L. (1999) Subjective well-being: three decades of progress. *Psychological Bulletin* 125: 276–302.

Fehr, E., and Rangel, A. (2011). Neuroeconomic foundations of economic choice—recent advances. *Journal of Economic Perspectives*, 25(4), 3-30.

Ferreira, S., and Moro, M. (2010). On the use of subjective well-being data for environmental valuation. *Environmental and Resource Economics*, 46(3), 249-273.

Ferrer-i-Carbonell, A., and Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness?. *The Economic Journal*, 114(497), 641-659.

Finkelstein, A., Luttmer, E. F., and Notowidigdo, M. J. (2013). What good is wealth without health? The effect of health on the marginal utility of consumption. *Journal of the European Economic Association*, 11(suppl_1), 221-258.

Fleming, C., and Ambrey, C. (2017). The Life Satisfaction Approach to Environmental Valuation. *Oxford Research Encyclopedia of Environmental Science*. Retrieved 18 Apr. 2018, from http://environmentalscience.oxfordre. com/view/10.1093/ acrefore/ 9780199389414.001.0001/ acrefore-9780199389414-e-530.

Freeman III, A.M. (2003), *The Measurement of Environmental Resource Values: Theory and Methods*, Resources for the Future, Washington.

Frey, B. S., Luechinger, S., and Stutzer, A. (2009). The life satisfaction approach to valuing public goods: The case of terrorism. *Public Choice*, 138(3), 317-345.

Frey, B. S., Luechinger, S., and Stutzer, A. (2010). The life satisfaction approach to environmental valuation. *Annual Review of Resource Economics* 2010 2:1, 139-160.

Köke, S., and Perino, G. (2017). For "better" or "worse": a direct approach to elicit preference rankings from life-satisfaction data. *WiSo-HH Working Paper Series*, No. 43.

Krekel, C., Kolbe, J., and Wüstemann, H. (2016). The greener, the happier? The effect of urban land use on residential well-being. *Ecological economics*, 121, 117-127.

Levinson, A. (2012). Valuing public goods using happiness data: The case of air quality. *Journal of Public Economics*, 96(9), 869-880.

Luechinger, S. (2009). Valuing air quality using the life satisfaction approach. *The Economic Journal*, 119(536), 482-515.

Luttmer, Erzo F. P. (2005). Neighbors as Negatives: Relative Earnings and Well-Being. *Quarterly Journal of Economics*, 120(3), 963-1002.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic theory* (Vol. 1). New York: Oxford university press.

Palmquist, R. B., and Smith, V. K. (2002). The use of hedonic property value techniques for policy and litigation. *The international yearbook of environmental and resource economics*, 115-64.

Pigou, A.C. (1932). *The Economics of Welfare.* 4th ed., London.

Powdthavee, N. (2008). Putting a price tag on friends, relatives, and neighbours: Using surveys of life satisfaction to value social relationships. *The Journal of Socio-Economics*, 37(4), 1459-1480.

Rayo, L., and Becker, G. S. (2007a). Evolutionary efficiency and happiness. *Journal of Political Economy*, 115(2), 302-337.

Rayo, L., and Becker, G. S. (2007b). Habits, peers, and happiness: An evolutionary perspective. *American Economic Review*, 97(2), 487-491.

Rehdanz, K., and Maddison, D. (2005). Climate and happiness. *Ecological Economics*, 52(1), 111-125.

Rehdanz, K., and Maddison, D. (2008). Local environmental quality and life-satisfaction in Germany. *Ecological economics*, 64(4), 787-797.

Smith, V. K. (2008). Reflections on the Literature. *Review of Environmental Economics and Policy*, 2, 292–308.

University of Essex. Institute for Social and Economic Research (2010). British Household Panel Survey: Waves 1-18, 1991-2009. [data collection]. 7th Edition. UK Data Service. SN: 5151, http://doi.org/10.5255/UKDA-SN-5151-1

Welsch, H., and Ferreira, S. (2014). Environment, well-being, and experienced preference. *International Review of Environmental and Resource Economics*, 7(3), 205–239.

Welsch, H., and Kühling, J. (2009). Using happiness data for environmental valuation: issues and applications. *Journal of Economic Surveys*, 23(2), 385-406.

Appendix A

Appendix Chapter 2

A.1 Proofs

Proof of Lemma 1: Equation (2.4) immediately shows that the probability π is weakly decreasing in β^R as a larger β^R reduces the probability of an individual country ratifying the agreement. Also, β^R is increasing in $\hat{q}$:

$$\frac{\partial \beta^R}{\partial \hat{q}} = \frac{\partial \beta^R(\hat{q}, \hat{r})}{\partial \hat{q}} = \frac{C'(\hat{q})(\hat{q} - q^N) - (C(\hat{q}) - C(q^N))}{\hat{r}(\hat{q} - q^N)^2} > 0$$

where we used the convexity of the cost function $C(\cdot)$. Similarly β^R is decreasing in $\hat{r}$:

$$\frac{\partial \beta^R}{\partial \hat{r}} = \frac{\partial \beta^R(\hat{q}, \hat{r})}{\partial \hat{r}} = -\frac{\beta^R}{\hat{r}} < 0$$

which already proves the claim regarding the impact of $\hat{q}$ (weakly negative) and – by continuity – regarding the impact of $\hat{r}$ (weakly positive) on π for small variance of the country-specific shocks ϵ_i (see (2.4)).

Increases in $\hat{r}$ have, however, also a directly declining effect on π in case country-specific shocks ϵ_i are present (even when β^R is kept fixed). This is because the more countries are required to ratify before the agreement enters into force the higher the probability that one of the required countries does not succeed in ratifying (see (2.4)).

The potentially decreasing effect of $\hat{r}$ on π can be easily seen by comparing the extreme cases of $\hat{r} = 1$ and $\hat{r} = n$ for a case in which no global shock exists ($\sigma_\theta^2 = 0$):

$$\begin{aligned} \pi(\hat{q}, 1) &= 1 - G(\beta_1^R - \lambda b^R)^n \\ \pi(\hat{q}, n) &= (1 - G(\beta_n^R - \lambda b^R))^n \end{aligned}$$

where $\beta_r^R = \beta^R(\hat{q}, 1)/r$. Here $\pi(\hat{q}, 1) > \pi(\hat{q}, n)$ would be obtained if $1 > G(\beta_1^R - \lambda b^R)^n + (1 - G(\beta_n^R - \lambda b^R))^n$ for which examples can easily be created. □

Proof of Proposition 1:

(i)

We already have shown in the text that $\hat{r} = N$ under certainty. We now note that under certainty pivotal agents will accept the commitment level $\hat{q} = q^{*R}$ which is optimal for the representatives ($C'(q^{*R}) = nb^R$) if $b^P = \lambda b^R \geq \beta(q^{*R}, n)$. Otherwise, the largest $\hat{q}$ that will be accepted is chosen, i.e. $b^P = \lambda b^R = \beta(\hat{q}, n)$, thereby implying $\hat{q} < q^{*R}$.

(ii)

For positive levels of the variance ($\sigma_\theta^2 > 0$ or $\sigma_\epsilon^2 > 0$), (2.7) immediately implies that the optimal level of $\hat{q}$ lies below q^{*R} as $\partial\pi/\partial\hat{q} < 0$. For $\sigma_\epsilon^2 = 0$, (2.6) implies that $\hat{r} = n$ as the probability is increasing in $\hat{r}$. To show that $\hat{r} < n$ can result if for $\sigma_\epsilon^2 >$, it suffices to prove one example: consider $\sigma_\theta^2 = 0$ and a symmetric distribution with $\sigma_\epsilon^2 > 0$. For $\sigma_\epsilon^2 \to \infty$, we have $G(\beta - \lambda b^R) \to 0.5$ as $G(\beta - \theta) = G(-(\beta - \theta)) = 0.5(1 - \int_{-(\beta-\theta)}^{\beta-\theta} G'(x)dx) \to= 0.5$ since G' uniformly converges to zero when $\sigma_\epsilon^2 \to \infty$.

Thus, (2.4) with (2.6) immediately implies that the optimal minimum participation level converges to $\bar{r} = \arg\max_{\hat{r}} \hat{r} \sum_{r=\hat{r}}^{n} \binom{n}{r}$. This leads to $\hat{r} < n$ for $n > 2$ as evaluated at $\hat{r} = n$ it evaluates to n while the value for $\hat{r} = n - 1$ is already larger: $(n-1)(1+n)$. In fact, the relationship between the limit of the participation threshold $\bar{r}$ and n is surprisingly linear: the best linear fit is given by $\bar{r} = -1.08777 + 0.440294n$ and describes the data well for $n > 20$ (derivation using Wolfram Mathematica 8). That is, the participation threshold in a completely uncertain world would require a minimum of about 44% of negotiating countries to ratify the agreement before it enters into force.

(iii)

Under certainty, $\hat{q} = q^{*R}$ if $\beta^R(q^{*R}, n) \leq \lambda b^R$, while otherwise $\hat{q}$ must be chosen at level

$\beta^R(\hat{q},n) = \lambda b^R$, to ensure ratification. For $\lambda < 1/n$, however, representatives would unilaterally abate more: $\hat{q} < q^N$ $(\lim_{\hat{q}\downarrow q^N} \beta^R(\hat{q},n) = C'(q^N)/n = b^R/n > \lambda b^R)$.

For small variance, continuity of the objective function implies that $\hat{r} = n$ just as under certainty. As such, small variance for $\lambda \leq 1/n$, leads to an increase in the implemented $\hat{q}$. Since $\hat{q} = q^{*R}$ if $\beta^R(q^{*R},n) \leq \lambda b^R$ the claim in this range follows immediately from (ii). For $1/n < \lambda < \beta^R(q^{*R},n)/b^R$, we have that at the commitment level chosen under certainty ($\lambda b^R < \beta^R(\hat{q},n)$) the marginal impact of $\hat{q}$ on probability $\partial\pi(\hat{q},n)/\partial\hat{q}) \to -\infty$ when the variance approaches zero as the distributions are assumed to be continuous for any positive variance. As such, it follows that for small variance the commitment level is chosen *below* the one that results under certainty.

The last statement follows immediately from (2.7) and $\partial\pi/\partial\hat{q} \to 0$ if $\sigma_\theta^2 \to \infty$ or $\sigma_\epsilon^2 \to \infty$ (since $\partial\pi/\partial\beta \to 0$).

□

Proof of Proposition 2

Noting $\hat{\pi} = 1$ under certainty, condition (2.8) is thus always fulfilled if $\lambda \leq 1/(n-1)$ as then no commitment above the unilateral level would be accepted in a coalition of $n-1$ players, i.e. the right hand side of (2.8) is zero. For $\lambda > 1/n$, (2.8) is equivalent to

$$
\begin{aligned}
& 1 - \frac{C(\hat{q}(n)) - C(q^N)}{nb^R(\hat{q}(n) - q^N)} \geq \frac{(n-1)b^R(\hat{q}(n-1) - q^N)}{nb^R(\hat{q}(n) - q^N)} \\
\Leftrightarrow \quad & 1 - \frac{c(\hat{q}(n) + q^N)}{2nb^R} \geq \frac{(n-1)(\hat{q}(n-1) - q^N)}{n(\hat{q}(n) - q^N)} \\
\Leftrightarrow \quad & \begin{cases} 1 - \frac{n+1}{2n} \geq \frac{(n-1)(n-2)}{n(n-1)} & \text{if} \quad \lambda \geq \frac{n}{2(n-1)} \\ 1 - \frac{n+1}{2n} \geq \frac{(n-1)2((n-1)\lambda-1)}{n(n-1)} & \text{if} \quad \frac{n+1}{2n} \leq \lambda < \frac{n}{2(n-1)} \\ 1 - \lambda \geq \frac{(n-1)((n-1)\lambda-1)}{n(n\lambda-1)} & \text{if} \quad \frac{1}{n} < \lambda < \frac{n+1}{2n} \end{cases}
\end{aligned}
\tag{A.1}
$$

In a first step this simplifies to

$$
n \leq \begin{cases} 3 & \text{if} \quad \lambda \geq \frac{n}{2(n-1)} \\ \frac{4\lambda+3}{4\lambda-1} & \text{if} \quad \frac{n+1}{2n} \leq \lambda < \frac{n}{2(n-1)} \\ \frac{3}{2\lambda} + \frac{1}{\lambda}\sqrt{\frac{5}{4} - \lambda} & \text{if} \quad \frac{1}{n} < \lambda < \frac{n+1}{2n} \end{cases} \tag{A.2}
$$

which only endogenously determines the threshold levels for n. We therefore need to reinsert the solutions for the maximal n given in (A.2) to check for which λ the respective conditions on the right hand side hold. For example: $n = 3$ satisfies $\lambda \geq n/(2(n-1))$ only if $\lambda \geq 3/4$. $n = \frac{3}{2\lambda} + \frac{1}{\lambda}\sqrt{\frac{5}{4} - \lambda}$ only satisfies $\lambda < \frac{n+1}{2n}$ if $\lambda < \frac{1+\sqrt{17}}{8}$. Using these thresholds directly leads to (2.12). The corresponding abatement levels $q^* = \hat{q}(n^*)$ are thus obtained using (2.18) and lead to (2.13).

