ESSAYS IN DEVELOPMENT ECONOMICS

RACHEL CASSIDY



University of Oxford

Department of Economics

Submission for the degree of Doctor of Philosophy (DPhil)

August 2017

Rachel Cassidy: *Essays in development economics,* University of Oxford, © August 2017

In loving memory of Anthony Cassidy,

1948–2005.

ABSTRACT

The thesis comprises three chapters on the microeconomics of development.

The first chapter consists of a lab-in-field experiment in rural Pakistan. I test whether anticipated changes in individuals' liquidity constraints can cause individuals spuriously to appear "present-biased" or "future-biased". The design uses experimental windfalls to create fully exogenous variation in subjects' liquidity constraints. The results suggest that the effect of liquidity constraints on "present-bias" and "future-bias" in monetary tasks exists, is large, and does not operate via changing cognitive function or risk preferences. Importantly, and in contrast with other recent studies, I establish that the causation runs from tighter liquidity constraints to appearing "present-biased" — rather than truly present-biased individuals making choices which lead to tighter liquidity constraints. The findings have important implications for the measurement of time preferences, and for interventions offering commitment.

The second chapter — co-authored with Marije Groot Bruinderink, Wendy Janssens and Karlijn Morsink — describes a field experiment in the slums of Matola, Mozambique. We model and test how intra-household bargaining affects adoption of dual-protection contraceptive technologies, when women place a greater weight than men on the health costs of unprotected sex. To do so, we evaluate an intervention which provides women with information about and free access to male and female condoms. We find strong impacts on adoption of female condoms, using both survey data and data from weekly coital diaries. As predicted by the model, take-up of female condoms is higher among women with low household bargaining power at baseline, including women who are unable to convince their partner to use male condoms.

The third chapter — co-authored with Marcel Fafchamps — reports the findings from a survey of Village Savings and Loan Associations (VSLAs) in Malawi. We ask whether some individuals adopt VSLAs as a commitment savings technology, and whether they sort into groups with borrowers. Using dyadic regression analysis we find evidence of negative assortative matching on a measure of "present-bias", which suggests that this is the case. This may increase financial intermediation, and welfare, in villages with low access to formal banking.

Page count: 198 (excluding bibliography)

ACKNOWLEDGMENTS

I am eternally grateful to Simon Quinn for his supervision and constant support throughout the length of the DPhil. I also thank Marcel Fafchamps and Stefan Dercon for their invaluable guidance in the earlier and later years respectively. For very helpful comments on drafts and presentations of particular chapters, I thank Johannes Abeler, Dan Anderberg, Abi Barr, James Fenske, Glenn Harrison, Climent Quintana-Domeque, and Chris Roth.

None of my fieldwork would have been possible without Rose Page and the CSAE team, to whom I am deeply indebted. For supporting my fieldwork in Pakistan I also thank Tahir Waqar, Ghulam Rasool, Tazeem Khan and our fantastic team of enumerators at NRSP, as well as Mahniya Zafar, Ahmed Ayub, Mehreen Irshad, and Tawfiq Hamid. In Mozambique my co-authors and I thank Lene Boehnke, Emilie Berkhout, Balthazar Chilundo, Pathfinder, Oxfam Novib, and WeConsult. For assisting with the project in Malawi, my co-author and I thank James Mkandawire and all at Invest in Knowledge Initiative, Waluza Munthali, and Smit Gade. We also thank Helene Bie Lilleør, Christopher Ksoll, Jonas Helth Lønberg and Ole Dahl Rasmussen for sharing their panel data, and Sam Asher and Paul Novosad for sharing their "fuzzy matching" algorithm. Finally, I thank all of the respondents in all three countries, who were willing to give their time and share information about their lives in order to further the causes of research.

A DPhil is a long and lonely road, but I have had the privilege of encountering many brilliantly smart and genuinely good-hearted people along the way. Karlijn Morsink has been not only a fantastic co-author but also a deeply supportive friend. I am not sure I would have made it through without Daniela Scur and Kate Orkin, whose steadfast friendship and professional advice has helped me survive even the toughest of times. Damian Clarke, Pavel

Luengas Sierra and Hannah Mülrad have been like a DPhil family to me, providing Stata code and many a lively diversion whenever needed. I am incredibly grateful to Muhammad Meki for his excellent company and thoughtful cultural guidance in Pakistan. More broadly, I been fortunate enough to be part of a great PRS cohort, and a lovely group of development economics DPhil students. I especially thank the members of Firms Group, whose presentation advice and camaraderie have made the economics department a much more welcoming place to be. I also thank the countless other friends whose warmth and shared adventures have kept me going along the way. I thank Davin Lunz for his patient companionship and LaTeXtroubleshooting in the final months. I send love to my sisters, for never quite understanding what I was doing or what on earth a DPhil was, but trusting me that it was going to be alright, our kid. And finally, I thank my mum, for everything she has ever done for me.

DECLARATION OF FINANCIAL SUPPORT

My work as a DPhil student was supported by the Economic and Social Research Council, grant number ES/J500112/1. The project in chapter one was funded by the International Growth Centre (IGC) Pakistan. The IGC aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research. The views expressed here do not necessarily reflect those of IGC Pakistan nor the IGC as a whole. The project in chapter two was funded by the Netherlands Organisation for Scientific Research (NWO) WOTRO Science for Global Development. The project in chapter three was funded by the UK Department for International Development (DFID) as part of the research consortium Improving Institutions for Pro-Poor Growth (iiG), which studies how to improve institutions in Africa and South Asia. The views expressed are not necessarily those of DFID.

CONTENTS

I INTRODUCTION 1

1 INTRODUCTION 2

1.1 Present-bias and technology adoption . . . 2

1.2 Intra-household bargaining and technology adoption . . . 4

1.3 Local institutions as a vehicle for savings and credit technologies . . . 6

II CHAPTERS 9

2 ARE THE POOR REALLY SO PRESENT-BIASED? EXPERIMENTAL EVIDENCE FROM PAKISTAN 10

2.1 Introduction . . . 10

2.1.1 Summary of experiment and results . . . 12

2.1.2 Implications for policy and explaining stylized facts . . . 14

2.2 Theoretical Framework . . . 17

2.2.1 Measuring "present-bias" . . . 17

2.2.2 Consumption and liquidity constraints . . . 19

2.2.3 Experimental choices . . . 20

2.2.4 Predicted effects of income shocks . . . 23

2.3 Experimental Design . . . 25

2.3.1 Surveys and activities . . . 25

2.3.2 Time preference activities . . . 26

2.3.3 Participation fee timing experiment . . . 28

2.3.4 Survey timing experiment . . . 29

2.3.5 Experimental predictions . . . 30

2.4 Data . . . 31

2.4.1 Sample and setting . . . 31
2.4.2 Descriptive statistics and balance . . . 33
2.4.3 Measured "time-inconsistency" . . . 38
2.5 Results . . . 46
2.5.1 Participation fee timing . . . 46
2.5.2 Survey timing . . . 51
2.5.3 Interaction of treatments . . . 55
2.5.4 Heterogeneity by existing liquidity constraints . . . 58
2.6 Robustness . . . 60
2.6.1 Alternative explanations . . . 60
2.6.2 Salience and frame order effects . . . 65
2.6.3 Alternative specifications . . . 66
2.7 Conclusion . . . 68
2.8 Appendices . . . 70
2.8.1 Proof of predicted effects of windfalls . . . 70
2.8.2 Experimental materials . . . 72
2.8.3 Additional results . . . 75
2.8.4 Additional robustness checks . . . 79
3 THE POWER TO PROTECT: HOUSEHOLD BARGAINING AND FEMALE CONDOM USE 81
3.1 Introduction . . . 81
3.2 Context and intervention . . . 87
3.2.1 HIV and condom use in Maputo . . . 87
3.2.2 Female condom intervention . . . 89
3.3 Model . . . 90
3.3.1 Preferences . . . 91
3.3.2 Optimal contraceptive choice and bargaining power . . . 92
3.3.3 Choice between male condoms and unprotected sex . . . 94
3.3.4 Introduction of female condoms . . . 95
3.3.5 Model predictions . . . 97
3.4 Study design and data sources . . . 98

3.4.1 Study design . . . 98
3.4.2 Survey data . . . 101
3.4.3 Diary data . . . 105
3.4.4 Bargaining power . . . 107
3.5 Results . . . 109
3.5.1 Main impact results . . . 109
3.5.2 Diary results . . . 115
3.5.3 Heterogeneity by bargaining power . . . 120
3.5.4 Heterogeneity by HIV status and risk perceptions . . . 124
3.5.5 Effects on knowledge, STIs and well-being . . . 130
3.5.6 Robustness . . . 131
3.6 Conclusion . . . 133
3.7 Appendices . . . 135
3.7.1 Proofs . . . 135
3.7.2 Additional descriptive data . . . 136
3.7.3 Additional analyses . . . 141
4 CAN SELF-FUNDED MICROFINANCE GROUPS MATCH SAVERS WITH BORROWERS? EVIDENCE FROM MALAWI 151
4.1 Introduction . . . 151
4.1.1 Contribution to existing literature . . . 155
4.2 Conceptual framework . . . 157
4.2.1 Design of VSLAs . . . 157
4.2.2 Sorting across VSLAs . . . 159
4.3 Survey and Data . . . 162
4.3.1 Description of VSLA survey . . . 162
4.3.2 Matching to household survey data . . . 164
4.3.3 Group-level data on saving and borrowing . . . 165
4.3.4 Individual member characteristics . . . 166
4.3.5 Group composition by member characteristics . . . 173
4.3.6 Dyad characteristics . . . 175

4.4 Empirical strategy and results . . . 175
4.4.1 Dyadic regression framework . . . 175
4.4.2 Sorting on occupation . . . 179
4.4.3 Sorting on time preferences . . . 181
4.5 Robustness . . . 184
4.6 Conclusion . . . 189
4.7 Appendices . . . 191

III CONCLUSION 196

5 CONCLUSION 197

BIBLIOGRAPHY 199

LIST OF FIGURES

Figure 1 Income in the last two & next two weeks, by baseline survey date 37
Figure 2 Effects of windfall timing on day one switch-points 40
Figure 3 Effects of survey timing on day one switch-points 41
Figure 4 Effects of windfall timing on day one static inconsistency 48
Figure 5 Effects of survey timing on day one static inconsistency 52
Figure 6 Effects of windfall timing, by survey timing 56
Figure 7 Structure of experimental sessions 72
Figure 8 Multiple price list demonstrations 73
Figure 9 Payment vouchers 74
Figure 10 Dual-protection technology frontier 93
Figure 11 Location of Karonga District within Malawi 163

LIST OF TABLES

Table 1 Balance – participation fee timing . . . 34
Table 2 Descriptive statistics by survey timing . . . 36
Table 3 Static inconsistency, day one & day fifteen . . . 39
Table 4 Static inconsistency on day one & revisions on day fifteen . . . 43
Table 5 Static inconsistency on day one & demand for commitment . . . 45
Table 6 Revisions on day fifteen & demand for commitment . . . 45
Table 7 Treatment effects on day one static inconsistency – windfall timing . . . 49
Table 8 Treatment effects on day one static inconsistency – survey timing . . . 53
Table 9 Treatment effects – interaction of windfall & survey timing . . . 57
Table 10 Treatment effects of windfall timing – heterogeneity . . . 59
Table 11 Treatment effects of windfall timing – trust controls . . . 62
Table 12 Estimated treatment effects – activity & frame order . . . 67
Table 13 Treatment effects on switch-points – windfall timing . . . 75
Table 14 Treatment effects on day fifteen inconsistencies – windfall timing . . . 76
Table 15 Treatment effects on switch-points – survey timing . . . 77
Table 16 Treatment effects on day fifteen inconsistencies – survey timing . . . 78
Table 17 Treatment effects of windfall timing – additional controls . . . 79
Table 18 Treatment effects of survey timing – additional controls . . . 80
Table 19 Predictors of attrition – treatment & control . . . 100
Table 20 Baseline balance of covariates & use – full sample . . . 102
Table 21 Comparison of sample to DHS representative sample . . . 104
Table 22 Bargaining power – summary statistics . . . 108
Table 23 Assets brought to the relationship – principal component analysis . . . 109
Table 24 Bargaining power – principal component analysis . . . 110

Table 25 Treatment effects – primary outcome variables . . . 112
Table 26 Treatment effects on female condom use, by baseline male condom use 114
Table 27 Impacts on proportion of all sex acts protected – diary subsample . . . 116
Table 28 Impacts on number of sex acts per respondent per week – diary subsample . . . 118
Table 29 Impacts on % sex acts protected per respondent per week – diary subsample . . . 119
Table 30 Impacts on % sex acts with discussions per respondent per week – diary subsample . . . 121
Table 31 Current use of female condoms at endline, by baseline bargaining power 123
Table 32 Current use of female condoms at endline – women using male condoms at baseline . . . 125
Table 33 Current use of female condoms at endline – women not using male condoms at baseline . . . 126
Table 34 Treatment effects, by baseline HIV status . . . 127
Table 35 Treatment effects – heterogeneity by baseline HIV risk beliefs . . . 129
Table 36 Treatment effects – other outcome variables . . . 131
Table 37 Baseline balance excluding attritters . . . 136
Table 38 Baseline balance on covariates – diary subsample . . . 137
Table 39 Baseline balance on use – diary subsample . . . 138
Table 40 Diary sample representativeness of full sample – covariates . . . 139
Table 41 Diary sample representativeness of full sample – baseline use . . . 140
Table 42 Balance – principal components of bargaining power . . . 140
Table 43 Treatment effects on male condom use, by baseline male condom use . 141
Table 44 Treatment effects on female condom use, by baseline contraceptive use . 142
Table 45 Treatment effects – heterogeneity by relationship status . . . 143
Table 46 Impacts on current use of female condoms – heterogeneity by bargaining power . . . 144
Table 47 Impacts on male condom use – heterogeneity by baseline HIV status . . 145

Table 48 Treatment effects – heterogeneity by baseline bargaining power & risk beliefs . . . 146
Table 49 Lee bounds – primary outcome variables . . . 147
Table 50 Treatment effects – survey variables, diary subsample . . . 148
Table 51 Impacts on condom use – fixed effects panel estimator, diary subsample 149
Table 52 Impacts on condom use – ANCOVA specification, diary subsample . . . 150
Table 53 Individual member characteristics – 2013 member census . . . 167
Table 54 Individual economic activity & schooling – 2013 member census . . . 168
Table 55 Individual member characteristics – matched subsample . . . 172
Table 56 VSLA groups per village . . . 173
Table 57 Group composition by member characteristics . . . 174
Table 58 Dyad characteristics – 2013 member census . . . 176
Table 59 Dyad characteristics – matched subsample . . . 177
Table 60 Dyadic regressions – 2013 member census . . . 180
Table 61 Dyadic regressions – matched subsample . . . 183
Table 62 Dyadic regressions – time preference measures, matched subsample . . . 185
Table 63 Dyadic regressions – effects of controls, matched subsample . . . 187
Table 64 Test of representativeness – matched dyads vs. 2013 census dyads . . . 191
Table 65 Dyadic regressions – 2013 member census and matched subsample . . . 192
Table 66 Dyadic regressions – 2013 member census, weighted . . . 193
Table 67 Dyadic regressions – time preference measures, matched subsample, weighted . . . 194
Table 68 Dyadic regressions – effects of controls, matched subsample, weighted . 195

Part I

INTRODUCTION

1

INTRODUCTION

This thesis comprises three chapters on the microeconomics of development. Each chapter seeks to answer a different research question, and reports the findings of a separate fieldwork project. Nonetheless, a theme which runs through all three chapters is that of understanding the preferences that drive demand for new technologies in developing countries. The chapters study this theme from the perspective of individual preferences alone, the mediation of these preferences through intra-household bargaining, and their interaction with local informal institutions respectively. To do so, they use a combination of lab-in-field experiments, field experiments, and collection of survey data.

1.1 PRESENT-BIAS AND TECHNOLOGY ADOPTION

The first chapter focuses on present-bias in individuals' time preferences, in the context of the demand for commitment savings technologies. Recent literature has suggested that the poor in particular may have a disproportionate desire to consume now rather than later, for example due to reduced bandwidth (Mani et al., 2013) or declining temptations (Banerjee and Mullainathan, 2010). Present-bias does appear to be a contributing factor to some of the cases of technology under-adoption observed in the developing world, for example that of fertilizer (Duflo et al., 2011). Models of present-bias may also be a good characterisation of sub-optimal individual behaviours in domains such as effort choices (Kaur et al., 2015).

However, it is not obvious that decision-making by the poor will generally be more myopic than that by wealthier individuals. This is because the mistakes implied by allowing present-biased preferences to dictate choices are much costlier for those living close to subsistence constraints. Thus in many cases the poor may be more likely to exercise costly self-control, or to use informal commitment devices. Furthermore, in developing countries in particular, many decisions about consumption and technology adoption are made at the household level. Thus the extent to which an individual's present-bias actually matters for choices and welfare may vary widely, depending on household budgeting arrangements.[1]

Moreover, even the leading proponents of present-bias theory have emphasised that present-bias should not be relied upon as the best or the only candidate explanation of many phenomena (O'Donoghue and Rabin, 2015). In many cases of under-investment or technology under-adoption, equally plausible alternative explanations to present-bias exist. As the first chapter of this thesis makes clear, individuals may appear "present-biased" in their consumption or financial choices because the utility function is concave and they are anticipating an improving consumption path, reducing their marginal rate of inter-temporal substitution over time (Epper, 2015). Similar comments apply when the cost of investment is one of effort: people may defer a task that involves effort if they anticipate that their time constraints will ease in the future. In some cases, failure to adopt beneficial technologies may also arise from lack of attention (Haushofer, 2014). All of this may be perfectly rational; although optimism may play a role if expectations are biased, and irrational inattention may occur.

Little evidence exists to distinguish these explanations from present-bias, except for attempts to measure present-bias via monetary or effort tasks and then to correlate them with observed behaviour. Some attempts yield the predicted correlation (Ashraf et al., 2006), while others do not despite a very similar setting (John, 2017). Yet whilst subjects' choices in incentivized versions of such tasks may produce a very close approximation to their real behaviour, the

1 Conversely, perfectly time-consistent individuals with different discount rates may bargain in a way that leads to "present-biased" decision-making at the household level (Adams et al., 2014); but in that case the story is one of bargaining, not of present-biased preferences.

downside is that almost all of the alternative explanations for "present-biased" behaviour may also explain away "present-biased" choices in such tasks.[2]

The first chapter describes a lab-in-field experiment in Pakistan, which tests for causal evidence that poor individuals who are expecting their liquidity constraints to improve may spuriously appear "present-biased". The experiment gives participants experimentally-timed liquidity shocks, and also exploits quasi-experimental shifts in liquidity around the harvest. I find that the proportion of individuals making "present-biased" and indeed "future-biased" choices is highly responsive to these exogenous changes in liquidity constraints. Thus much of the poor's apparently "present-biased" financial behaviour may in fact reflect rational expectations that their liquidity will increase in the future. The findings also raise questions as to whether other apparently "present-biased" behaviour is in fact due to rational or optimistic expectations, not only about liquidity constraints, but also about time constraints or risk. Important examples may be failures to invest in preventative health measures, or decisions to engage in risky sexual behaviour.