It is obvious that both n^* as well as q^* are (weakly) decreasing in λ.

$\square$

Proof of Proposition 3

See main text.

Proof of Proposition 4

The proof is given by example of uniform distribution in the main text. Below, we report simulation results for normal distributions ($\sigma_\epsilon^2 > 0$) as further evidence for the robustness of the result. We specify the quadratic abatement cost $C(q) = \frac{c}{2}q^2$ with $c = 50$ and assume a benefit of the representative at $b^R = 50$. We assume a normal distribution preference parameter b^P due to country-specific shocks only ($\sigma_\theta^2 = 0$). We simulate the optimal choices and maximal stable coalition size for $\sigma_\epsilon^2 \in \{0, 10, 20, 100\}$ and use specific values $\lambda = E(b^P)/b^R \in \{1.1, 1, 0.25\}$. For a positive variance under a normal distribution, we know from the discussion in Section 2.2.4 that $\pi < 1$ and $\hat{q} < q^{*R}$. In Table A.1, we report n^*, the resulting content of the agreement $\hat{q}^*$ and $\hat{r}^*$, as well as the ratification probability $\hat{\pi}^*$, the expected abatement levels above the non-cooperative solution, $\hat{\pi}^*\hat{r}^*\hat{q}^*$, the expected number of ratifying countries ($\hat{\pi}^*\hat{r}^*$) as well as the expected payoff gains to representatives above the fully non-cooperative solution ($q_i = q^N$ for all i) in percent for $N = 100$ (for the rest of the analysis N does not matter). The latter is given by

$$\Delta E[U^R] = \hat{\pi}^*\hat{r}^* \left[b^R N(\hat{q}^* - q^N) - (C(\hat{q}^*) - C(q^N)) \right]$$

and are reported relative to the non-cooperative payoff level $U^{RN} = N\left[b^R N q^N - C(q^N)\right]$.

Table A.1 shows that an increasing variance σ_ϵ^2 in the ratifiers' preference parameter b^P initially can increase the number of countries who join the negotiations, before n^*

λ	$b^R/E(b^P)$	σ_ϵ	n^*	$\hat{\pi}^*$	$\hat{q}^*$	$\hat{r}^*$	$\hat{r}^*\hat{\pi}^*$	$\hat{q}^*\hat{r}^*\hat{\pi}^*$	$\Delta E[U^R]/U^{RN}$ in %
1.1	0.91	0	3	1	3	3	3	9	5.9
1.1	0.91	10	4	0.98	3.51	3	2.95	10.35	7.27
1.1	0.91	20	5	0.80	3.53	4	3.21	11.32	7.97
1.1	0.91	100	4	0.51	3.33	2	1.01	3.37	2.32
1	1	0	3	1	3	3	3	9	5.9
1	1	10	4	0.95	3.10	3	2.86	8.86	5.9
1	1	20	5	0.71	3.25	4	2.84	9.23	6.3
1	1	100	4	0.48	3.29	2	0.96	3.16	2.1
0.25	4	0	10	1	4	10	10	40	29.4
0.25	4	10	26	0.76	4.30	15	11.4	49.02	36.8
0.25	4	20	22	0.62	4.23	10	6.2	26.23	19.8
0.25	4	100	4	0.27	2.97	2	0.54	1.06	1.1

Table A.1: *Stable coalition sizes n^*, optimal decisions of $\hat{q}^*$ and $\hat{r}^*$, and expected utility gains $\Delta E[U^R]/U^{RN}$ for representatives ($\lambda = E(b^P)/b^R \in \{1.1, 1, 0.25\}$ ($\sigma_\theta^2 = 0$), normal distribution of country-specific shocks $\sigma_\epsilon^2 \in \{0, 10, 20, 100\}$, $b^R = 50$, $c = 50$).*

the number declines again. A similar non-monotonic effect is obtained for the minimum participation level, and also for the effective commitment level $\hat{q}^*$. We also see that uncertainty may also increase the expected size of the coalition, the expected abatement level, as well as expected payoff gains $\Delta E[U^R]$. The corresponding probability of entry into force $\pi(n^*)$ is continuously decreasing in σ_ϵ^2.

For large levels of uncertainty, however, negotiators return to proposing full internalization as has been shown in the previous section and, as a consequence, a small number of countries joins the negotiations. Here, uncertainty worsens the chances of meaningful cooperation and of achieving payoff gains relative to the non-cooperative solution as it is less likely that countries ratify.

□

Detrimental effect of large variance on prospects of cooperation

We first rewrite condition (2.8) as

$$\begin{aligned} & \frac{\hat{\pi}(n)\hat{r}(n)}{n}\left[nb^R(\hat{q}(n) - q^N) - (C(\hat{q}(n)) - C(q^N))\right] \\ \geq \; & \frac{\hat{\pi}(n-1)\hat{r}(n-1)}{n-1}\left[(n-1)b^R(\hat{q}(n-1) - q^N)\right] \end{aligned} \tag{A.3}$$

We further note that Proposition 1 (iii) already implies $q^*(n) = nb^R/c$ if $\sigma_\epsilon^2 + \sigma_\theta^2 \to \infty$.

Thus the terms in the brackets converge to $((b^R)^2/(c))[n(n-1)-(n^2-1)/2]$ on the left hand side and $((b^R)^2/(c))[(n-1)(n-2)]$ on the right hand side. Without the probability weights, this leads to the well-known result that the maximal size of a stable coalition is 3.

Noting that for $\sigma_\epsilon^2 = 0$, we have $\hat{r} = n$ and $\pi(n) \to 0.5$ for symmetric distributions as seen in the proof of Proposition 1. As such (A.3) in the limit coincides with the condition under certainty besides both sides being weighted with 0.5.

For $\sigma_\epsilon^2 \to \infty$ ($\sigma_\theta^2 = 0$), the proof of Proposition 1 shows that $\hat{r}(n) \to \arg\max_{\hat{r}} \hat{r} \sum_{r=\hat{r}}^{n} \binom{n}{r}$. This can be used to easily show that $\hat{r}(2) = 1$, $\hat{r}(3) = \hat{r}(4) = 2$. Straightforward calculations imply that $n^* = 2$ and $\hat{r}^* = 1$, with the probability of the agreement entering into force being given by 0.75.

Proof of Proposition 5

Using (2.18), we can rewrite the stability condition (2.19):

$$\begin{aligned}
& nb^R(\hat{q}(n) - q^N) - (C(\hat{q}(n)) - C(q^N)) = (n-1)b^R(\hat{q}(n-1) - q^N) \\
\Leftrightarrow \quad & 1 - \frac{C(\hat{q}(n)) - C(q^N)}{nb^R(\hat{q}(n) - q^N)} = \frac{(n-1)b^R(\hat{q}(n-1) - q^N)}{nb^R(\hat{q}(n) - q^N)} \\
\Leftrightarrow \quad & 1 - \frac{c(\hat{q}(n) + q^N)}{2nb^R} \geq \frac{(n-1)(\hat{q}(n-1) - q^N)}{n(\hat{q}(n) - q^N)} \\
\Leftrightarrow \quad & \begin{cases} 1 - \frac{n+\lambda}{2n} \geq \frac{(n-1)(n-1-\lambda)}{n(n-\lambda)} & \text{if} \quad \lambda \geq \frac{n-1}{2(n-1)-1} \\ 1 - \frac{n+\lambda}{2n} \geq \frac{(n-1)2(n-2)\lambda}{n(n-\lambda)} & \text{if} \quad \frac{n}{2n-1} \leq \lambda < \frac{n-1}{2(n-1)-1)} \\ 1 - \lambda \geq \frac{(n-1)2(n-2)\lambda}{n2(n-1)\lambda} & \text{if} \quad \frac{1}{n} < \lambda < \frac{n}{2n-1} \end{cases}
\end{aligned} \tag{A.4}$$

In a first step this simplifies to

$$n \leq \begin{cases} 2 + \sqrt{\lambda^2 - 2\lambda + 2} & \text{if} \quad \lambda \geq \frac{n-1}{2(n-1)-1} \\ \frac{5\lambda + 2\sqrt{\lambda^3 - 2\lambda^2 + 2\lambda}}{4\lambda - 1} & \text{if} \quad \frac{n}{2n-1} \leq \lambda < \frac{n-1}{2(n-1)-1)} \\ \frac{2}{\lambda} & \text{if} \quad \frac{1}{n} < \lambda < \frac{n}{2n-1} \end{cases} \tag{A.5}$$

which only endogenously determines the threshold levels for n. We therefore need to reinsert the solutions for the maximal n given in (A.5) to check for which λ the respective conditions on the right hand side hold. For example: $n = 2/\lambda$ satisfies $\lambda < n/(2n-1)$

only if $\lambda < 1/(2-\lambda/2)$, i.e. if $\lambda < 2-\sqrt{2}$. The other threshold is obtained numerically. Using these thresholds directly leads to (2.20). The corresponding abatement levels $q^* = \hat{q}(n^*)$ are thus obtained using (2.18) and lead to (2.21).

Considering (2.20), n^* is obviously decreasing in $\lambda \leq 2-\sqrt{2}$. For $\leq 2-\sqrt{2} \leq \lambda < 0.66067$, we need to consider

$$\begin{aligned}
& \frac{\partial}{\partial\lambda}\frac{5\lambda+?\sqrt{\lambda^3-2\lambda^2+2\lambda}}{4\lambda-1} \\
= \; & \frac{[5+\frac{3\lambda^2-4\lambda+2}{\sqrt{\lambda^3-2\lambda^2+2\lambda}}](4\lambda-1)-4[5\lambda+2\sqrt{\lambda^3-2\lambda^2+2\lambda}]}{(4\lambda-1)^2} \\
= \; & \frac{-5+\frac{3\lambda^2-4\lambda+2}{\sqrt{\lambda^3-2\lambda^2+2\lambda}}(4\lambda-1)-8\sqrt{\lambda^3-2\lambda^2+2\lambda}}{(4\lambda-1)^2} \\
= \; & \frac{-5+\frac{(4\lambda-1)(3\lambda^2-4\lambda+2)-8\lambda^3+16\lambda^2-16\lambda}{\sqrt{\lambda^3-2\lambda^2+2\lambda}}}{(4\lambda-1)^2} \\
= \; & \frac{-5+\frac{4\lambda^3-3\lambda^2-4\lambda-2}{\sqrt{\lambda^3-2\lambda^2+2\lambda}}}{(4\lambda-1)^2} \\
< \; & 0
\end{aligned}$$

where the last inequality follows from $\lambda < 1$.

For $\lambda > 0.66067$, it is given by $2+\sqrt{\lambda^2-2\lambda+2}$. Here, $\lambda^2-2\lambda+2$ takes its minimum in $\lambda = 1$ such that the minimal n^* is given at this level by 3. Identical arguments show the claimed relationship between q^* and λ.

□

Appendix B

Appendix Chapter 3

B.1 Derivation of conditions for cooperative equilibria

Derivation of condition (3.2):

It is obvious that the game has a subgame perfect equilibrium in which *all* players *always* defect: as in the one-shot prisoner's dilemma game, no player individually has an incentive to cooperate. To show the minimal requirement for sustained cooperation to exist, we rely on modified grim trigger strategies: we assume that a set of $Q \leq n$ players follow a modified grim trigger strategy: they cooperate as long as at least $Q-1$ other players cooperate, otherwise they defect in all subsequent periods. The remaining $n-Q$ players always defect. This strategy can sustain cooperation as a subgame perfect equilibrium if

$$
\begin{aligned}
& \sum_{t=0}^{\infty}(1-\delta)^t \left[E - c - p_0(D_0 - dQ)\right] \\
& \geq \left[E - p_0(D_0 - d(Q-1))\right] + \sum_{t=1}^{\infty}(1-\delta)^t (E - p_0 D_0) \\
\Leftrightarrow \quad & \frac{1}{\delta}\left[E - c - p_0(D_0 - dQ)\right] \\
& \geq \left[E - p_0(D_0 - d(Q-1))\right] + \frac{1-\delta}{\delta}(E - p_0 D_0) \qquad \text{(B.1)}
\end{aligned}
$$

Here, the left-hand side states the expected payoff of any cooperating player i if all Q players continue to cooperate forever. The first expression of the right hand side states the payoff of a deviator i in the period in which he deviates, while the second term states the expected continuation payoff if all players play defect, starting in the next period. Note that given the defection of other players, the deviating player does not have an incentive to return to cooperation. Therefore, if condition (B.1) is satisfied, Q players playing the modified grim trigger strategy and $n-Q$ players always defecting establishes a subgame perfect equilibrium.