1.2 INTRA-HOUSEHOLD BARGAINING AND TECHNOLOGY ADOPTION

The second chapter turns its focus to risky sexual behaviour, but this time considers how individual preferences are mediated through intra-household bargaining. Women ostensibly have stronger preferences than men for risk reduction (Agnew et al., 2008), and for investing in household public goods such as children (Duflo, 2003). Economic literature has consistently rejected the unitary model of intra-household bargaining (Udry, 1996; Fortin and Lacroix, 1997; Browning and Chiappori, 1998), finding that when women gain greater control over resources, household choices better reflect their preferences (Attanasio and Lechene, 2002; Duflo and Udry, 2004; Attanasio and Lechene, 2014). Practitioners have recognised the implications of this for the design of interventions such as microfinance and conditional cash

2 "Temptation" models of present-bias may also explain behaviour better than the "beta-delta" model in some circumstances, and may not be captured by existing measures of "present-bias" (Gul and Pesendorfer, 2001; Banerjee and Mullainathan, 2010; Toussaert, 2015).

transfer programmes: namely, if the policy objective is to increase household spending on these categories, then it may be better to place funds in the hands of women rather than men.

In the domain of sexual and reproductive health, very similar arguments apply. Women are likely to have a stronger preference for safe sex, given that they are more risk-averse and disproportionately bear the cost of HIV/AIDS and child-rearing.[3] Previous studies have shown that young women are responsive to HIV risk information (Dupas, 2011), and that increasing young women's bargaining power increases their ability to avoid unwanted sex and pregnancy (Bandiera et al., 2015). However, the only widely-promoted technology that protects against HIV/AIDS and other STIs is still the male condom — a technology marketed to men, sold in places more accessible to men (at least in more conservative societies), and associated by many men with discomfort and stigma. Coupled with the fact that men tend to hold greater bargaining power than women, this may help explain the chronic under-adoption of male condoms: in 2015, an estimated 3.3 billion risky sex acts took place without condoms in Sub-Saharan Africa alone, leading to 910,000 new HIV infections (UNAIDS, 2016b).

Female condoms have long been seen as a more expensive alternative to male condoms, and as potentially more difficult to introduce because of lower familiarity with the technology, and taboo or ignorance about female anatomy. Yet new, lower-cost versions are available (Dowdy et al., 2006), and have been successfully introduced with fairly minimal information programmes (Mantell et al., 2015). There is also evidence that female condoms provide lower displeasure and stigma to men than male condoms (Koster et al., 2015). Female condoms may therefore have important potential for boosting condom adoption.

The second chapter — co-authored with Marije Groot Bruinderink, Wendy Janssens and Karlijn Morsink — seeks to test this idea via a field experiment in Mozambique. We evaluate an intervention that provides women with information about contraception and free access to both male and female condoms. We find strong take-up of female condoms, which are other-

3 In 2015, women represented 59% of HIV-positive individuals in Sub-Saharan Africa, and the rate of new infections among young women aged 16-25 was double that among young men (UNAIDS, 2016b). This reflects the tendency of women to have older partners, women's lower access to sexual and reproductive health services, and women's higher biological risk of becoming infected (UNAIDS, 2016a).

wise difficult to obtain in the study area. Crucially, take-up is particularly high among women with low baseline bargaining power, including women who are not using male condoms at baseline. In terms of implications for policies to combat HIV/AIDS and other STIs, our results suggest that female condoms may be the only viable dual protection method for some women with low bargaining power. Thus for policymakers considering whether to subsidise distribution of female condoms, the correct cost comparison in contexts where women have low bargaining power is not to male condoms, but rather to antiretroviral therapies and other costs associated with unprotected sex. The results also highlight how intra-household bargaining should be taken into account more broadly when designing and marketing technologies to improve household welfare. Possible policies include information and social norm campaigns targeted specifically at men, providing versions of technologies that are more acceptable to men (as in the case of female condoms), or bundling technologies with goods for which men have a high demand.

1.3 LOCAL INSTITUTIONS AS A VEHICLE FOR SAVINGS AND CREDIT TECHNOLOGIES

The third chapter studies the interaction of individual preferences with the introduction of a new type of informal financial institution — Village Savings and Loan Associations (VSLAs). Two billion people in the world still live in households that are unbanked (Demirgüç-Kunt et al., 2015). Until recently, microcredit was widely touted as the solution, and the joint-liability lending model as introduced by Grameen Bank has been replicated all over the developing world. There has been some shift towards individual-liability lending especially in urban areas,[4] but the product has remained one of credit with frequent repayment,[5] and the role of the group has remained key in rural settings (Feigenberg et al., 2013). However, there is now widespread evidence that such lending models produce little or no impact on consumption,

4 By 2012, 46% of lending by major microfinance institutions was individual-liability (Quidt et al., 2016). Grameen Bank itself no longer uses joint-liability contracts.

5 Microfinance lenders such as Grameen now ask borrowers to contribute a small amount of savings. However, given that the amount is a tiny fraction of the loan size, this is likely about gaining psychological buy-in rather than a meaningful deposit.

investment or a host of welfare indicators (Banerjee et al., 2015). In contrast, several studies have found positive intent-to-treat effects from offering formal savings products (Ashraf et al., 2006; Dupas and Robinson, 2013a). Yet finding a financially sustainable way for banks to offer savings accounts to the poor appears to be elusive.

One possible solution could be to bundle savings and credit products at the level of a microfinance group, or to create peer-to-peer lending systems such that the cost of providing a savings technology is borne by borrowers. ROSCAs can be seen as a basic and ubiquitous version of this idea. Moreover, adding additional savings components onto ROSCAs has been shown experimentally to be very successful (Dupas and Robinson, 2013b). More recently, NGOs have come to champion the VSLA or "Self-Help Group" model. This is essentially a more flexible form of a ROSCA, in which members can decide individually and week-by-week how much to contribute over and above a minimum amount, and if and when they wish to ask the group for a loan. Unlike ROSCAs, VSLAs can also yield interest on savings, since borrowers repay with interest which is then shared out among members at the end of the year.

Recent studies have shown positive impacts of VSLAs on food security (Beaman et al., 2014; Ksoll et al., 2016). Yet what remains unknown is the extent to which these groups really are able to attract savers, and to promote financial intermediation between savers and borrowers. Instead, it may be the case that these groups overwhelmingly attract borrowers, and thus act more like regular microfinance groups. Indeed, the existing literature on VSLAs has analysed them with a focus almost exclusively on credit (Greaney et al., 2016; Burlando et al., 2017).

Understanding how these groups function in terms of savings and financial intermediation is thus the focus of the third chapter of this thesis, co-authored with Marcel Fafchamps. We highlight how VSLAs have the features not only of a borrowing but also of a commitment savings technology. We therefore ask whether VSLAs are adopted both by present-biased individuals with a demand for commitment savings and by individuals with a demand for credit, and if such individuals sort into groups with one another. On the other hand, such intermediation may be impeded by limited liability, or broader social frictions. The results

show evidence that commitment savers and borrowers do sort into groups with one another, which may increase welfare for both of these types of individuals. However, there are unrealised intermediation possibilities between farmers and non-farmers who do not group together. The latter implies that VSLAs may be heavily exposed to covariate shocks such as a bad harvest, and may benefit from being linked into larger credit systems or government funds, as has been the case with Self-Help Groups in India.

Part II

CHAPTERS

2

ARE THE POOR REALLY SO PRESENT-BIASED? EXPERIMENTAL EVIDENCE FROM PAKISTAN

2.1 INTRODUCTION

In both developed and developing countries, questions over time-dated monetary payments have long been the most prevalent method for eliciting discount rates and identifying time-inconsistent preferences (Thaler, 1991; Tanaka et al., 2010; Halevy, 2015). The simplest examples are tasks in which a respondent is asked to choose between receiving $X today or an increasing amount in the future, with the trade-off between waiting and obtaining a larger amount taken to reveal her discount rate.[1] In wealthier settings, such tasks have faced the critique that subjects may arbitrage experimental payments, although the evidence on arbitrage is mixed (Coller and Williams, 1999; Harrison et al., 2002; Cubitt and Read, 2007; Andreoni et al., 2017).[2] In light of this, recent lab experiments have moved towards asking subjects to trade off exerting a short amount of effort now or a longer amount of effort later (Augenblick et al., 2015; Augenblick and Rabin, 2015). In contrast, in poorer settings subjects are often severely credit-constrained and hold little to no liquid savings; thus their opportunities to arbitrage experimental payments are usually much more limited. Monetary tasks have therefore continued to dominate, in particular due to their portability for large-scale household

1 See Section 2.3.2 for a discussion of variations on this task and estimation of discount rates from such data.

2 Andreoni et al. design tests of arbitrage and find no direct evidence; although they interpret the fact that they see no present-bias in their sample as indirect evidence. Similarly, Augenblick et al. find that Berkeley students do not appear present-biased in monetary tasks but do appear so in effort tasks, suggesting that they may arbitrage monetary tasks. In contrast, Halevy still finds substantial evidence of present-bias in monetary tasks among students at UBC, suggesting they do not arbitrage. Harrison et al. find modest effects of controlling for the interest rates that Danish subjects face on saving and borrowing. Coller and Williams show that the extent to which subjects arbitrage is sensitive to priming about interest rates.

surveys and field experiments (Ashraf et al., 2006; Tarozzi and Mahajan, 2011; Giné et al., 2016; Janssens et al., 2017).

Yet recent literature has highlighted a different potential problem with eliciting discount rates in poor contexts. The issue arises because the poor's income and expenditures are often highly variable, seasonal, and subject to shocks. If subjects do not make experimental choices under narrow framing, but instead partially or fully consider changes in their background consumption when making experimental decisions, then these fluctuations may bias measures of discount rates and time-inconsistency (Noor, 2009; Dean and Sautmann, 2014; Epper, 2015; Gerber and Rohde, 2015; Carvalho et al., 2016). In particular, if a subject expects to have higher consumption in the long run, and cannot smooth this into the present because she is credit constrained, then she may have a higher marginal rate of intertemporal substitution in the short run than in the long run. She may therefore appear more impatient over the short run than the long run, and researchers may mistakenly attribute this to present-biased preferences.[3] The true extent of present-bias amongst the poor may therefore be lower than estimates suggest, if poor subjects are on average expecting to have higher consumption in the future. This is likely to be the case particularly if experiments are conducted during the lean season, or if individuals hold optimistic expectations about their future consumption,[4] or indeed if poor subjects are more credit-constrained than savings-constrained.

However, there is still a lack of causal evidence as to whether changes in background consumption really do make poor individuals appear "present-biased", and whether this effect is large enough to create an important bias in measures of time preferences. Recent studies have provided compelling suggestive evidence. Balakrishnan et al. (2015) show that subjects in Kenya do appear partially to integrate their background expenditure into decisions over time-dated payments in the lab. In a field setting, Dean and Sautmann (2014) find that individuals' responses to such tasks are correlated with household savings and expenditures in a panel dataset in Mali. Moreover, Carvalho et al. (2016) show that poor individuals in the

3 Thus whilst Augenblick et al. show that *not* finding "present-bias" over monetary payments is insufficient to conclude that a subject or sample is *not* truly present-biased; this argument implies that *finding* "present-bias" over monetary payments is insufficient to conclude that a subject or sample *is* truly present-biased.

4 Browning and Tobacman (2007) discuss more general equivalence — and thus lack of separate identification — between common discounting functions and optimistic expectations.

US who take part in a phone interview randomly before rather than after pay-day exhibit higher rates of "present-bias" over money. The authors find no evidence of reduced cognitive function or increased risk-taking before pay-day, and thus the increase in "present-biased" choices seems to be a direct result of being more liquidity-constrained now compared to the near future. Meanwhile Ambrus et al. (2015) find that Icelandic subjects whose income remained the same one year after a lab experiment — and indeed those who had correctly anticipated no income growth — had not exhibited "present-bias" in the lab.[5]

Yet despite their compelling results, these studies cannot rule out that expenditure shocks or anticipated income changes may be endogenous to whether subjects truly have present-biased preferences. For example, Dean and Sautmann cannot rule out that subjects' ability to build up precautionary savings, or to invest in risk-mitigating technologies or insurance, may be endogenous to true present-bias. Similarly, if Carvalho et al.'s sample contains truly present-biased individuals, then they are the most likely to have failed to smooth their pay-cheque to the end of the month. These individuals should therefore experience the largest jump in liquidity from before to after pay-day, and thus the strongest intensity of the survey timing treatment. The finding that individuals appear more "present-biased" just before pay-day therefore cannot rule out that individuals truly are present-biased, and may in fact be caused by this. Finally Ambrus et al.'s results could be explained by truly present-biased individuals also having more optimistic expectations about their future income. This is not implausible, especially since naïve present-bias could be interpreted as another form of optimism: namely, an optimistic belief that one's future self will be more patient than one's current self.

2.1.1 *Summary of experiment and results*

This paper therefore seeks to provide causal evidence that expecting their liquidity constraints to ease in the future can cause poor individuals to appear "present-biased", and that this

5 Balakrishnan et al. find that partial asset integration does not dramatically affect their estimates of present-bias. In contrast, Ambrus et al. find that correcting for subjects' anticipated income growth removes all estimates of "present-bias" from their data.

effect is large enough to cause concern for experimental findings. To my knowledge, I offer the first experimental test of this question, by generating completely exogenous variation in subjects' liquidity constraints. The results therefore offer experimental confirmation of the quasi-experimental relationship observed by Carvalho et al. Importantly, I establish that the causation runs from tighter liquidity constraints to appearing "present-biased" — rather than truly present-biased individuals making choices which lead to tighter liquidity constraints, especially after shocks or at the end of the month.

The experimental design is as follows. Just over 500 subjects each participate in two sessions of incentivized activities: a baseline, and a follow-up two weeks later.[6] Subjects receive a fee for taking part. The experiment randomises at the individual level whether a subject receives this windfall during the first or the second session. Subjects who receive the participation fee during the first session thereby experience an immediate and unanticipated positive windfall, easing whatever liquidity constraints they currently face. Meanwhile, subjects who are told during the first session that they will receive the participation fee during the second session thereby experience an anticipated positive windfall dated two weeks into the future. Given that subjects are highly credit-constrained and hold low savings, this exacerbates any liquidity constraints they currently face, as they are unable to smooth this future consumption into the present.

The results show that even this relatively modest, experimentally-induced change in liquidity constraints can indeed affect measures of "time-inconsistency". Specifically, receiving the windfall during the first session makes subjects 7.1 percentage points less likely to appear "present-biased", and 8.0 percentage points more likely to appear "future-biased". This is compared to 29.1% of subjects appearing "present-biased" and 32.1% appearing "future-biased" when they are told during the first session that they will receive the participation fee during the second session. This temporary experimental treatment by definition cannot alter subjects' "deep" underlying time preferences. Nor does the treatment's effect on time choices appear to operate via effects on subjects' cognitive functioning, risk preferences, opti-

6 Subjects are female clients of the NGO National Rural Support Programme, which is considering offering commitment savings products. See Section 2.4 for details.

mism or trust, detailed measures of which are also included as robustness checks. The results therefore provide experimental evidence that an assumption of "narrow framing" when subjects respond to time-preference questions appears to be violated. Instead, standard monetary measures of time preference also capture subjects' liquidity constraints.

To test that the results also hold for non-experimental sources of liquidity, I also randomly schedule each village's baseline sessions either before or after the main onset of the wheat harvest. This generates quasi-experimental variation in subjects' natural liquidity constraints, in a similar spirit to Carvalho et al. Specifically, subjects who receive their first session prior to the main onset of the harvest are interviewed in a period of tighter liquidity constraints, with the expectation that their liquidity constraints will ease in the near future once harvesting has begun. They may also face greater uncertainty about harvest yields. Results show that just 31.0% of subjects appear "time-consistent" in baseline sessions held before the harvest, whilst the level is 14.6 percentage points higher among subjects whose baseline session takes place after the main onset of the harvest. This is in line with Carvalho et al.'s finding that subjects appear much more "time-consistent" over money after pay-day. However, the effect is mainly driven by levels of "future-bias" being significantly higher prior to the harvest.[7] A possible explanation is that harvest yields carry much more uncertainty than pay-cheques, and thus subjects interviewed prior to the harvest may use the experiment as a way to make precautionary savings.

2.1.2 *Implications for policy and explaining stylized facts*

The belief that many poor individuals are present-biased has led to a proliferation of development interventions with commitment features, most notably savings products (Ashraf et al., 2006; Dupas and Robinson, 2013b; Dupas and Robinson, 2013a; John, 2017). Yet if measurements of "present-bias" are in part spuriously driven by income changes, then the

7 Levels of "present-bias" also appear somewhat higher prior to the harvest, but this is not significant. Carvalho et al. do not report the levels of "future-bias" in their study.

potential benefits of commitment may actually be smaller and less widespread than first thought. Moreover, such benefits may not outweigh the important downside of "hard" commitment products: namely, that tying up poor individuals' liquidity reduces their capacity to self-insure against shocks.[8] The low use and re-adoption rates that have often been observed, after high initial take-up, may thus in part reflect rational learning rather than naïveté.[9] Relatedly, monetary measures of time preference have been shown to be uncorrelated with take-up of commitment in some settings (Tarozzi and Mahajan, 2011) and correlated but far from perfectly in others (Ashraf et al., 2006). On one hand, this may be because individuals who appear present-biased in monetary tasks are naïve. On the other hand, if a significant number of those who appear "present-biased" do so spuriously, they may be rationally uninterested in taking up commitment.

Causal evidence that changes in subjects' liquidity can lead them to appear "time-inconsistent" could also help explain two consistent puzzles in the literature on time preferences. The first arises from studies which compare both "static" inconsistency in choices across different time horizons, as described above, and "dynamic" revision of prior choices at a later date (Halevy, 2015; Giné et al., 2016; Janssens et al., 2017). Many individuals are found to exhibit "time-inconsistency" in one sense but not the other, and the correlation between types of inconsistency is often limited.[10] Standard models of time-inconsistency predict that if an individual's time preference parameters are what drive her to exhibit a static inconsistency in one direction, then she must later exhibit a dynamic inconsistency in the same direction (Phelps and Pollak, 1968; Elster, 1979; Laibson, 1997). A model in which choices over time-dated monetary payments capture time preferences alone thus cannot straightforwardly explain

8 The trade-off between commitment and flexibility may of course also lead individuals who are truly present-biased not to take up commitment (Laibson, 2015).

9 Ashraf et al. see 28.4% take-up for their commitment saver products in the Philippines; but only 50% of those who opened an account made a deposit beyond the opening balance, and only 10% of those who set a savings target met that goal within a year. John sees 25% take-up of a very similar product, but 55% of those who take up completely default on the savings instalment plan that they commit to, despite having designed their own plan. Dupas and Robinson (2013b) offer zero-interest accounts with withdrawal fees to micro-enterpreneurs in Kenya and observe 87% take-up. However, only 47% of those offered the accounts actually made a single deposit within first six months.

10 Giné et al. do find that static "present-biased" choices are correlated with dynamic revisions in the direction of present-bias, suggesting that some individuals in their Malawian population may be truly present-biased and their choices may reflect this. The authors find no such correlation between static "future-bias" and revisions in the direction of "future-bias", thus they suggest that measurements "future-bias" may be more susceptible to changes in background consumption.

the many cases in which individuals appear "present-biased" in one set of choices but not another, at least not without invoking additional assumptions about changing levels of sophistication over time. In contrast, as Janssens et al. point out, a model of fluctuating background liquidity constraints can straightforwardly explain "inconsistent" choices at one point in time but not another. Specifically, individuals who expect to be less liquidity-constrained in the future, and thus who spuriously appear "present-biased" in the static sense, may be no more or less likely to experience a negative liquidity shock prior to revising their choices, and thus spuriously to appear "present-biased" in the dynamic sense.