Solving condition (B.1) for Q immediately leads to condition (3.2). □

Derivation of conditions that are used in Figure 3.1:

We rewrite condition (B.1) for the different treatments to see when a player with $u_i(\pi) = \pi^{1-\sigma}/(1-\sigma)$ does not have an incentive to deviate from cooperation under the assumption that all other cooperating players play a modified grim trigger strategy. For *CertDam* this is the case if:

$$\begin{aligned} &\frac{1}{\delta}\left[u_i(E-c-p_0(D_0-dQ))\right] \\ &\geq \left[u_i(E-p_0(D_0-d(Q-1)))\right] + \frac{1-\delta}{\delta}u_i(E-p_0D_0) \end{aligned} \tag{B.2}$$

For *DamRed* this is the case if:

$$\begin{aligned} &\frac{1}{\delta}\left[p_0u_i(E-c-(D_0-dQ)) + (1-p_0)u_i(E-c)\right] \\ &\geq \left[p_0u_i(E-(D_0-d(Q-1))) + (1-p_0)u_i(E)\right] \\ &+\frac{1-\delta}{\delta}\left[p_0u_i(E-D_0) + (1-p_0)u_i(E)\right] \end{aligned} \tag{B.3}$$

while for *ProbRed* we obtain:

$$\begin{aligned} &\frac{1}{\delta}\left[(p_0-xQ)u_i(E-c-D_0) + (1-p_0+xQ)u_i(E-c)\right] \\ &\geq \left[(p_0-x(Q-1))u_i(E-D_0)) + (1-p_0+x(Q-1))u_i(E)\right] \\ &+\frac{1-\delta}{\delta}\left[p_0u_i(E-D_0) + (1-p_0)u_i(E)\right] \end{aligned} \tag{B.4}$$

For each of the treatments we can define $Q^{min}(\sigma)$ as the value of Q that satisfies the respective condition (B.2), (B.3) or (B.4) with equality. As an analytical solution proves impossible, Figure 3.1 displays the simulation results.

B.2 Experimental Instructions for the DamRed Treatment (English translation)

In the following, we report an English translations of the experimental instructions for the *DamRed* treatment.

General instructions for the participants

You are now taking part in an economic science experiment. If you carefully read the following instructions, you can - depending on your decisions - earn a not inconsiderable amount of money. Therefore, it is very important that you carefully read the following instructions.

The instructions that we gave you are solely meant for your private information. **During the experiment, communication is completely prohibited.** If you have any questions, please raise your hand out of the cabin. Someone will then come to you to answer your question. Violation of this rule leads to exclusion from the experiment and from all payments.

During the experiment we do not have Euro but Taler. Your total income will first be computed in Taler. The total amount of Taler that you earned during the experiment will be converted into Euro in the end, such that

100 Taler = 1 Euro.

At the end of the experiment you will be paid in cash the total amount of Taler that you earned (converted into Euro) plus 5 Euro for participation. We will conduct the payment such that no other participant will see your payment.

The experiment is divided into two parts. Here, we give the instructions for the 1st part. You will get the instructions for the 2nd part on your computer screen after the 1st part is finished. The two parts are not related with respect to their content.

Explanations for the 1st part of the experiment

The 1st part of the experiment is divided into **phases**. You do not know, however, how many phases there are in total. Each phase is divided into **rounds**. The number of rounds in a phase is random. After each round, the phase ends with a probability of 20%.

More concretely, this means that: after the first round there is a second round with a probability of 80% (which is on average in four cases out of five). So, with a probability of 20% (which is on average in one case out of five) the phase ends after the first round. After the second round (if there is one) there is a third one with a probability of 80%. So, with a probability of 20%, the phase ends after the second round and so on...

At the beginning of each phase, participants are randomly assigned into groups of four. Thus, your group has three other members in addition to you. **During one phase, the constellation of the group remains unchanged. It only gets randomly rematched at the beginning of a new phase.**

Information on the structure of a round

All rounds in all phases are always structured in the exact same way. In the following we describe the structure of one round.

At the beginning of each round, every participant gets an **income** of 25 Taler.

At the end of each round, a **damage** might occur, which reduces the income by 20 Taler.

The damage occurs with a probability of 50% (which is on average in one of two rounds). For this, in each round, the computer randomly determines whether the damage occurs. The occurrence of the damage is only valid in the respective round and does not influence the probability of the next rounds. The occurrence of the damage is determined jointly for the whole group, such that either all or no group members suffer the damage.

All group members are able to reduce the potential damage through their decisions. **For this, at the beginning of each round, i.e. before the damage occurs, each group member has to decide whether it does or does not carry out a damage-reducing action** (see Figure 4 at the end of the instruction).

Each damage-reducing action costs the group member taking the action 5 Taler (independent of whether the damage occurs or not). Each damage-reducing action reduces the personal damage of **each** group member (not only of the group member taking the action) by 4 Taler. For you, personally, this means that each damage-reducing action that has been carried out in your group reduces your damage (if it occurs) by 4 Taler, independent of whether you have taken such an action yourself. **A damage-reducing action which you carry out costs you 5 Taler for sure. In return, you reduce your damage and the damage of each other group member by 4 Taler, if the damage occurs.**

The personal damage, if it occurs, amounts to **20 Taler** if no one in your group carried out an action, **16 Taler** if one person carried out an action, **12 Taler** if two persons took the action, **8 Taler** if three persons took the action and **4 Taler** if all group members took the action.

Your round income (in Taler) is calculated as follows

- If the damage **does not** occur and you **did not** take the damage-reducing action:

  25

- If the damage **does not** occur and you did take the damage-reducing action:

  $25 - 5 = 20$

- If the damage occurs and you **did not** take the damage-reducing action:

  $25 - 20 + 4*$ [sum of all damage-reducing actions in the group],

- If the damage occurs and and you did take the damage-reducing action:

  $25 - 5 - 20 + 4*$ [sum of all damage-reducing actions in the group],

4 examples:

The damage probability always is 50%.

- **You** and **one other** group member take a damage-reducing action in your group, the damage **does not** occur. Your round income is $25 - 5 = \mathbf{20}$ Taler.

- **Only you** take a damage-reducing action in your group, the damage occurs. Your round income is $25 - 5 - 20 + 4 * 1 = \mathbf{4}$ Taler.

- **You** and **two other** group members take a damage-reducing action, the damage occurs. Your round income is $25 - 5 - 20 + 4 * 3 = \mathbf{12}$ Taler.

- **Two other** group members take a damage-reducing action, but you do not, the damage occurs. Your round income is $25 - 20 + 4 * 2 = \mathbf{13}$ Taler.

At the end of a round, each participant receives information on whether he/she took an action him- or herself, how many other group members took an action, if the damage occurred and what the round income is. Then, a new round starts in the same group constellation or in a new group constellation if a new phase begins.

The **sum of all your round incomes** will be paid out to you in private at the end of the experiment.

Before the experiment starts, we would like to ask you to answer some control questions on the computer to make sure you understand the rules.

Figure B.1: Decision screen for taking the action in a round in Part 1 of the experiment.

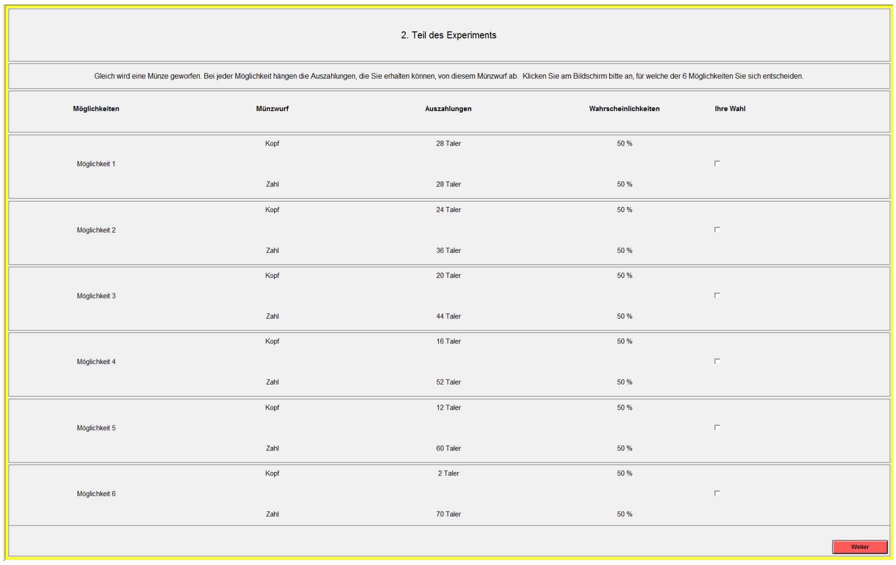

Figure B.2: Decision screen for the risk-assessment task in Part 2 of the experiment.

Appendix C

Appendix Chapter 4

C.1 Description of variables

GESIS and BHPS

Here we describe the explanatory variables used in the regressions. The following variables[1] are available and summarized in Table C.1 and Table C.2:

Table C.1: Summary Statistics of BHPS variables

	count	mean	sd	min	max
signdLM	117244	-.0148408	.732213	-1	1
signRM	117244	.1101122	.6397049	-1	1
difference	117244	.1249531	.8655879	-2	2
female	115413	.5476419	.4977272	0	1
age	115404	46.91606	18.12847	17	100
change_fin_situation	116830	2.041051	.6940744	1	3
hh_inc_month	117244	2.676986	2.134119	-.009	86.70329
d_hh_inc_month	117244	.0820844	1.704277	-65.97318	86.54298
fin_trend3	116449	2.041383	.461874	1	3
inc_trend3	104883	2.351201	.639902	1	3
fin_trend3_lead	85219	2.031789	.4571693	1	3
inc_trend3_lead	104654	2.344583	.6417428	1	3
d_employment_status	117203	2.100305	1.3621	1	6
d_marital_status_legal	116846	3.386466	.6048385	1	4

d_hh_income_month is the change in monthly household income from past to current year in 1,000 EUR for Gesis and 1,000 GBP for BHPS. In GESIS net monthly

[1]The codebooks are available under http://www.gesis.org/unser-angebot/daten-erheben/gesis-panel/gesis-panel-data-usage/ and https://www.iser.essex.ac.uk/bhps/documentation/volb where we use the “individual-level data for respondents” in waves F to R except K.

Table C.2: Summary Statistics of GESIS variables

	count	mean	sd	min	max
signdLM	9553	.0059667	.8239416	-1	1
signRM	9553	.2788653	.7277624	-1	1
difference	9553	.2728986	.9995578	-2	2
female	9553	.5174291	.4997223	0	1
age	9523	49.893	14.11716	20	74
change_fin_situation	9465	3.191337	.8915453	1	5
hh_income	7812	3.185625	1.656383	.45	7
d_hh_inc	6656	.0282527	1.011679	-6.55 6.55	
d_employment_status	8604	1.900628	1.454252	1	6
d_marital_status_legal	8849	3.337326	.600452	1	4

household income is computed from changes in reported categories of income ranges of which we took the middle value to describe the category. The question was: "‘If you take a look at the total income from all members of the household: how high is the monthly average household income today? I.e. the sum of all incomes including pensions and social benefits? Please use the list below"’[2] There are 14 answer categories ranging from "‘700 Euro and less"’ to "‘6000 Euro and more"’. In BHPS, the household monthly income is computed from household income in different categories (labor income, capital income...) that are asked separately. The answer scales for the subcategories are open.

Change_fin_situation is a subjective evaluation of the change in financial situation. In GESIS, the question is: "In your opinion, has your Life improved or deteriorated in the following domains? Financial situation:"’ [3]. There are five answer options: "Considerably improved", "Slightly improved", "Stayed the same", "Slightly deteriorated", "Considerably deteriorated" and "Don’t know". In BHPS, the question was "‘Would you say that you yourself are better off or worse off financially than you were a year ago?"’ and the answer options were "‘Better off"’, "‘Worse off"’, "‘About the same"’ and "‘Don’t know"’.