The second puzzle is that almost all studies measuring time preference in developing countries find a significant proportion of individuals to be "future-biased". This holds both in static choices, i.e. subjects exhibit a *lower* discount rate in the near frame than in the far frame,[11] and in dynamic choices, i.e. subjects revise their far-frame choice in a more *patient* direction if they are asked to choose again closer to the date of the far frame.[12] Such behaviour is difficult to interpret as a widespread preference for delayed gratification.[13] In contrast, a model in which subjects take into account changes in their liquidity constraints straightforwardly predicts systematic observation of "future-biased" choices even if few individuals truly have future-biased preferences. Specifically, a subject may make static "future-biased" choices if she expects to be *more* liquidity-constrained in the far frame. Meanwhile she may make dynamic "future-biased" choices if she experiences an unanticipated *positive* liquidity shock after making her initial choices.[14]

11 Ashraf et al. measure 26% of their sample as "present-biased" and 14.6% as "future-biased". Dupas and Robinson (2013b) find 23% of the women in their sample to be "present-biased" and 18% to be "future-biased"; whilst among men 27% are "present-biased" and 3% are "future-biased". Meanwhile Dupas and Robinson (2013a) measure 16% of their sample to be "present-biased" and 18% to be "future-biased". John finds 16.6% of her sample to be "present-biased" and 18.9% to be "future-biased".

12 Giné et al. find that 34% of subjects revise in a "present-biased" direction but 31% revise in a "future-biased" direction. Janssens et al. find that 32% revise in a "present-biased" direction whereas 11% revise in a "future-biased" direction.

13 Another possible explanation is that "future-bias" is evidence of sophisticated present-biased individuals using the experiment as a commitment device. However, there is no reason why sophisticates should wish to defer payment more in the near frame than in the far frame, as in theory sophisticates should seek to limit the liquidity held by their future self (Harris and Laibson, 2001).

14 Consistent with this, Giné et al. show that "future-biased" revisions are correlated with households experiencing a (non-experimental) positive income shock.

The rest of this paper proceeds as follows. Section 2.2 outlines the theoretical framework and derives the hypotheses that the experiment is designed to test. Section 2.3 describes the experimental design. Section 2.4 outlines the sampling process, and presents descriptive statistics and checks of balance. Section 2.5 presents the estimating equations and results. Section 2.6 details a series of robustness checks. Section 2.7 concludes.

2.2 THEORETICAL FRAMEWORK

This section outlines a simple theoretical framework, to illustrate how "time-inconsistent" choices in standard monetary tasks are insufficient to identify time-inconsistent preferences. Specifically, I highlight how decreases in an individual's expected marginal rate of substitution over time may cause her spuriously to appear "present-biased". In doing so, I draw together insights from both Epper (2015) and Dean and Sautmann (2014). In particular, like Epper I adopt a buffer stock model of savings with partial asset integration. However, in contrast to Epper and following Dean and Sautmann, I allow for the possibility that the subject saves future experimental payments, and that she is truly present-biased ($\beta < 1$). Unlike Dean and Sautmann, I do not assume that the agent is perfectly sophisticated.

2.2.1 *Measuring "present-bias"*

The canonical, static way to measure present-bias is to interview a subject once and to ask for her choices over a near frame and a far frame. These choices can be used to generate the following measures:[15]

15 Activities such as multiple price lists and Becker-DeGroot-Marschak auctions generate such measures directly. More complex tasks such as convex time budgets, if understood by subjects, can be used to generate such measures, and also to infer information about the curvature of the utility function. See section 2.3.2 for a discussion. The intuition of the argument presented here still holds for such tasks.

- **The near-frame switch-point $x_{1.0}$:** at $t = 0$, the amount of money $x_{1.0}$ that makes the subject indifferent between receiving a given, fixed sum x now at $t = 0$ or $x_{1.0}$ at $t = 1$.

- **The far-frame switch-point $x_{2.0}$:** at $t = 0$, the amount of money $x_{2.0}$ that makes the subject indifferent between receiving x at $t = 1$, or receiving $x_{2.0}$ at $t = 2$.

If a subject's choices imply $x_{1.0} \neq x_{2.0}$, this is taken to violate "stationarity" (Halevy, 2015). In particular, if $x_{1.0} > x_{2.0}$ — i.e. her near-frame switch-point is greater than her far-frame switch-point at $t = 0$ — she is conventionally assumed to have a higher discount rate in the near frame than in the far frame, and is therefore labelled as "present-biased".

Recent "dynamic" attempts to measure time-inconsistency comprise both an initial session at $t = 0$ and a follow-up session at $t = 1$. This adds decisions over two further time frames — a near frame at $t = 1$ and a far frame at $t = 1$ — and thereby generates the following additional measures:

- **The near-frame switch-point $x_{2.1}$:** at $t = 1$, the amount of money $x_{2.1}$ that makes the subject indifferent between receiving x at $t = 1$ (which is now the present) or receiving $x_{2.1}$ at $t = 2$.

- **The far-frame switch-point $x_{3.1}$:** at $t = 1$, the amount of money $x_{3.1}$ that makes the subject indifferent between receiving x at $t = 2$ or receiving $x_{3.1}$ at $t = 2$.

Since the far frame at $t = 0$ becomes the near frame at $t = 1$, observing $x_{2.1}$ allows us see whether a subject re-allocates money between $t = 1$ and $t = 2$ now that these dates are in the immediate future. If a subject's choices imply $x_{2.0} \neq x_{2.1}$, this is taken to violate "time-consistency". In particular, if $x_{2.0} > x_{2.1}$ — i.e. her near-frame switch-point at $t = 1$ is greater than the far-frame switch-point at $t = 0$ — then she this is again taken as evidence that she has higher discount rate in the near frame than in the far frame. She is therefore labelled as "present-biased" in a dynamic sense.[16]

16 If a subject chooses $x_{1.0} \neq x_{2.1}$ and/or $x_{2.0} \neq x_{3.1}$, this is taken to constitute a violation of "time-invariance". This is interpreted to mean that an individual's time preference parameters about the exact same time delay and

2.2.2 *Consumption and liquidity constraints*

Let the subject's utility maximisation problem be as follows:

$$\max_{s_t \in [0, w_t)} U_0 = u(c_0) + \beta E_0 \sum_{t=1}^{\infty} \delta^t u(c_t) \tag{1}$$

$$\text{s.t.} \quad c_t = y_t + (1+r)s_{t-1} - s_t = w_t - s_t > 0 \ \ \forall t \tag{2}$$

$$\text{and} \quad s_t \geqslant 0 \ \ \forall t. \tag{3}$$

Equation 1 describes utility of the canonical "beta-delta" form,where $c_t \geqslant 0$ is her consumption in period t, δ is her discount factor as applied to period t, and β is her degree of present-bias ($\beta < 1$ describes present-bias, while $\beta = 1$ yields time-consistent preferences).[17] For simplicity, assume that present-biased subjects are fully naïve about β, i.e. they implicitly believe that $\beta = 1$.[18] Equation 2 describes the agent's budget constraint: y_t is income in period t, which is taken to be exogenous, and s_t are savings at the end of period t. Equation 3 is a no-borrowing constraint, which is a standard assumption in buffer-stock savings models used to characterise household behaviour in developing countries (Deaton, 1991).[19]

distance from the present may be non-constant over time. This is a separate property from declining or increasing discount rates with time horizon, i.e. present-bias or future-bias.

17 Equation 1 contains the common implicit assumption that the agent's atemporal utility function remains constant over time, and is perceived by the agent as remaining constant over time. See Andersen et al. (2011) for a discussion of the possible implications of "projection bias".

18 Dean and Sautmann, building on the results of Harris and Laibson (2001), show that a perfectly sophisticated subject's choices across a near and a far frame do not identify β, and indeed only deviate due to a statistical bias. Thus an assumption of at least partial naïveté is needed for a subject's choices over a near and a far frame to identify anything about her present-bias.

19 Section 2.4.2 provides evidence that this is a reasonable assumption in the context of the experiment presented below. Dean and Sautmann adopt a different approach, assuming that the interest rate $r(s_t)$ is decreasing and concave in s. This captures the idea of "soft" credit constraints, wherein borrowing becomes increasingly expensive with loan size once more accessible credit sources are exhausted, and saving becomes decreasingly profitable once higher-interest options are full. This effectively nests the case here in which the interest rate on borrowing is infinite, and the interest rate on saving is zero or some constant positive number.

2.2.3 *Experimental choices*

Let the first experimental session take place at $t = 0$. Further assume that savings at $t = 0$ are pre-determined with respect to experimental payments, i.e. a subject has already chosen her current savings to optimise discounted expected utility and does not adjust this when offered x by the experiment at $t = 0$.[20] This assumption is highly plausible for many developing contexts, since most subjects' savings are low and and illiquid and thus unlikely to respond immediately to experimental payments. Furthermore, poor subjects are rarely able to access a significant positive interest rate on saving outside of the experiment. Thus it is highly unlikely that a subject would accept a sooner payment in the hope of saving it outside the experiment, at a higher interest rate than that offered by the experiment.