In both GESIS and BHPS, the question about current employment status contains categories like employed part time, employed full time, unemployed, student, old-age pensioner, unable to work etc. We computed 6 categories for the change in employment situation **d_employment_status** depending on current and previous answer:

[2]The German original version is: "Wenn man nun die Einkünfte aller Mitglieder Ihres Haushalts zusammen nimmt: Wie hoch ist das durchschnittliche monatliche Nettoeinkommen aller Haushaltsmitglieder also die Summe aller Einkünfte einschließlich aller Bezüge und Sozialleistungen insgesamt? Benutzen Sie bitte wieder die Liste".

[3]German original: "Hat sich Ihr Leben Ihrer Meinung nach in den folgenden Bereichen in den letzten 12 Monaten verschlechtert oder verbessert? Finanzielle Situation:"

- **got unemployed**: switched to "‘unemployed"’ from any other category,
- **got employed**: switched to "‘employed"’ or "‘part time employed"’ from any other category,
- **exit labor market**: switched from "‘emplpoyed"’ or "‘unemployed"’ or "‘part time employed"’ to any but those categories,
- **stayed employed**: was in any of the categories "‘emplpoyed"’ or "‘part time employed"’,
- **stayed unemployed**: stayed in "‘unemployed"’,
- **stayed out of labor market**: was in anything but "‘emplpoyed"’ or "‘unemployed"’ or "‘part time employed"’.

In both, GESIS and BHPS, marital status contains different categories like married, divorced, never married, widowed, and separated. We computed the 4 following categories of the change in marital status **d__marital__status__legal**:

- **got married**: switch to married from any other category,
- **marriage ended**: switch to divorced from any other category,
- **stayed married**: stayed in category married,
- **stayed not married**: stayed in any group or switch between these groups: divorced, widowed, separated.

To compute the trend in financial situation or in income for BHPS (in Gesis there are not enough waves in the panel yet) we computed **fin__trend3** and **inc__trend3**. They both are categorical variables with three categories computed for both **change__fin__situation** and **d__hh__income__month**:

- **Negative trend** if there was improvement in t and t-1
- **No trend** if there was anything else than positive or negative trend

- **Positive trend** if there was deterioration in t and t-1

For the expected trend **inc_trend3_lead** and **fin_trend3_lead** we used $t+1$ instead of $t-1$. We also include **gender** and **age**.

Lightspeed data

The following table summarizes the Lightspeed data:

Table C.3: Summary Statistics of Lightspeed Variables

	(1)				
	count	mean	sd	min	max
LM1	1580	6.499367	2.164396	0	10
LM2	1580	6.474051	2.186367	0	10
LM1reassess	389	6.421594	2.291208	0	10
LM1recall	451	6.629712	1.99174	0	10
signdLM	1580	.0050633	.8021344	-1	1
RM	1580	.135443	.7511929	-2	2
signRM	1580	.1208861	.6132952	-1	1
difference	1580	.1158228	.9100816	-2	2
Female	1580	.4892405	.5000425	0	1
Age	1580	45.12975	13.07276	16	65
ch_job	1580	.0696203	.571844	-1	1
ch_money	1580	-.021519	.6119293	-1	1
ch_leisure	1580	.1392405	.5226759	-1	1
ch_relationship	1580	.0987342	.5099376	-1	1
ch_health	1580	-.0727848	.556287	-1	1
ch_family	1580	.1291139	.5312649	-1	1
ch_friends	1580	.1	.4855088	-1	1
ch_neighbour	1580	.0398734	.4588777	-1	1

The questions asked to participant are as follows:

LM1 and **LM2**:

Question: “We would like to ask you how you evaluate your life. Everything taken together, how satisfied are you with your CURRENT life situation?”

Answer: 10 categories from “completely dissatisfied” to “completely satisfied”, with “don’t know” option.

German question: “Im Folgenden würden wir gerne von Ihnen wissen, wie Sie Ihr Leben bewerten. Alles in allem betrachtet, wie zufrieden sind Sie mit Ihrer GEGENWÄRTIGEN Lebenssituation?”

Answer German: “Ganz und gar unzufrieden” – “Ganz und gar zufrieden, weiß nicht”

LM1recall: “How well can you remember the answer you gave during the first questionnaire 3 months ago concerning the satisfaction with your situation back then?”

Answer: “I can remember exactly, it was:”, “I am not sure but I think it was:” and “I cannot remember.”

German question: “Wie gut können Sie sich daran erinnern, welche Antwort Sie in der ersten Befragung vor 3 Monaten bezüglich Ihrer Zufriedenheit mit Ihrer damaligen Lebenssituation gegeben haben?”

Answer German: “Ich kann mich genau erinnern, es war:”, “Ich bin mir nicht sicher, aber ich denke es war:” and “Ich kann mich nicht erinnern.”

LM1reassess:

Question: “We would like to ask you how you evaluate your life AT THE TIME OF THE FIRST QUESTIONNAIRE (beginning of August 2015). Everything taken together, how satisfied were you with your life situation 3 months ago?”

German question: “Im Folgenden würden wir gerne von Ihnen wissen, wie Sie Ihr Leben ZUM ZEITPUNKT DER ERSTEN BEFRAGUNG (Anfang Juli 2015) bewerten. Alles in allem betrachtet, wie zufrieden waren Sie mit Ihrer Lebenssituation vor 3 Monaten?”

RM:

Question: ‘We would like to ask you how you evaluate the CHANGES in your life since the first questionnaire (beginning of August 2015). In your opinion, has your life as a whole since then improved or become worse? Please evaluate any changes from today’s perspective.”

Answer: “Considerably deteriorated”, “Slightly deteriorated”, “Stayed the same”, “Slightly imporved”, “Considerably imporved” and “Don’t know”

German question: “Im Folgenden würden wir gerne von Ihnen wissen, wie Sie die VERÄNDERUNGEN Ihres Lebens seit der ersten Befragung (Anfang Juli 2015) bewerten. Hat sich Ihr Leben Ihrer Meinung nach in dieser Zeit insgesamt verbessert oder verschlechtert? Bewerten Sie die Veränderungen aus heutiger Sicht.”

Answer German: “Deutlich verschlechtert”, “Leicht verschlechtert”, “Gleich geblieben”, “Leicht verbessert”, “Deutlich verbessert” and “Weiß nicht”

C.2 Additional regression results

Table C.4: Lightspeed: Ordered logit regressions on changes in perceived changes in socio-economic circumstances (Treatments: Control and RecallExPost).

	(1) difference	(2) signdLM	(3) signRM
Age	-0.0161**	0.0120*	-0.0111
	(0.00551)	(0.00528)	(0.00653)
Female	0.106	-0.228	-0.204
	(0.135)	(0.134)	(0.156)
ch_job	0.372**	0.161	1.151***
	(0.126)	(0.141)	(0.200)
ch_money	0.309*	0.210	1.059***
	(0.129)	(0.127)	(0.177)
ch_leisure	0.380*	-0.160	0.565**
	(0.167)	(0.141)	(0.208)
ch_relationship	0.108	0.253	0.803***
	(0.150)	(0.145)	(0.236)
ch_health	0.176	0.155	0.689***
	(0.152)	(0.148)	(0.204)
ch_family	0.130	0.220	0.735**
	(0.150)	(0.147)	(0.227)
ch_friends	-0.326	0.0218	-0.598**
	(0.178)	(0.154)	(0.232)
ch_neighbour	0.0202	0.0658	0.182
	(0.160)	(0.159)	(0.249)
Constant cut1	-4.289***	-0.357	-3.028***
	(0.354)	(0.261)	(0.363)
Constant cut2	-1.818***	1.254***	0.973**
	(0.281)	(0.264)	(0.346)
Constant cut3	0.0761		
	(0.274)		
Constant cut4	2.139***		
	(0.303)		
Observations	794	795	796
r2_p	0.0313	0.0168	0.245
p	5.98e-09	0.00311	1.26e-30

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.5: BHPS: Regressions on key socio-economic variables (financial situation) without lags. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects.

	(1) deviation OLS	(2) signdLM OLS	(3) signRM OLS	(4) signRM FE
female	0.0171**	0.00406	0.0212***	0
	(0.00554)	(0.00284)	(0.00510)	(.)
age	-0.0232***	0.00622***	-0.0170***	-0.0106
	(0.000929)	(0.000530)	(0.000832)	(0.0233)
age2	0.000171***	-0.0000555***	0.000115***	0.00000305
	(0.00000925)	(0.00000532)	(0.00000823)	(0.0000181)
change_fin_situation	0.132***	0.0689***	0.201***	0.162***
	(0.00402)	(0.00309)	(0.00341)	(0.00292)
stayed unemployed	-0.204***	0.0353	-0.169***	-0.0996***
	(0.0263)	(0.0189)	(0.0209)	(0.0232)
stayed out of labour market	-0.0448***	0.00543	-0.0393***	0.0227**
	(0.00752)	(0.00430)	(0.00673)	(0.00876)
got employed	-0.00485	0.0612***	0.0563***	0.0821***
	(0.0132)	(0.0113)	(0.0100)	(0.00964)
got unemployed	-0.0861***	-0.0960***	-0.182***	-0.114***
	(0.0206)	(0.0174)	(0.0164)	(0.0148)
exited labor market	0.0469***	0.0472***	0.0941***	0.110***
	(0.0139)	(0.0116)	(0.0111)	(0.0103)
got married	0.205***	0.0513**	0.257***	0.122***
	(0.0193)	(0.0162)	(0.0146)	(0.0147)
marriage ended	-0.110***	-0.0228	-0.133***	-0.196***
	(0.0266)	(0.0242)	(0.0235)	(0.0187)
stayed married	0.0163**	-0.0209***	-0.00459	-0.116***
	(0.00621)	(0.00333)	(0.00571)	(0.0101)
Constant	0.507***	-0.290***	0.218***	0.327
	(0.0250)	(0.0166)	(0.0216)	(0.948)
Year dummies	Yes	Yes	Yes	Yes
Observations	114568	114568	114568	114568
R^2	0.046	0.008	0.107	0.049
Adjusted R^2	0.045	0.008	0.107	-0.175
AIC	286652.3	252794.3	209947.1	163799.3
BIC	286864.6	253006.5	210159.3	164001.9

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.6: BHPS: Regressions on key socio-economic variables (income) without lags. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects.

	(1) deviation OLS	(2) signdLM OLS	(3) signRM OLS	(4) signRM FE
female	0.0154**	0.00323	0.0186***	0
	(0.00568)	(0.00283)	(0.00533)	(.)
age	-0.0262***	0.00454***	-0.0216***	-0.00448
	(0.000940)	(0.000524)	(0.000855)	(0.0237)
age2	0.000194***	-0.0000415***	0.000153***	0.0000305
	(0.00000937)	(0.00000527)	(0.00000845)	(0.0000183)
d_hh_inc_month	-0.000520	0.00310*	0.00258*	0.00232*
	(0.00156)	(0.00153)	(0.00103)	(0.00103)
stayed unemployed	-0.250***	0.0129	-0.237***	-0.172***
	(0.0265)	(0.0190)	(0.0215)	(0.0235)
stayed out of labour market	-0.0705***	-0.00797	-0.0784***	-0.0157
	(0.00761)	(0.00424)	(0.00692)	(0.00886)
got employed	0.0145	0.0695***	0.0840***	0.102***
	(0.0133)	(0.0113)	(0.0104)	(0.00978)
got unemployed	-0.163***	-0.132***	-0.295***	-0.209***
	(0.0205)	(0.0174)	(0.0170)	(0.0149)
exited labor market	-0.00545	0.0207	0.0152	0.0461***
	(0.0138)	(0.0116)	(0.0111)	(0.0104)
got married	0.207***	0.0524**	0.260***	0.123***
	(0.0193)	(0.0163)	(0.0148)	(0.0149)
marriage ended	-0.153***	-0.0401	-0.193***	-0.243***
	(0.0265)	(0.0241)	(0.0240)	(0.0189)
stayed married	0.0121	-0.0223***	-0.0102	-0.124***
	(0.00636)	(0.00332)	(0.00594)	(0.0103)
Constant	0.877***	-0.0953***	0.781***	0.388
	(0.0226)	(0.0142)	(0.0198)	(0.964)
3.wave	Yes	Yes	Yes	Yes
Observations	114980	114980	114980	114980
R^2	0.035	0.004	0.064	0.018
Adjusted R^2	0.035	0.004	0.064	-0.214
AIC	288942.9	254197.4	216222.7	168243.2
BIC	289155.3	254409.7	216435.0	168445.9

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.7: BHPS: Regressions on key socio-economic variables (financial situation) including only the three first years of the panel, making it comparable to the current time horizon of GESIS. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects.

	(1) deviation OLS	(2) signdLM OLS	(3) signRM OLS	(4) signRM FE
female	0.00461	0.00857	0.0132	0
	(0.0126)	(0.00940)	(0.0104)	(.)
age	-0.0219***	0.00483**	-0.0171***	-0.0198
	(0.00232)	(0.00174)	(0.00183)	(0.0264)
age2	0.000175***	-0.0000539**	0.000121***	-0.000124
	(0.0000229)	(0.0000174)	(0.0000177)	(0.000264)
change_fin_situation	0.101***	0.0821***	0.184***	0.162***
	(0.00998)	(0.00862)	(0.00799)	(0.0118)
L.change_fin_situation	0.100***	-0.0591***	0.0411***	0.00365
	(0.00963)	(0.00842)	(0.00750)	(0.0120)
stayed unemployed	-0.114	-0.000554	-0.115*	-0.177
	(0.0663)	(0.0577)	(0.0550)	(0.116)
stayed out of labour market	-0.0216	-0.00405	-0.0257	0.0146
	(0.0189)	(0.0143)	(0.0151)	(0.0684)
got employed	0.0255	0.0917**	0.117***	0.160**
	(0.0363)	(0.0325)	(0.0275)	(0.0560)
got unemployed	-0.0618	-0.0774	-0.139**	-0.0947
	(0.0547)	(0.0483)	(0.0463)	(0.0699)
exited labor market	0.103**	0.0278	0.131***	0.170***
	(0.0359)	(0.0318)	(0.0289)	(0.0456)
lag got employed	0.0311	0.0190	0.0501*	0.0248
	(0.0316)	(0.0280)	(0.0240)	(0.0397)
lag got unemployed	-0.0645	0.0118	-0.0527	-0.0619
	(0.0572)	(0.0517)	(0.0439)	(0.0666)
lag exited labor market	0.142***	-0.0633	0.0783**	0.0225
	(0.0378)	(0.0328)	(0.0271)	(0.0454)
got married	0.237***	0.0352	0.272***	0.107
	(0.0520)	(0.0449)	(0.0392)	(0.0694)
marriage ended	-0.156**	-0.00368	-0.159**	-0.300**
	(0.0593)	(0.0565)	(0.0544)	(0.0949)
stayed married	-0.0372*	0.00653	-0.0307*	-0.289**
	(0.0148)	(0.0112)	(0.0121)	(0.0992)
lag got married	0.119*	0.0570	0.176***	0.210***
	(0.0511)	(0.0417)	(0.0380)	(0.0629)
lag marriage ended	-0.0802	0.117*	0.0370	-0.0799
	(0.0621)	(0.0556)	(0.0477)	(0.0728)
Constant	0.273***	-0.112*	0.161**	1.156
	(0.0620)	(0.0463)	(0.0510)	(0.653)
Year dummies	Yes	Yes	Yes	Yes
Observations	16426	16426	16426	16426
R^2	0.050	0.017	0.112	0.046
Adjusted R^2	0.049	0.016	0.111	-1.231
AIC	40450.0	36069.9	29755.0	12099.0
BIC	40604.2	36224.0	29909.1	12237.7

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.8: BHPS: Regressions on key socio-economic variables (income) including only the three first years of the panel, making it comparable to the current time horizon of GESIS. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects.

	(1) deviation OLS	(2) signdLM OLS	(3) signRM OLS	(4) signRM FE
female	0.00271	0.00892	0.0116	0
	(0.0128)	(0.00940)	(0.0107)	(.)
age	-0.0261***	0.00423*	-0.0219***	-0.0197
	(0.00232)	(0.00173)	(0.00186)	(0.0267)
age2	0.000207***	-0.0000477**	0.000160***	-0.000136
	(0.0000230)	(0.0000173)	(0.0000180)	(0.000268)
d_hh_inc_month	0.00252	0.00174	0.00426	0.00206
	(0.00394)	(0.00377)	(0.00287)	(0.00380)
L.d_hh_inc_month	0.000442	0.000271	0.000713	-0.00325
	(0.00401)	(0.00347)	(0.00290)	(0.00466)
stayed unemployed	-0.187**	-0.0107	-0.198***	-0.216
	(0.0659)	(0.0585)	(0.0563)	(0.116)
stayed out of labour market	-0.0645***	-0.0113	-0.0758***	-0.0478
	(0.0189)	(0.0141)	(0.0153)	(0.0690)
got employed	0.00764	0.121***	0.129***	0.180**
	(0.0364)	(0.0327)	(0.0285)	(0.0566)
got unemployed	-0.138*	-0.105*	-0.243***	-0.181*
	(0.0545)	(0.0489)	(0.0480)	(0.0705)
exited labor market	0.0468	0.00380	0.0506	0.103*
	(0.0358)	(0.0316)	(0.0293)	(0.0458)
lag got employed	0.0548	0.0121	0.0668**	0.0393
	(0.0321)	(0.0282)	(0.0249)	(0.0403)
lag got unemployed	-0.0920	0.0306	-0.0614	-0.0518
	(0.0583)	(0.0523)	(0.0458)	(0.0675)
lag exited labor market	0.112**	-0.0514	0.0604*	0.0223
	(0.0379)	(0.0326)	(0.0273)	(0.0457)
got married	0.253***	0.0345	0.288***	0.100
	(0.0522)	(0.0452)	(0.0400)	(0.0699)
marriage ended	-0.195**	-0.0297	-0.225***	-0.373***
	(0.0601)	(0.0565)	(0.0570)	(0.0960)
stayed married	-0.0351*	0.00698	-0.0281*	-0.280**
	(0.0150)	(0.0111)	(0.0124)	(0.100)
lag got married	0.116*	0.0510	0.167***	0.180**
	(0.0513)	(0.0419)	(0.0389)	(0.0637)
lag marriage ended	-0.123	0.123*	0.0000408	-0.122
	(0.0630)	(0.0553)	(0.0492)	(0.0736)
Constant	0.824***	-0.0509	0.773***	1.543*
	(0.0528)	(0.0391)	(0.0426)	(0.658)
3.wave	Yes	Yes	Yes	Yes
Observations	16506	16506	16506	16506
R^2	0.034	0.010	0.067	0.015
Adjusted R^2	0.033	0.009	0.066	-1.299
AIC	40931.4	36359.6	30748.7	12757.2
BIC	41085.6	36513.8	30902.9	12896.0

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.9: GESIS: Regressions on key socio-economic variables (income). Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects.

	(1) deviation OLS	(2) signdLM OLS	(3) signRM OLS	(4) signRM FE
female	-0.0374	0.00256	-0.0349	0
	(0.0318)	(0.0236)	(0.0266)	(.)
age	-0.0280*	0.00232	-0.0257**	-0.0796
	(0.0114)	(0.00879)	(0.00939)	(0.102)
age2	0.000188	-0.0000316	0.000156	0.000643
	(0.000116)	(0.0000890)	(0.0000960)	(0.000951)
d_hh_inc	0.0318	0.00866	0.0405**	0.0403*
	(0.0199)	(0.0170)	(0.0140)	(0.0186)
L.d_hh_inc	0.0215	0.00218	0.0236	0.0154
	(0.0173)	(0.0147)	(0.0133)	(0.0170)
stayed unemployed	-0.106	-0.516**	-0.623**	-0.440
	(0.277)	(0.177)	(0.214)	(0.430)
stayed out of labour market	-0.179**	0.0368	-0.142**	-0.0884
	(0.0567)	(0.0428)	(0.0461)	(0.179)
got employed	0.0480	-0.0862	-0.0382	-0.212
	(0.0960)	(0.0784)	(0.0726)	(0.144)
got unemployed	-0.343	0.0343	-0.309*	-0.454*
	(0.185)	(0.149)	(0.150)	(0.217)
exited labor market	-0.0492	0.0432	-0.00595	-0.0195
	(0.0859)	(0.0666)	(0.0644)	(0.106)
lag got employed	0.0713	0.0276	0.0989	0.0864
	(0.0777)	(0.0667)	(0.0548)	(0.0935)
lag got unemployed	-0.0303	0.0523	0.0221	-0.0401
	(0.222)	(0.179)	(0.159)	(0.241)
lag exited labor market	0.116	-0.00752	0.108	0.0726
	(0.0845)	(0.0699)	(0.0614)	(0.110)
got married	0.00865	0.0611	0.0698	-0.185
	(0.149)	(0.125)	(0.108)	(0.207)
marriage ended	0.0473	0.0365	0.0838	0.0593
	(0.166)	(0.128)	(0.120)	(0.225)
stayed married	-0.0319	-0.0120	-0.0440	-0.301
	(0.0410)	(0.0323)	(0.0335)	(0.254)
lag got married	0.0222	0.0794	0.102	0.0267
	(0.114)	(0.0995)	(0.0846)	(0.134)
lag marriage ended	-0.00834	0.0705	0.0621	0.158
	(0.137)	(0.108)	(0.104)	(0.165)
Year dummies	yes	yes	yes	yes
Constant	1.308***	-0.0595	1.248***	2.827
	(0.268)	(0.210)	(0.220)	(2.703)
Observations	3442	3442	3442	3442
R^2	0.032	0.005	0.064	0.018
Adjusted R^2	0.027	-0.001	0.059	-1.535
AIC	9486.4	8260.7	7301.9	2899.1
BIC	9609.3	8383.6	7424.8	3009.7

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.10: BHPS: Regressions on key socio-economic variables (financial situation). Variable description in Appendix A.1. Results from ordered logistic regressions including year fixed effects.

	(1) deviation	(2) signdLM	(3) signRM
female	0.0366**	0.00695	0.0758***
	(0.0132)	(0.0133)	(0.0142)
age	-0.0469***	0.0123***	-0.0586***
	(0.00236)	(0.00237)	(0.00253)
age2	0.000354***	-0.000119***	0.000410***
	(0.0000233)	(0.0000235)	(0.0000248)
change_fin_situation	0.226***	0.222***	0.655***
	(0.0100)	(0.0101)	(0.0113)
L.change_fin_situation	0.230***	-0.156***	0.154***
	(0.0101)	(0.0101)	(0.0110)
stayed unemployed	-0.271***	-0.0340	-0.435***
	(0.0728)	(0.0742)	(0.0782)
stayed out of labour market	-0.0821***	0.00896	-0.121***
	(0.0190)	(0.0191)	(0.0204)
got employed	0.113**	0.115**	0.378***
	(0.0380)	(0.0384)	(0.0417)
got unemployed	-0.198***	-0.264***	-0.648***
	(0.0514)	(0.0527)	(0.0570)
exited labor market	0.166***	0.111**	0.419***
	(0.0356)	(0.0360)	(0.0397)
lag got employed	0.0417	0.0120	0.0910*
	(0.0329)	(0.0330)	(0.0358)
lag got unemployed	-0.356***	0.147*	-0.353***
	(0.0573)	(0.0584)	(0.0626)
lag exited labor market	0.246***	-0.103**	0.234***
	(0.0365)	(0.0365)	(0.0395)
got married	0.427***	0.139**	0.861***
	(0.0509)	(0.0507)	(0.0566)
marriage ended	-0.308***	-0.0668	-0.609***
	(0.0664)	(0.0694)	(0.0763)
stayed married	-0.0000336	-0.0234	-0.0439**
	(0.0151)	(0.0152)	(0.0163)
lag got married	0.434***	-0.0792	0.547***
	(0.0515)	(0.0512)	(0.0562)
lag marriage ended	-0.112	0.249***	0.155*
	(0.0643)	(0.0654)	(0.0699)
Constant cut1	-3.975***	-0.720***	-1.971***
	(0.0695)	(0.0668)	(0.0721)
Constant cut2	-1.592***	1.339***	1.044***
	(0.0666)	(0.0669)	(0.0720)
Constant cut3	0.505***		
	(0.0664)		
Constant cut4	2.889***		
	(0.0680)		
Year dummies	Yes	Yes	Yes
Observations	83533	83533	83533
percent_correct		0.414	0.520
Log likelihood	-103336.9	-87923.9	-74544.3
chi2	4253.6	1004.5	10233.9
chi2type	LR	LR	LR
df_m	25	25	25
p	0	6.47e-196	0

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.11: Deviation between measures on direction of change and trends in financial situation using leads instead of lags to capture expected trends (BHPS). Description of variables given in Appendix A.1. Results from pooled OLS regression including year fixed effects.