Let the subject place weight $\kappa \in [0, 1]$ on her outside wealth when making experimental decisions. This nests the cases of narrow framing ($\kappa = 0$), partial asset integration ($\kappa \in (0, 1)$), and full asset integration ($\kappa = 1$). At $t = 0$, for an individual to be indifferent between receiving x at $t = 0$ versus the chosen amount $x_{1.0}$ at $t = 1$, it must be the case that:

$$u(\kappa c_0 + x) + \beta\delta E_0 V_1(\kappa w_1) = u(\kappa c_0) + \beta\delta E_0 V_1(\kappa w_1 + x_{1.0}),$$

which rearranges to:

$$\frac{u(\kappa c_0 + x) - u(\kappa c_0)}{E_0 \left[V_1(\kappa w_1 + x_{1.0}) - V_1(\kappa w_1)\right]} = \beta\delta, \tag{4}$$

where the value function $V_t(w)$ is given by:

$$V_t(w) = \max_{s_t \in [0, w_t)} u(w - s) + \delta E_t V_{t+1}(w_{t+1}). \tag{5}$$

20 Cubitt and Read (2007) establish that if the subject at least partially saves or borrows against the experimental payment x, this causes problems for identifying time preferences. However, the point of the exercise here is to show that similar problems arise even if the subject does not actively seek to arbitrage experimental payments in this way.

Similarly, at $t = 0$, for her to be indifferent between x at $t = 1$ versus $x_{2.0}$ at $t = 2$ it must be the case that:

$$E_0[V_1(\kappa w_1 + x) + \delta V_2(\kappa w_2)] = E_0[V_1(\kappa w_1) + \delta V_2(\kappa w_2 + x_{2.0})],$$

which rearranges to:

$$\frac{E_0\left[V_1(\kappa w_1 + x) - V_1(\kappa w_1)\right]}{E_0\left[V_2(\kappa w_2 + x_{2.0}) - V_1(\kappa w_2)\right]} = \delta. \tag{6}$$

Combining Equations 4 and 6, a static observation of "present-bias" ($x_{1.0} > x_{2.0}$) will occur if:

$$\frac{u(\kappa c_0 + x) - u(\kappa c_0)}{\beta E_0\left[V_1(\kappa w_1 + x_{2.0}) - V_1(\kappa w_1)\right]} > E_0\left[\frac{V_1(\kappa w_1 + x) - V_1(\kappa w_1)}{V_2(w_2 + x_{2.0}) - V_2(w_2)}\right]. \tag{7}$$

For a sufficiently small x, the left-hand side of Equation 7 becomes the marginal rate of intertemporal substitution between $t = 1$ and $t = 0$, and the right-hand side is the marginal rate of intertemporal substitution between $t = 2$ and $t = 1$.

Consistent with the traditional interpretation of static reversals as indicating present-biased preferences, condition 7 may hold if $\kappa = 0$ and $\beta < 1$. However, condition 7 may hold even if $\beta = 1$, if the subject is anticipating higher income from $t = 1$ onwards which she is unable to smooth into the present, due to the no-borrowing constraint.[21] If so, the subject appears "present-biased" because she is experiencing a higher marginal rate of intertemporal substitution now at $t = 0$ than she excepts to experience from $t = 1$ onwards. Moreover, note that it is her subjective expectation that matters: if a subject is optimistic about her future prospects, then she may appear "present-biased" even if objectively her liquidity constraints are likely to remain constant. Thus overall, measured "present-bias" could reflect true present-bias,

21 Alternatively, she may be experiencing a low draw of income at $t = 0$ that she was unable fully to smooth via prior saving, or was lower than expected.

but could also reflect a rational response to liquidity constraints, or indeed overly optimistic income expectations.

Turning to choices at the follow-up session at $t = 1$, for the subject to be indifferent between x at $t = 1$ versus $x_{2.1}$ at $t = 2$ it must be the case that:

$$\frac{u(\kappa c_1 + x) - u(\kappa c_1)}{E_1 \left[V_2(\kappa w_2 + x_{2.1}) - V_2(\kappa w_2)\right]} = \beta\delta. \tag{8}$$

Combining Equations 6 and 8, a dynamic observation of "present-bias" ($x_{2.1} > x_{2.0}$) will occur if:

$$\frac{u(\kappa c_1 + x) - u(\kappa c_1)}{\beta E_1 \left[V_2(\kappa w_2 + x_{2.0}) - V_2(\kappa w_2)\right]} > E_0 \left[\frac{V_1(w_1 + x) - V_1(\kappa w_1)}{V_2(\kappa w_2 + x_{2.0}) - V(\kappa w_2)}\right]. \tag{9}$$

Once more, this may be driven by truly present-biased preferences, i.e. $\beta < 1$ and $\kappa = 0$. However, again if $\kappa > 0$ then such choices may be observed even if $\beta = 1$. For example, it may be that the subject's marginal utility of consumption at $t = 1$ is higher than she originally anticipated it to be, perhaps because she received a low draw from her subjective probability distribution over y_1. Alternatively, it could be that she has revised her expectation of the marginal utility of consumption at $t = 2$ downwards, for example if she has received a signal between $t = 0$ and $t = 1$ that y_2 is going to be higher than originally anticipated.

The reverse arguments also apply for static and dynamic choice reversals in the direction of "future-bias". Such choices will be observed if the reverse of Equations 7 and 9 hold, respectively. Intuitively, a static observation of "future-bias" could be explained by the subject anticipating a *downward* trend in her background consumption between $t = 0$ and $t = 1$. Meanwhile, an observation of dynamic "future-bias" could be explained by her receiving a

subjectively *high* draw of consumption at $t = 1$,[22] or a *downward* revision of her consumption forecast for $t = 2$.

However, recall that the model's assumptions capture the stylized fact that poor subjects are often able to save, albeit imperfectly, but not to borrow. Therefore, subjects are partially able to smooth high current income or positive income shocks into the future via saving; whereas they are unable to smooth high future income into the present, or to smooth away negative income shocks via borrowing. This implies that conditions for observing spurious "future-bias" are less likely to hold than the conditions for observing spurious "present-bias". Thus the model also predicts that, as long as subjects are more credit-constrained than savings-constrained, more spurious "present-bias" than spurious "future-bias" is likely to be observed.

2.2.4 *Predicted effects of income shocks*

It remains an empirical question whether the magnitude of κ is large enough to generate a significant proportion of apparently "present-biased" and "future-biased" choices. Identifying κ — and more broadly whether changes in background consumption can have a causal effect on measures of "present-bias" and "future-bias" — requires exogenous variation in the size and timing of income shocks. If such variation can be generated, the framework suggests a number of predictions that can be used to test between $\kappa = 0$ and $\kappa > 0$. In particular, the experiment described in the rest of this paper tests the following key predictions:

Proposition 1. If a subject receives a positive and unanticipated windfall at $t = 0$:

a. If $\kappa = 0$, there should be no effect on her experimental choices, which are a pure reflection of her preference parameters β and δ.

22 If an agent is risk-averse, then even a draw of $y_1 = E_0[y_1]$ would leave an agent with higher utility at $t = 1$ than the expected utility at $t = 0$ of the draw over y_1. Thus technically, even a draw of y_1 equal to its expectation could be sufficient to induce lower marginal utility at $t = 1$ than was expected at $t = 0$, and thus an observation of dynamic "future-bias".

b. If $\kappa > 0$, from Equation 7 she is less likely to appear statically "present-biased" at $t = 0$ ($x_{1.0} > x_{2.0}$) and more likely to appear statically "future-biased" at $t = 0$ ($x_{1.0} < x_{2.0}$).

Intuition: the positive income shock at $t = 0$ increases consumption at $t = 0$, which decreases the marginal utility of consumption at $t = 0$ and decreases the marginal rate of intertemporal substitution between $t = 0$ and $t = 1$.[23] This decreases the likelihood of Equation 7 holding, and increases the likelihood of its reverse holding. See Appendix 2.8.1 for proof.

Proposition 2. If a subject is told at $t = 0$ that she will receive a positive windfall at $t = 1$:

a. If $\kappa = 0$, there should be no effect on her experimental choices, which are a pure reflection of her preference parameters β and δ.

b. If $\kappa > 0$, from Equation 7 she is more likely to appear statically "present-biased" at $t = 0$ ($x_{1.0} > x_{2.0}$) and less likely to appear statically "future-biased" at $t = 0$ ($x_{1.0} < x_{2.0}$).

Intuition: The announced additional income at $t = 1$ decreases the expected marginal utility of consumption at $t = 1$. Since the subject is credit-constrained, this increases the marginal rate of intertemporal substitution between $t = 1$ and $t = 0$.[24] This increases the likelihood of Equation 7 holding, and decreases the likelihood of its reverse holding. See Appendix 2.8.1 for proof.

The next section describes how the experimental design generates these carefully-timed income shocks, and uses these to test for this evidence of $\kappa > 0$.

23 Recall that, by assumption, savings at $t = 0$ are pre-determined and thus the windfall is entirely consumed at $t = 0$. However, the same qualitative results hold if this assumption is relaxed, as long as the subject's marginal propensity to consume out of windfall income is sufficiently high such that she does not perfectly smooth the windfall into future periods. Thus technically, the joint null for the experiment is that $\kappa = 0$ and/or the agent does perfectly smooth the windfall into future periods via saving.

24 Thus again, technically the joint null for the experiment is that $\kappa = 0$ and/or the agent is actually able to perfectly smooth the anticipated windfall into the present via borrowing.

2.3 EXPERIMENTAL DESIGN

2.3.1 *Surveys and activities*

The design consists of two activity sessions per subject: a baseline on day one, and follow-up two weeks later on day fifteen. Both sessions and all experimental payments take place at the subject's home. This ensures that the subject makes decisions in a familiar setting, and also minimises transaction costs for future-dated payments. Figure 7 in Appendix 2.8.2 illustrates the sequence of activities within each session. The sessions begin with short surveys: a baseline survey at the start of the day one session only, and an income and expenditure survey at the start of both sessions. Subjects then receive an explanation of when and how their participation fee will be paid; see Section 2.3.3 for details. All subjects then receive the same explanation of the activity incentivization procedures: namely that they will be paid according to one of their choices in one of the risk and time preference activities, in one of the sessions, and that the exact question and session will be determined by drawing a series of balls from a bag at the end of each session.[25] Each subject is thereby made to understand that any single one of her forthcoming responses in the risk and time activities has a chance of being selected for payment.[26] The draws for the activity payments are conducted at the end of the session.[27]

25 Such "randomised lottery incentive" procedures may in principle bias subjects' responses, if subjects' behaviour does not obey reduction of compound lotteries. However, evidence suggests that the use of lottery incentives has no significant effect on measured discount rates and time-inconsistency (Laury, 2005; Andersen et al., 2014), except for individuals who exhibit rank-dependent utility (Andersen et al., 2015).

26 While the time-preference and risk-preference activities are incentivized, the probability beliefs and cognitive tasks are not. This is to avoid differential incentives to get the answer right — effectively a form of wage work — across subjects who have just been paid the participation fee versus those who have not.

27 The exact payment protocol is as follows. At the end of the day one session, the individual first draws a ball from a bag to determine whether she will be paid for her responses to today's activities or for her responses to the day fifteen activities. The probabilities are tilted 10%-90% towards being paid for the day fifteen activities, so that most subjects do not end up answering hypothetically at follow-up (since those who are paid for their responses on day one know at follow-up that they will not be paid for their responses at follow-up). This in principle leads to more "high-powered" incentives on day fifteen for the subjects who have not yet been paid for an activity on day one, as these subjects know that they will be paid for one of their responses at follow-up. However, this should not affect the results as there should be no difference in this effect across treatment arms. Next — or at the end of the follow-up session, if the subject draws to be paid at follow-up — the subject draws another ball to determine whether she will be paid for the risk or the time activities; and if she draws for the time activities, she draws further balls to determine which frame (near or far) and question she will be paid for. Once the exact

The enumerator then conducts the activities with the subject. The time preference activities are detailed below in Section 2.3.2. The control activities consist of a certainty equivalent task to measure risk preferences; a probability beliefs exercise to measure optimism; and a digit span test, maths test and numerical Stroop test to measure cognitive functioning. Subjects are pre-randomised by computer into receiving either the block of time preference activities first ("salience" treatment) or the block of control activities first ("control" treatment). The purpose of this is to randomly vary the salience of existing liquidity constraints and the participation fee whilst subjects are answering the time-preference activities.

2.3.2 *Time preference activities*

The time preference activities are multiple price lists, which have long been widely used in the developing-country and developed-country literature (Coller and Williams, 1999; Ashraf et al., 2006; Halevy, 2015).[28] In the near frame, the enumerator asks the subject to choose between 400 PKR (≈ 4 USD) today and increasing amounts in two weeks; whilst in the far frame, the enumerator asks the subject to choose between 400 PKR in two weeks and increasing amounts in four weeks. The order of the frames is also pre-randomised by computer at the individual level.[29] 400 PKR is roughly equivalent to one day of the legal minimum wage in Pakistan, which stands at 12000 PKR/month for formal employment. It also far exceeds the

risk or time activity has been determined, she then draws a further ball to determine which question number within the activity she will be paid for. The enumerator then displays her response and pays her accordingly. If the response involves payments in the future, the enumerator writes out a payment voucher clearly stating the time and amount of the future payment, and a member of the survey team returns to the household on that date to make the payment.

28 Recent literature has seen a shift towards use of "convex time budget" tasks (Andreoni and Sprenger, 2012a) to allow estimation of the discount rate and the curvature of the utility function from one task. However, such tasks are cognitively demanding and from piloting they appeared too complex for subjects in the context of this study. Harrison et al. (2013) suggest that lack of comprehension may even drive the high proportion of corner solutions observed in the original US data of Andreoni and Sprenger. However, corner solutions could also be evidence of arbitrage (Sprenger, 2015). Balakrishnan et al. (2015) find that when subjects in Kenya do appear to understand convex time budgets tasks, responses are highly correlated with responses in multiple price lists. Moreover, since the central concern of this paper is estimation of the effects of liquidity changes on the difference between subjects' near-frame and far-frame choices, multiple price lists are adequate for this purpose.

29 Enumerators also carefully explain to subjects that only one frame will ever pay out, thus concerns that subjects might integrate the later date in the near frame and the earlier date in the far frame should not prove an issue. Nor are they a threat to identification given the randomised frame order.

daily earnings of the women in the sample, who are small business owners or home workers. The time-preference activities therefore involve meaningful stakes, and so any evidence of high levels of non-constant discounting cannot plausibly be attributed to magnitude effects over small stakes (Andersen et al., 2013). Moreover, to avoid rounding effects, the amounts offered are all denominated in multiples of major Rupee bills. To aid comprehension, for each choice the enumerator places the corresponding cash amounts on calendars to demonstrate, as shown in Figure 8 in Appendix 2.8.2.[30] The data suggest very high levels of understanding, with fewer than 2% of subjects exhibiting multiple-switching behaviour in either the first or the second session.[31]

The time discounting activities carry no front-end delay, as the earlier payment in the near frame is dated "today" rather than a future date. This is deliberate, since if subjects' non-constant discounting applies only to the very immediate future, front-end delays may lead to under-estimation of true present-bias. On the other hand, front-end delays are often introduced to mitigate concerns that subjects may not trust experimenters to deliver future-dated payments. In particular, if subjects view future-dated payments as having a positive probability of default, and combine this with some kind of convex weighting probability weighting or taste for certainty, then this may lead them disproportionately to prefer payments dated today and thus spuriously to appear "present-biased" (Halevy, 2008; Andreoni and Sprenger, 2012b). Trust issues could also potentially confound the treatment effect, if receiving the fee on day one rather than day fifteen makes subjects have more trust that future-dated payments will be made. To mitigate this concern, all subjects are given official vouchers for both the participation fees and the activity payments, bearing the logo of the well-known local

30 Whilst subjects are not presented with the full list of choices at the start due to low literacy levels, subjects are told at the start of each activity that there will be sixteen questions in total, in case uncertainty about when the activity will end might lead subjects to switch sooner. The order of the rows is not randomised, as this is cognitively very difficult for subjects and furthermore does not genuinely mitigate concerns about anchoring towards the middle amounts (Andersen et al., 2006).

31 There is a greater degree of multiple-switching in the risk-preference activities, which are certainty equivalent tasks but administered in multiple price list format. As an alternative additional time preference activity, I also attempted to include a present-equivalent task (Abdellaoui et al., 2013). However, according to enumerators this task were very poorly understood; and consistent with this, the data showed a very high degree of multiple-switching behaviour. The treatment effects of the participation fee and the harvest are insignificant for the measures of time-inconsistency constructed from the present-equivalent tasks; tables available on request.

implementing partner.[32] Further, subjects receiving the participation fee on day one receive a receipt which is virtually identical to the vouchers; see Figure 9 in Appendix 2.8.2. This ensures that all subjects receive the same amount of official paperwork at the same point in each session.[33]

2.3.3 *Participation fee timing experiment*

Subjects receive a participation fee of 1000 PKR (approximately 10 USD) for taking part in the two sessions. This is equivalent to a day's income for the average household in the sample (which comprises six individuals), and thus represents a sizeable windfall. However, it is still well within the average range of uncertainty in household income that subjects report facing over the next four weeks.[34] The main experiment randomises at the individual level whether a subject receives her participation fee during the day one session or the day fifteen session. If a subject is going to be paid during the day fifteen session, she is informed on day one that she will be paid on day fifteen.[35]

Subjects learn their treatment status is as follows. When gaining subjects' consent to participate at the start of the day one session, subjects are informed that there will be two sessions — one that day and one exactly two weeks from that day — and that they will receive a 1000 PKR combined participation fee for participating in both sessions. It is not mentioned when

32 Andersen et al. (2014) find that after establishing trust in a similar way, introduction of a front-end delay has statistically insignificant or statistically significant but economically negligible effects on estimated discount rates.

33 The fact that the date of the later payment in the near frame and the earlier payment in the far frame corresponds to the date of the day fifteen session should also reassure subjects that the enumerator will at least deliver any time-preference payouts allocated to those dates, since subjects are told at the start of the day one session how important completion of both sessions is to our study. Admittedly this may lead subjects to be more willing to defer payment in the near frame than in the far frame, since the later date in the far frame is not an session date. This could therefore increase the likelihood that subjects appear "future-biased". However, this effect should not be differential across treatment arms.

34 As part of the income and expenditure survey, subjects are asked their maximum and minimum expected household income for the next four weeks. The average difference between the two reported values, after winsorizing each value at the 95th percentile, is 3920 PKR.

35 Note that there is no "pure control" group who receive no participation fee at any time. This was done for ethical reasons, as participants spend a considerable amount of time participating in the two interviews, and may also hear from other participants who were paid a participation fee.

exactly this fee will be paid.[36] Once the day one session is under way, subjects are given one of the two participation fee and survey timing treatments: half of the subjects in each village are paid the 1000 PKR fee instantly and are given a receipt, whereas the other half are informed that they will be paid during the follow-up session and are given a voucher (as described above). The instructions revealing a subject's treatment status are read out after the surveys but before the activities, so that treatment status cannot influence responses in the day one surveys, but so that all subjects are informed of the timing of their participation fee prior to the activities. For transparency, and in case subjects communicate after the day one sessions, subjects are told that other subjects in their village might be paid either on day one or on day fifteen, and that this is due to administrative reasons and has been determined entirely randomly by computer.

2.3.4 *Survey timing experiment*

All individuals in a village receive their baseline and their follow-up interview on the same date. As a cross-cutting quasi-experiment, the order in which villages are surveyed is also randomised. The timing of the fieldwork is designed such that the start of the wheat harvest falls in the middle of the survey period. Therefore, subjects whose village is assigned to the first half of the survey period receive their baseline session (and possibly their follow-up session) prior to the main onset of the harvest. They are thereby interviewed in a period of tighter liquidity constraints, with the expectation that their liquidity constraints will ease in the near future once harvesting has begun. They may also face greater uncertainty about harvest yields. Meanwhile, subjects whose village is assigned to the second half of the survey period receive their baseline (and follow-up) session after the main onset of the harvest. They are

36 It is reasonable to assume that if subjects infer anything about the timing of the participation fee from this description, it is that the fee will be paid at the end of the second session. Thus those who are eventually paid on day one are better-off than their prior, whereas those who are paid on day fifteen are no less well-off ex post than they expect to be when consenting to participate. However, being paid on day one was deliberately not framed as a "positive surprise" in order to minimise the possible effect of the treatment status on subjects' mood.

thereby interviewed at a time when liquidity constraints have begun to relax and uncertainty has begun to be resolved.