	(1)	
	deviation OLS	
female	0.0186**	(0.00700)
age	-0.0208***	(0.00126)
age2	0.000160***	(0.0000125)
financial deterioration	-0.0491***	(0.0114)
lag financial deterioration	-0.0985***	(0.00930)
financial improvement	0.132***	(0.0105)
lag financial improvement	0.0977***	(0.00879)
negative trend (lead)	-0.0219	(0.0152)
positive trend (lead)	0.0329*	(0.0136)
stayed unemployed	-0.0891*	(0.0388)
stayed out of labour market	-0.0412***	(0.0100)
got employed	0.0369	(0.0199)
got unemployed	-0.0893**	(0.0275)
exited labor market	0.0562**	(0.0187)
lag got employed	0.0191	(0.0174)
lag got unemployed	-0.138***	(0.0297)
lag exited labor market	0.114***	(0.0191)
got married	0.212***	(0.0264)
marriage ended	-0.162***	(0.0340)
stayed married	0.00147	(0.00797)
lag got married	0.173***	(0.0269)
lag marriage ended	-0.0389	(0.0335)
Constant	0.621***	(0.0312)
Year dummies	Yes	
Observations	60133	
R^2	0.049	
Adjusted R^2	0.049	
AIC	148716.0	
BIC	148968.1	

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.12: BHPS: Regressions on key socio-economic variables (log of income). Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects.

	(1) deviation OLS	(2) signdLM OLS	(3) signRM OLS	(4) signRM FE
female	0.0156*	0.00232	0.0179**	0
	(0.00637)	(0.00338)	(0.00594)	(.)
age	-0.0252***	0.00403***	-0.0212***	-0.0133
	(0.00111)	(0.000648)	(0.000997)	(0.0331)
age2	0.000191***	-0.0000402***	0.000151***	0.0000181
	(0.0000109)	(0.00000642)	(0.00000968)	(0.0000235)
d_log_hh_inc_month	0.00728	0.00867	0.0159***	0.0173***
	(0.00545)	(0.00493)	(0.00431)	(0.00398)
L.d_log_hh_inc_month	0.0126*	-0.00974*	0.00283	0.00302
	(0.00552)	(0.00490)	(0.00424)	(0.00400)
stayed unemployed	-0.173***	-0.0282	-0.201***	-0.165***
	(0.0339)	(0.0251)	(0.0290)	(0.0336)
stayed out of labour market	-0.0827***	-0.0000113	-0.0827***	-0.00344
	(0.00902)	(0.00517)	(0.00817)	(0.0123)
got employed	0.0389*	0.0694***	0.108***	0.135***
	(0.0178)	(0.0151)	(0.0141)	(0.0143)
got unemployed	-0.164***	-0.124***	-0.289***	-0.205***
	(0.0254)	(0.0217)	(0.0211)	(0.0191)
exited labor market	0.0210	0.0173	0.0383**	0.0757***
	(0.0168)	(0.0142)	(0.0135)	(0.0129)
lag got employed	0.0360*	0.00525	0.0412***	0.0483***
	(0.0155)	(0.0131)	(0.0121)	(0.0118)
lag got unemployed	-0.194***	0.0785***	-0.115***	-0.0621**
	(0.0280)	(0.0238)	(0.0216)	(0.0205)
lag exited labor market	0.0835***	-0.0373**	0.0462***	0.0295*
	(0.0171)	(0.0143)	(0.0129)	(0.0129)
got married	0.200***	0.0527**	0.253***	0.115***
	(0.0233)	(0.0195)	(0.0177)	(0.0183)
marriage ended	-0.186***	-0.0389	-0.225***	-0.301***
	(0.0317)	(0.0289)	(0.0284)	(0.0238)
stayed married	-0.00820	-0.00946*	-0.0177**	-0.186***
	(0.00729)	(0.00404)	(0.00674)	(0.0140)
lag got married	0.198***	-0.0365	0.162***	0.153***
	(0.0237)	(0.0199)	(0.0184)	(0.0181)
lag marriage ended	-0.0823**	0.0970***	0.0147	-0.0397
	(0.0316)	(0.0266)	(0.0236)	(0.0224)
Constant	0.808***	-0.0483**	0.760***	0.795
	(0.0267)	(0.0168)	(0.0236)	(1.391)
Year dummies	Yes	Yes	Yes	Yes
Observations	82908	82908	82908	82908
R^2	0.036	0.005	0.064	0.020
Adjusted R^2	0.035	0.005	0.064	-0.262
AIC	206721.6	182117.6	154677.8	116036.7
BIC	206964.0	182360.0	154920.3	116269.8

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.13: BHPS: Regressions on key socio-economic variables (financial situation) excluding observations where the current LM is either 1 or 7. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects.

	(1) deviation OLS	(2) signdLM OLS	(3) signRM OLS	(4) signRM FE
female	0.0222**	-0.00645	0.0158**	0
	(0.00703)	(0.00421)	(0.00596)	(.)
age	-0.0243***	0.0102***	-0.0141***	-0.0131
	(0.00129)	(0.000833)	(0.00103)	(0.0374)
age2	0.000200***	-0.000115***	0.0000844***	-0.0000260
	(0.0000129)	(0.00000849)	(0.0000100)	(0.0000271)
change_fin_situation	0.115***	0.0746***	0.189***	0.170***
	(0.00513)	(0.00425)	(0.00411)	(0.00385)
L.change_fin_situation	0.111***	-0.0736***	0.0376***	0.0161***
	(0.00507)	(0.00426)	(0.00379)	(0.00385)
stayed unemployed	-0.0968*	-0.0519	-0.149***	-0.0863*
	(0.0377)	(0.0281)	(0.0289)	(0.0369)
stayed out of labour market	-0.0362***	-0.00166	-0.0379***	0.0428**
	(0.0103)	(0.00660)	(0.00830)	(0.0137)
got employed	0.0605**	0.0303	0.0908***	0.124***
	(0.0191)	(0.0160)	(0.0147)	(0.0156)
got unemployed	-0.0978***	-0.0913***	-0.189***	-0.124***
	(0.0274)	(0.0228)	(0.0212)	(0.0206)
exited labor market	0.0946***	0.0278	0.122***	0.153***
	(0.0187)	(0.0155)	(0.0145)	(0.0144)
lag got employed	0.0218	0.00142	0.0232	0.0405**
	(0.0163)	(0.0137)	(0.0124)	(0.0127)
lag got unemployed	-0.169***	0.0756**	-0.0933***	-0.0542*
	(0.0302)	(0.0253)	(0.0223)	(0.0224)
lag exited labor market	0.117***	-0.0414**	0.0754***	0.0424**
	(0.0191)	(0.0157)	(0.0137)	(0.0144)
got married	0.205***	0.0206	0.226***	0.104***
	(0.0255)	(0.0209)	(0.0190)	(0.0201)
marriage ended	-0.139***	-0.0115	-0.151***	-0.235***
	(0.0338)	(0.0304)	(0.0298)	(0.0260)
stayed married	0.00380	-0.0292***	-0.0254***	-0.185***
	(0.00808)	(0.00497)	(0.00681)	(0.0152)
lag got married	0.192***	-0.0437*	0.149***	0.148***
	(0.0257)	(0.0213)	(0.0194)	(0.0199)
lag marriage ended	-0.0494	0.0971***	0.0477	-0.00475
	(0.0341)	(0.0290)	(0.0247)	(0.0246)
Constant	0.276***	-0.197***	0.0785**	0.422
	(0.0350)	(0.0236)	(0.0287)	(1.517)
Year dummies	Yes	Yes	Yes	Yes
Observations	71449	71449	71449	71449
R^2	0.051	0.016	0.110	0.055
Adjusted R^2	0.050	0.015	0.110	-0.252
AIC	180607.9	157794.0	131606.7	99764.0
BIC	180846.5	158032.6	131845.3	99993.4

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.14: BHPS: Regressions on key socio-economic variables (financial situation) excluding observations where the lag of LM is either 1 or 7. Variable description in Appendix A.1. Results from pooled OLS and fixed effects regressions including year fixed effects.

	(1) deviation OLS	(2) signdLM OLS	(3) signRM OLS	(4) signRM FE
female	0.00286	0.0159***	0.0187**	0
	(0.00652)	(0.00420)	(0.00604)	(.)
age	-0.0143***	-0.000441	-0.0147***	-0.0248
	(0.00117)	(0.000841)	(0.00105)	(0.0362)
age2	0.0000730***	0.0000183*	0.0000914***	-0.0000126
	(0.0000116)	(0.00000862)	(0.0000102)	(0.0000274)
change_fin_situation	0.0956***	0.0975***	0.193***	0.171***
	(0.00494)	(0.00428)	(0.00414)	(0.00386)
L.change_fin_situation	0.0866***	-0.0475***	0.0391***	0.0172***
	(0.00484)	(0.00426)	(0.00382)	(0.00387)
stayed employed	0	0	0	0
	(.)	(.)	(.)	(.)
stayed unemployed	-0.161***	0.0125	-0.148***	-0.0817*
	(0.0359)	(0.0288)	(0.0289)	(0.0375)
stayed out of labour market	-0.0489***	0.00945	-0.0395***	0.0454***
	(0.00925)	(0.00655)	(0.00845)	(0.0138)
got employed	0.0396*	0.0587***	0.0983***	0.128***
	(0.0185)	(0.0161)	(0.0148)	(0.0157)
got unemployed	-0.100***	-0.100***	-0.201***	-0.126***
	(0.0257)	(0.0231)	(0.0215)	(0.0208)
exited labor market	0.0710***	0.0543***	0.125***	0.164***
	(0.0178)	(0.0155)	(0.0146)	(0.0143)
lag got employed	0.00345	0.0144	0.0179	0.0368**
	(0.0157)	(0.0138)	(0.0124)	(0.0128)
lag got unemployed	-0.139***	0.0428	-0.0957***	-0.0660**
	(0.0287)	(0.0253)	(0.0225)	(0.0225)
lag exited labor market	0.0922***	-0.0278	0.0644***	0.0386**
	(0.0182)	(0.0158)	(0.0141)	(0.0146)
got married	0.150***	0.0823***	0.232***	0.109***
	(0.0240)	(0.0207)	(0.0190)	(0.0201)
marriage ended	-0.132***	-0.00616	-0.138***	-0.222***
	(0.0330)	(0.0310)	(0.0303)	(0.0264)
stayed married	-0.0299***	0.00964	-0.0203**	-0.182***
	(0.00749)	(0.00495)	(0.00692)	(0.0153)
stayed unmarried	0	0	0	0
	(.)	(.)	(.)	(.)
lag got married	0.170***	-0.0194	0.151***	0.155***
	(0.0250)	(0.0217)	(0.0199)	(0.0201)
lag marriage ended	-0.0529	0.0974***	0.0445	-0.0167
	(0.0322)	(0.0286)	(0.0249)	(0.0248)
Constant	0.122***	-0.0335	0.0888**	0.870
	(0.0328)	(0.0238)	(0.0292)	(1.464)
Year dummies	Yes	Yes	Yes	Yes
Observations	70759	70759	70759	70759
R^2	0.057	0.014	0.111	0.055
Adjusted R^2	0.057	0.014	0.111	-0.254
AIC	172994.9	156368.7	131099.4	98790.1
BIC	173233.3	156607.1	131337.8	99019.3

Standard error in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix D

Appendix Chapter 5

D.1 Summary statistics

Table D.1: Summary Statistics of BHPS Variables 1997-2003 (2001 missing)

	(1)				
	count	mean	sd	min	max
LM	143580	5.230882	1.286298	1	7
dLM	112550	-.0248512	1.170323	-6	6
signdLM	112550	-.0170413	.7318353	-1	1
RM	142262	.1266185	.645752	-1	1
Men	143580	.4540674	.4978875	0	1
Age	143580	45.94092	18.45143	16	100
Seeing friends and family	143580	3.307508	.759751	1	4
Talking to neighbours	143580	3.06285	.9405127	1	4
Personal income per month / 1000	143580	1.208245	1.190314	0	56.91667
Real household income per year and capita / 1000	138956	10.46172	7.454709	.0002583	99.02111
Health status	143580	3.76923	.9656901	1	5
Days in hospital	143580	.8963226	6.036168	0	320
Education: O- or A-level	143580	.4473813	.4972253	0	1
Education: university level	143580	.1991015	.399326	0	1
Own house	143580	.2763616	.4471994	0	1
number of own children in household	143580	.5149742	.928536	0	9
number of people in household	143580	2.859695	1.38858	1	16
married	143580	.5333403	.4988889	0	1
living as couple	143580	.11542	.3195292	0	1
widowed	143580	.0724405	.2592167	0	1
divorced	143580	.0557738	.2294852	0	1
separated	143580	.0172587	.1302341	0	1
job: unemployed	143580	.0332915	.1793974	0	1
job: self-employed	143580	.0673283	.2505906	0	1
job: retired	143580	.2091796	.4067243	0	1
job: student	143580	.0593397	.2362603	0	1
job: disabled	143580	.0439407	.2049638	0	1
job: look after home	143580	.0745438	.2626547	0	1
region / metropolitan area	143580	11.71494	6.398044	1	19
wave	143580	7.770093	3.433629	2	13

Table D.2: Personal income per month by LM

	LM							
Pers. inc. p. month	1	2	4	4	5	6	7	Total
0-10	1,988	2,970	8,336	19,328	40,829	45,529	19,784	138,764
10-20	1	1	3	19	55	76	9	164
20-30	0	0	0	1	4	9	5	19
30-40	0	0	0	1	1	4	2	8
50-60	0	0	0	0	0	1	0	1
Total	1,989	2,971	8,339	19,349	40,889	45,619	19,800	138,956

Source: BHPS 1997-2009

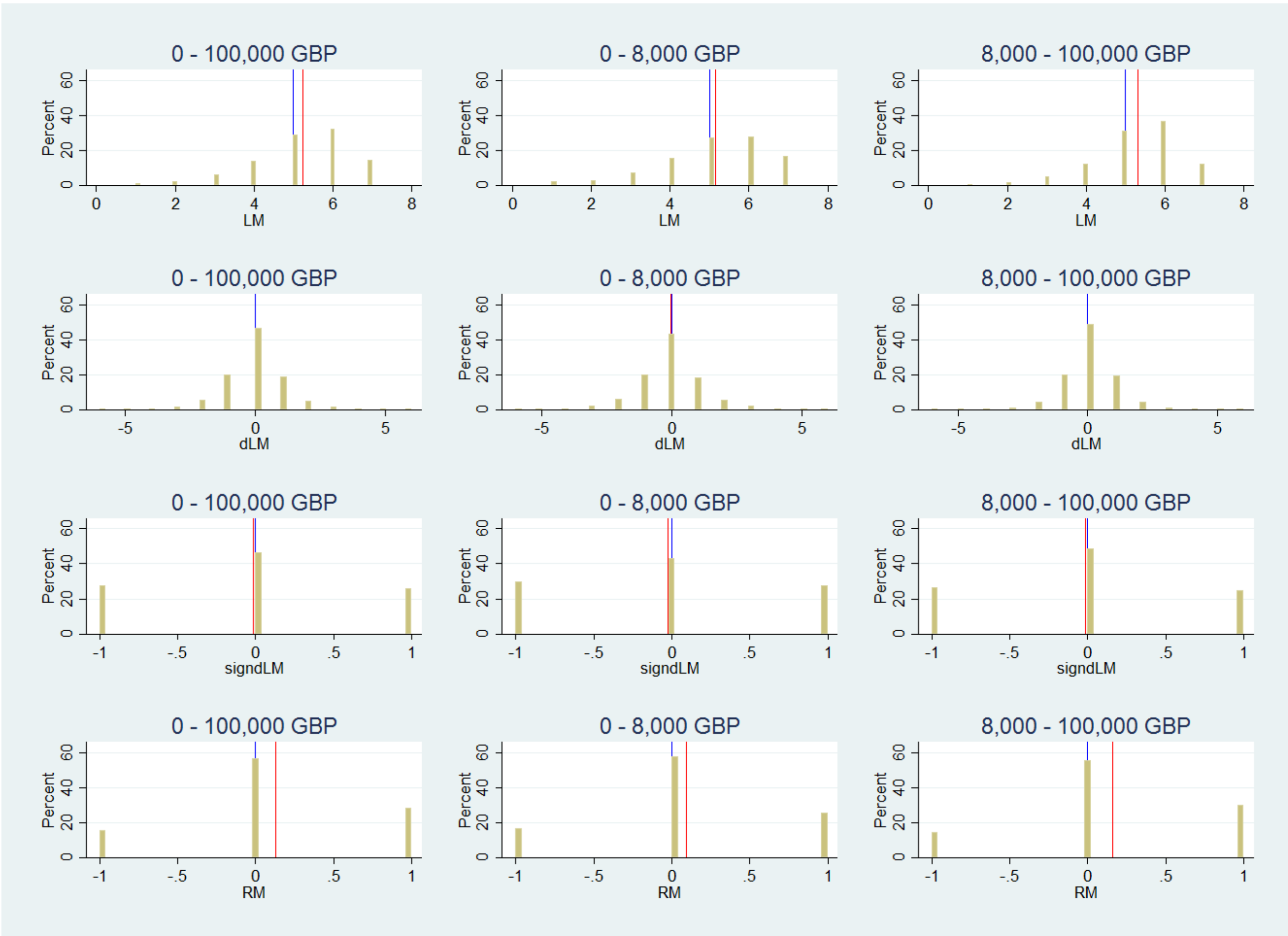

Figure D.1: Distribution of life satisfaction measures depending on levels of *Real household income per year and capita* (BHPS 1997-2009). Blue is median, red is the mean.

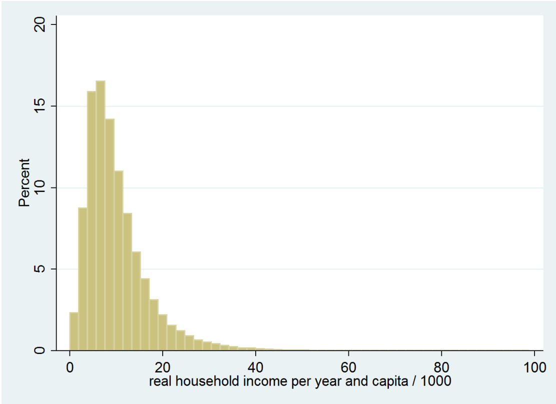

Figure D.2: Histogram of *Real household income per capita* in 1000 GBP (1997). 166,214 observations. Values below 0 and above 100,000 GBP were excluded (5,389 and 87 observations respectively).

D.2 Results of Powdthavee (2008)

Social networks and life satisfaction equations: pooled BHPS, 1997–2003

	Ordered probit				OLS	
	(1)		(2)		(3)	
	Beta	*Z*-ratio	Beta	*Z*-ratio	Beta	*T*-ratio
(i) Meet friends and relatives						
Once or twice a month	0.245***	6.50	0.155***	4.27	0.210***	4.85
Once or twice a week	0.297***	8.12	0.237***	6.69	0.303***	7.17
On most days	0.323***	8.74	0.312***	8.69	0.383***	8.94
(ii) Talk to neighbours						
Once or twice a month	0.068***	3.36	0.009	0.44	0.028	1.21
Once or twice a week	0.176***	9.11	0.114***	5.88	0.142***	6.47
On most days	0.202***	9.86	0.204***	9.85	0.236***	10.11
(iii) Socioeconomic variables						
Men	−0.022	−1.56	−0.062***	−4.42	−0.054***	−3.60
Age	−0.048***	−15.82	−0.058***	−13.75	−0.060***	−13.11
Age-squared/100	0.062***	16.23	0.073***	13.89	0.075***	13.17
Real household income/1000			0.005***	5.74	0.006***	6.31
Married			0.342***	14.94	0.384***	14.94
Living as couple			0.289***	12.82	0.323***	12.86
Separated			−0.270***	−6.29	−0.346***	−6.58
Divorced			−0.108***	−3.01	−0.147***	−3.47
Widowed			−0.100	−1.63	−0.104	−1.60
Unemployed			−0.332***	−11.00	−0.445***	−12.35
Self-employed			0.023	0.96	0.017	0.66
Retired			0.077**	2.26	0.044	1.23
Student			0.005	0.19	0.015	0.53
Disabled			−0.305***	−8.20	−0.432***	−9.61
Look after home			−0.030	−1.15	−0.075***	−2.60
Health: poor			0.440***	9.78	0.613***	10.57
Health: fair			0.819***	18.28	1.086***	18.88
Health: good			1.183***	26.18	1.503***	26.06
Health: excellent			1.515***	32.64	1.824***	31.23
Education: O-level, A-level			−0.087***	−4.53	−0.068***	−3.22
Education: University level			−0.126***	−6.31	−0.102***	−4.67
Household size			0.016**	2.48	0.017**	2.48
Own home outright			0.093***	4.90	0.097***	4.74
Days spent in hospital last year			−0.002**	−2.15	−0.004***	−2.67
Number of children			−0.061***	−6.29	−0.066***	−6.15
Constant					4.054***	32.57
Cut point_1	−2.536	(0.070)	−1.738	(0.110)		
Cut point_2	−2.144	(0.068)	−1.279	(0.110)		
Cut point_3	−1.644	(0.067)	−0.703	(0.109)		
Cut point_4	−1.050	(0.067)	−0.029	(0.109)		
Cut point_5	−0.232	(0.067)	0.871	(0.109)		
Cut point_6	0.825	(0.067)	1.994	(0.109)		
Regional dummies	No		Yes		Yes	
Wave dummies	No		Yes		Yes	
N	54,355		54,207		54,207	
Pseudo *R*-squared	0.0067		0.0583			
Log-likelihood	−85920		−81435			
R-squared					0.1848	

Note: Standard errors are in parentheses. Reference groups—meet with friends or relatives: less than once a month to never; female; single; health: very poor; education: no formal education; employed. Standard errors are clustered by personal identification in all specifications. Wave dummies represent the year of the interview. **<5%, ***<1%.

Figure D.3: *Table 2 from Powdthavee (2008).*

Valuations of social network status and other life events using life satisfaction equations (ordered probit and OLS regressions): BHPS, 1997–2003

	Ordered probit	OLS
(i) Meet friends and relatives		
Once or twice a month	£31,000	£35,000
Once or twice a week	£47,400	£50,500
On most days	£62,400	£63,833
(ii) Talk to neighbours		
Once or twice a month	n.s.	n.s.
Once or twice a week	£22,800	£23,667
On most days	£40,800	£39,333
(iii) Socioeconomic variables		
Married	£68,400	£64,000
Living as couple	£57,800	£53,833
Separated	−£54,000	−£57,667
Divorced	−£21,600	−£24,500
Widowed	n.s.	n.s.
Unemployed	−£66,400	−£74,167
Disabled	−£61,000	−£72,000
Health: good	£237,000	£251,000
Health: excellent	£303,000	£304,000

Note: n.s. = not significantly different from zero. The estimates are taken from the ordered probit and OLS regressions in Table 2. The valuations are annually figures, and are measured in per capita. Hence, a move from meeting friends or relatives less than once or twice a month to meeting friends or relatives on most day is worth around £62,000–64,000 a year, *ceteris paribus*. Average real annual household income per capita at the cross-section (in 1996 Pounds) is around £9800. At the time of writing, the value of one pound sterling is approximately 1.5 US$.

Figure D.4: *Table 3 from Powdthavee (2008).*

Social networks and life satisfaction equation: BHPS, 1991–2003 (fixed effects regression)

	All		Changes from OLS (in %)
	Beta	*T*-ratio	
(i) Meet friends and relatives			
Once or twice a month	0.115***	3.38	*−45*
Once or twice a week	0.139***	4.18	*−54*
On most days	0.170***	5.03	*−56*
(ii) Talk to neighbours			
Once or twice a month	0.013	0.63	*−54*
Once or twice a week	0.045**	2.18	*−68*
On most days	0.074***	3.35	*−67*
(iii) Socioeconomic variables			
Age	−0.069***	−3.67	*−15*
Age-squared/100	0.058***	5.27	*−23*
Real household income/1000	0.002**	1.96	*−67*
Married	0.101**	2.44	*−74*
Living as couple	0.165***	5.03	*−49*
Separated	−0.278***	−4.83	*+20*
Divorced	−0.072	−1.28	*+51*
Widowed	−0.417***	−4.51	*−301*
Unemployed	−0.286***	−10.02	*+36*
Self-employed	−0.012	−0.37	*−171*
Retired	−0.053	−1.47	*−220*
Student	0.077**	2.40	*+413*
Disabled	−0.331***	−8.38	*+23*
Look after home	−0.118***	−4.33	*−57*
Health: poor	0.403***	10.96	*−34*
Health: fair	0.657***	17.52	*−40*
Health: good	0.847***	22.03	*−44*
Health: excellent	0.967***	24.06	*−47*
Education: O-level, A-level	0.031	0.65	*+146*
Education: University level	0.058	1.22	*+156*
Household size	−0.017**	−2.00	*−200*
Own home outright	0.038	1.51	*−61*
Days spent in hospital last year	−0.001	−1.33	*+75*
Number of children	0.006	0.44	*+110*
Constant	5.795***	9.16	*+43*
Regional dummies	Yes		
Wave dummies	Yes		
N	54,207		
R-squared	0.0627		
F test that fixed-effects = 0	3.49	[0.000]	

Note: Standard errors are in parentheses. Change from OLS (in %) refers to the change in the coefficient size from the last column of Table 2. **<5%, ***<1%. Percentage changes are in italics.

Figure D.5: *Table 5 from Powdthavee (2008).*

Valuations of changes in social network status and other life events using life satisfaction equations (fixed effects regression): BHPS, 1997–2003

	Fixed effects	Differences from OLS (in %)
(i) Meet friends and relatives		
Once or twice a month	£57,500	64
Once or twice a week	£69,500	38
On most days	£85,000	33
(ii) Talk to neighbours		
Once or twice a month	n.s.	n.a.
Once or twice a week	£22,500	−5
On most days	£37,000	−6
(iii) Socioeconomic variables		
Married	£50,500	−21
Living as couple	£82,500	53
Separated	−£139,000	−141
Divorced	n.s.	n.a.
Widowed	−£208,500	n.a.
Unemployed	−£143,000	−93
Disabled	−£165,500	−130
Health: good	£423,500	69
Health: excellent	£483,000	59

Note: n.s. = not significantly different from zero; n.a. = not applicable. The estimates are taken from the fixed-effects regression in Table 6. The valuations are annually figures, and are measured in capita. Hence, a move from meeting friends or relatives less than once or twice a month to meeting friends or relatives on most day is worth around £85,000 a year when fixed personality traits have been controlled for in the regression. An average change in real annual household income per capita (in 1996 Pounds) is around £4,000. At the time of writing, the value of one pound sterling is approximately 1.5 US$.

Figure D.6: *Table 6 from Powdthavee (2008).*

Appendix E

Formal requirements

E.1 Zusammenfassung

Die meisten Umweltgüter und -leistungen werden aufgrund von Marktversagen nicht in effizienter Menge auf Märkten bereitgestellt. Um die gesellschaftlich optimale Menge zu erreichen, sind zwei Arten von Untersuchung nötig: erstens muss die Struktur des Problems verstanden werden, um passende Instrumente zu identifizieren, die das Marktversagen reduzieren oder sogar beheben können. Zweitens muss die Größe der Abweichung des Ist-Zustandes von der optimalen Situation geschätzt werden. Diese Dissertation trägt zu beiden Untersuchungsarten bei.

In den ersten beiden Artikeln werden die Auswirkungen zweier Arten von Unsicherheit auf die Bereitstellung öffentlicher Güter untersucht. Der erste Artikel (mit Andreas Lange) untersucht den Einfluss von Ratifizierungsunsicherheit auf die optimale Ausgestaltung internationaler Umweltabkommen über internationale öffentliche Güter (IUAs). Dazu wird die Literatur über IUAs mit der Public-Choice-Literatur über Two-Level-Games kombiniert. Durch die Modellierung unsicherer Präferenzen des pivotalen Entscheiders im Ratifizierungsprozess, können mehrere wichtige Beiträge zur Literatur geleistet werden: (i) die Modellierung der zeitlichen Struktur von IUAs wird realistischer dargestellt, (ii) wir finden einen überraschenden jedoch intuitiven nicht-monotonen Zusammenhang zwischen optimalen Beiträgen und Varianz der Ratifizierungsentscheidungen, (iii) wir zeigen, dass eine größere Anzahl Teilnehmerländer in den internationalen Verhandlungen zu erwarten ist als in den meisten Koalitionsmodellen mit repräsentativen Agenten. Ratifizierungsunsicherheit kann daher die Wohlfahrtsgewinne aus internationalen Abkommen erhöhen.

Der zweite Artikel (mit Andreas Nicklisch und Andreas Lange) untersucht Unsicherheit, die nicht im Prozess der Bereitstellung sondern im öffentlichen Gut selbst liegt. Beim Klimawandel, zum Beispiel, ist das Ergebnis einer bestimmten Emissionsreduktion mit Unsicherheit behaftet. Eine Emissionsreduktion könnte entweder die Wahrscheinlichkeit eines Sturmes reduzieren, oder sie könnte seine Stärke reduzieren. Dieser Artikel untersucht daher die Dynamik freiwilliger Beiträge, die entweder die Höhe oder die Wahrscheinlichkeit eines probabilistischen Schadens zur Folge haben. Mit Varianten eines wiederholten vier-personen Gefangenendilemma-Spiels, zeigen wir, dass die Kooperation höher und stabiler ist, wenn sie die Wahrscheinlichkeit statt die Höhe des Schadens beeinflusst. Wir finden folgende Verhaltensanpassungen: nicht-kooperative Spieler wechseln mit höherer Wahrscheinlichkeit zu Kooperation, nachdem sie den Schaden erlebt haben, während Kooperation verstärkt wird, wenn kein Schaden eintritt. Dieses Verhalten ist konsistent mit einfachen Lerndynamiken, die auf ex-post Evaluation der gewählten Strategie beruhen.

Der dritte und vierte Artikel gehören zur zweiten Art von Untersuchung, die die Größe des Marktversagens schätzt. Sie leisten einen methodischen Beitrag zum Lebenszufriedenheitsansatz zur Bewertung von Nicht-Marktgütern indem ein neues Maß für Lebenszufriedenheit vorgeschlagen wird. Der dritte Artikel (mit Grischa Perino) präsentiert konzeptionelle Gründe und empirische Befunde dafür, dass die Darstellung von Präferenzrelationen durch Änderungen von Lebenszufriedenheit systematisch verzerrt sind. Als Lösung, schlagen wir das Ranking Measure vor, wo Befragte angeben, ob ihr Leben "besser" oder "schlechter" geworden ist. Anhand dreier repräsentativer Datensätze zeigen wir, dass die beiden Maße öfter widersprüchliche Präferenzrelationen angeben als nicht. Das Ranking Measure wird besser durch sozio-ökonomische Variablen erklärt als Änderungen im Level. Adaptation (z.B. an Einkommen) ist nur bei Levels aber nicht bei Rankings zu beobachten, was darauf hindeutet, dass Adaptation durch die Messmethode statt durch tatsächliche Präferenzen zustande kommt.

Der vierte Artikel (ohne Ko-Autor) wendet das Ranking Measure im Lebenszufriedenheitsansatz zur Bewertung von Nicht-Marktgütern an. Diese Bewertungsmethode wird zunehmend verwendet. Sie misst die Grenzzahlungsbereitschaft (GZB) als Grenzrate der Substitution zwischen Einkommen und dem Nicht-Marktgut, bei konstanter Lebenszufriedenheit. Oft leiden die Ergebnisse unter unrealistisch kleinen Einkommenskoeffizienten und daraus folgend hoher GZB. Dieser Artikel untersucht zwei Quellen für Verzerrungen der Methode: Personen-Fixed-Effects und Adaptation der Antwortfunktion. Fixed Effects können durch Fixed-Effects oder First-Differenced Modelle berücksichtigt werden. Um Änderungen in der Antwortfunktion zu vermeiden, wird das Ranking Measure vorgeschlagen, welches die Notwendigkeit der Anpassung der Antwortskala vermeidet (Köke und Perino 2017). Die Ergebnisse zeigen, dass bei einem Vergleich von pooled OLS, fixed effects und first differences mit dem LM sowie dem RM die GZB auf ein zunehmend realistisches Niveau sinkt. Diese methodische Untersuchung ermöglicht ein besseres Verständnis der Vorzüge aber auch der Grenzen des Bewertungsansatzes.

Die vier Artikel sind also sehr unterschiedlich. Sie können alle in einem umweltökonomischen Kontext interpretiert werden, sind jedoch nicht darauf beschränkt. Die ersten beiden können allgemein auf öffentlichen Güter angewandt werden und der Lebenszufriedenheitsansatz zur Bewertung kann für eine Vielzahl von Nicht-Marktgütern verwendet werden. Die vier Artikel decken auch eine große Bandbreite an Methoden ab: Spieltheorie, Laborexperimente, Paneldatenanalyse. Dies spiegelt die Diversität der (umwelt-) ökonomischen Methoden wider.

E.2 Abstract

Most environmental goods and services are not efficiently provided through markets due to different types of market failure. In order to reach the socially optimal amount, two key lines of investigation are necessary: firstly the structure of the problem needs to be understood in order to find tools that can reduce or eliminate the market failure. Secondly, the magnitude of the discrepancy between the actual and the optimal situation needs to be estimated. This thesis contributes to both these types of investigations. In the first two papers, the effects of two types of uncertainty on public good provision are analysed.

The first paper (joint with Andreas Lange) analyses the impact of ratification uncertainty on the optimal terms of international environmental agreements on international public good provision (IEAs). For this, it combines the literature on IEAs with the public choice literature on two-level games. By also incorporating uncertain preferences of the pivotal voter in the ratification stage, it makes several important contributions: (i) it contributes to a more realistic modelling of the temporal structure of international agreements, (ii) it shows a surprising, yet intuitive non-monotonic relationship between the optimal commitment level and the variance of ratification decisions, (iii) it identifies reasons to expect a larger number of countries to join international negotiations than predicted by most of the coalition formation literature based on a representative agent model. Ratification constraints thereby can improve the welfare gains from stable international agreements.

The second paper (joint with Andreas Nicklisch and Andreas Lange) explores uncertainty structures in the public good itself. In the context of climate change for example, the outcomes when emitting a certain amount of green house gases is not certain. Reducing emissions might reduce the probability of a storm hitting or it might reduce the strength of the storm. This paper thus investigates the dynamics of voluntary cooperation to either reduce the size or the probability of stochastic losses. For variants of a repeated four-person prisoner's dilemma game, it shows that cooperation is larger and more stable when it affects the probability rather than the size of the adverse event. It provides crucial insights on behavioural adaptation: defecting players are more likely to switch to

cooperation after experiencing an adverse event, while existing cooperation is reinforced when the losses do not occur. This behavior is consistent with simple learning dynamics based on ex post evaluations of the chosen strategy.

The third and fourth paper belong to the second type of investigation which estimates the size of the market failure. They make a methodological contribution to the life satisfaction approach to valuation of non-market goods by introducing and using a new measure of life satisfaction. The third paper (joint with Grischa Perino) provides conceptual reasons and empirical evidence that preference rankings elicited from changes in levels of life satisfaction are systematically biased. As a remedy, it proposes the ranking measure where subjects state whether their life has become "better" or "worse". Using three representative datasets it shows that the two measures produce conflicting preference rankings more often than not. The ranking measure is better explained by socio-economic variables than changes in levels. Adaptation (e.g. to income) is only observed for levels but not for rankings, indicating that adaptation is driven by the elicitation method rather than changes in preferences.

The fourth paper (single authored) applies the ranking measure of life satisfaction to valuation of non-market goods in order to test its performance. The life satisfaction approach to valuation is increasingly used. It calculates the marginal willingness to pay (MWTP) as the marginal rate of substitution between income and the non-market good, keeping life satisfaction constant. It often suffers from unrealistically small income coefficients and therefore large MWTP. This paper investigates two possible sources for biases in the method: person-fixed effects and adaptation of the reporting function. Fixed effects are accounted for by fixed effects or first differenced models. To avoid changes in the reporting function, the ranking measure of life satisfaction is suggested, which avoids the need for adaptation of the reporting function (Köke and Perino 2017). The paper finds that when moving from pooled OLS, to fixed effects, to first differences with the LM to the RM, the inferred marginal willingness to pay is reduced to an increasingly realistic level. This methodological examination helps better understand the merits and limitations of the approach.

The four chapters are thus very different. They can all be interpreted in an environmental economics context but are not restricted to that. The first two can more generally be interpreted for all kinds of public good provision and the life satisfaction approach to valuation can be used for a variety of non-market goods. The four chapters cover a variety of methods: game theory, laboratory experiments, panel data analysis. This is reflective of the diversity of approaches used in (environmental) economics.

E.3 List of publications from this thesis

Köke, Sonja and Andreas Lange (2017). Negotiating environmental agreements under ratification constraints, *Journal of Environmental Economics and Management*, 83, 90-106.