2.3.5 *Experimental predictions*

The treatment arms provide a way to test whether $\kappa > 0$ as outlined in Section 2.2.4:

1. **Participation fee timing experiment:** if a subject receives the participation fee on day one (as opposed to day fifteen), then this constitutes a positive and unanticipated windfall at $t = 0$ (as opposed to information that she will receive a positive windfall at $t = 1$). From propositions 1 and 2:

    a) If $\kappa = 0$, this should have no effect on the likelihood that she appears statically "present-biased" or "future-biased" on day one.

    b) If $\kappa > 0$ she should be less likely to appear statically "present-biased" on day one and more likely to appear statically "future-biased" on day one.[37]

2. **Survey timing experiment:** if a subject receives her day one session before the main onset of the harvest (as opposed to after), then she is more likely to be anticipating an increase in income and consumption in the near future. By the logic of proposition 2, the following predictions also hold:

    a) If $\kappa = 0$, this should have no effect on the likelihood that she appears statically "present-biased" or "future-biased" on day one.

37 Note that this experimental prediction goes in the opposite direction to the prediction of a model in which measured "present-bias" is driven by the "cost of keeping track", i.e. in which subjects prefer to take payments today rather than expend bandwidth thinking about future payment appointments (Haushofer, 2014). Such a model would predict that subjects who receive the participation fee on day one should appear *more* "present-biased", as they would prefer to receive the time preference payment on day one as well.

b) If $\kappa > 0$, she should be more likely to appear statically "present-biased" on day one and less likely to appear statically "future-biased" on day one.

3. **Interaction of experiments:** if a subject receives her day one session before the main onset of the harvest (as opposed to after) then she is also less likely to be able to smooth the timing of the participation fee via her own liquidity or informal borrowing. The following predictions therefore also hold:

   a) If $\kappa = 0$, the interaction of the two treatments should have no effect on the likelihood that she appears statically "present-biased" or "future-biased" on day one.

   b) If $\kappa > 0$, the effect described in experimental prediction 1b should be stronger for individuals interviewed prior to the harvest.

These main predictions, along with secondary predictions and specifications for balance tests, treatment effect estimations and robustness checks were registered in a pre-analysis plan.[38] Where later estimations depart from the pre-analysis plan, this is clearly stated.

2.4 DATA

2.4.1 *Sample and setting*

I implemented this experiment across 53 villages in the Sargodha district of northern Punjab, Pakistan. The sampling frame consisted of female microfinance clients of National Rural Support Program (NRSP), a well-known Pakistani NGO. The rationale for sampling microfinance clients is that although NRSP is engaged in a range of activities and is known to most community members, microfinance clients have had the most frequent interaction with

38 http://www.socialscienceregistry.org/trials/1343.

NRSP and are thus the most likely to trust NRSP to deliver future-dated payments. Moreover, microfinance clients are also the first group to whom NRSP is considering rolling out commitment-savings-type products.[39] Ten women per village were randomly sampled from the list of all past and present NRSP microfinance clients.[40] Five subjects overall were absent at follow-up, in each case reportedly for medical reasons. As a conservative measure, these subjects are dropped from the sample, in case they already anticipated at baseline that they might attrit. This leaves a final sample of 525 women.[41]

The order in which the villages were surveyed was randomised by computer for 48 of the 53 villages.[42] The remaining five villages were brought into the sample later to boost power, thus their interview dates were not randomised. They are therefore excluded from all analyses involving the survey timing treatment.[43] In advance of the survey, local NRSP staff predicted that the modal date for farmers to start harvesting wheat would be 25th April. The baseline survey was then timed such that villages 1-24 received their day one sessions before this approximate harvest start date and villages 25-48 received their day one sessions after this approximate harvest start date.[44]

39 To the extent that group-liability microfinance loans with fixed, regular repayment instalments can themselves be considered commitment savings products, these women have already taken up a type of commitment. They might therefore be expected to exhibit higher rates of true present-bias than a random sample of the population (Bauer et al., 2012; Afzal et al., 2017).

40 Prior to the start of the baseline survey, I privately assigned all clients a random number in Stata using pure randomisation. Clients were then sorted on this number to determine their order of priority for surveying within each village. The first ten clients in the sorted list constituted the designated subjects, whilst the remainder constituted the reserve subjects and were to be used as replacements in order of their position in the sorted list: for example, if one designated subject was not available then enumerators would seek out the eleventh subject in the list, and so on. This procedure was used to obtain a random sample of NRSP borrowers and avoid "cherry-picking" of replacement subjects, for example if the survey team tried to select replacements out of those clients who had a particularly good relationship with NRSP, or who appeared particularly in need of the participation fee. This is an extension of the idea of a "waitlist design" (Hirshleifer et al., 2015).

41 Following an enumerator error, three subjects were incorrectly administered the wrong frame order for the time preference tasks on day fifteen. These subjects' data is kept for analyses involving day one data only, but dropped from all analyses involving day fifteen. A further forty-eight subjects received their follow-up interview on day fourteen rather than day fifteen due to a public holiday. Following the pre-analysis plan, all analyses were run both with and without these individuals, and since there is no significant difference these individuals are included in the full sample.

42 The exact same randomisation procedure was used as for individual clients.

43 This is in line with the pre-analysis plan. Including these villages does not change the results presented below, and only strengthens their significance; tables available on request.

44 The final sessions were concluded a week before the beginning of Ramadan on 5th June. Thus none of the sessions took place during Ramadan, or in the days just prior to it.

2.4.2 *Descriptive statistics and balance*

The experimental data for each subject are taken from the baseline survey conducted on day one, and from the activities which were all repeated on both day one and day fifteen. Table 1 presents the means of pre-specified characteristics from the baseline survey, split by those who received the participation fee on day one versus day fifteen. Important descriptives in terms of the experiment are as follows. First, subjects appear to have limited scope for saving and borrowing, and thus are unlikely to be able fully to smooth consumption or experimental payments over time. Whilst average savings are sizeable, 64% of subjects report having no cash savings at home (not shown). This comprises 37% of subjects who have no savings at all, and a further 27% who only have illiquid savings in ROSCAs or livestock.[45] Just 23% of subjects report that their household has a bank account. Moreover, whilst 97% report having access to formal borrowing — which makes sense given that the sample is drawn from individuals who have been NRSP microfinance clients — just 3% report that they have access to informal borrowing from family, friends or neighbours. Subjects' ability to borrow against participation fees and experimental earnings at short notice is therefore likely to be highly limited. Second, as proxies of trust that future payments will take place, subjects report high levels of trust that a female NRSP representative would keep a future appointment,[46] and indeed that they themselves would keep a future appointment with NRSP, both averaging over 4 on a Likert scale from 1-5.

The column "mean difference" in Table 1 is generated by regressing the variable in question on a dummy for receiving the participation fee on day one, with standard errors robust to individual heteroskedasticity. Since the baseline survey was conducted prior to subjects learning about the timing of their participation fee, "mean difference" therefore constitutes a test of balance for each variable. Only "harvests wheat" is unbalanced at the 10% level, and

45 No subjects report receiving positive nominal interest on their savings. Given inflation and other likely transaction costs and risks of informal saving, this suggests that subjects face a negative real interest rate on saving outside the experiment.

46 All enumerators were female, since all subjects were female. Moreover, piloting showed that some subjects were unwilling to answer this question unless the representative was specified to be female.

this variable is already included in the vector of pre-specified controls for regression analyses. A test that the mean differences are not jointly different from zero also cannot be rejected (McKenzie, 2017).[47]

Table 1: Balance – participation fee timing

	Mean	Pay day 1 Mean	Pay day 15 Mean	Mean Diff.	Total N	Pay day 1 N	Pay day 15 N
Liquidity							
HH income (100,000 PKR)	2.70	2.71	2.69	0.02	523	260	263
Savings (100,000 PKR)	0.44	0.41	0.47	-0.06	525	260	265
Bank account	0.23	0.24	0.22	0.02	525	260	265
Could borrow	0.98	0.98	0.99	-0.01	525	260	265
Could borrow formal	0.97	0.97	0.97	-0.01	525	260	265
Could borrow informal	0.03	0.03	0.03	0.00	525	260	265
Harvest							
Harvests wheat	0.37	0.33	0.40	-0.07*	525	260	265
Demographics							
Muslim	0.88	0.88	0.87	0.02	525	260	265
Education (years)	2.24	2.24	2.23	0.01	525	260	265
Housewife	0.75	0.74	0.75	-0.02	525	260	265
Age	37.85	38.23	37.48	0.74	525	260	265
Married	0.88	0.88	0.88	-0.00	525	260	265
HH size	6.30	6.27	6.33	-0.06	524	260	264
HH head	0.06	0.05	0.07	-0.02	525	260	265
HH decisions (index 0-1)	0.11	0.11	0.10	0.01	525	260	265
Trust							
Trust NRSP (1-5)	4.22	4.24	4.20	0.04	525	260	265
Trust self (1-5)	4.42	4.45	4.40	0.05	525	260	265

Notes: All variables are taken from the baseline survey, conducted at the start of the day one session prior to revelation of treatment status. Treatment status — receiving the participation fee on day one or on day fifteen — is computer-randomised prior to session. Mean diff. represents difference in means across the two treatment arms. In line with the pre-analysis plan, *, ** and *** indicate significance of this difference at the 10%, 5% and 1% levels respectively, as estimated from a regression of the variable of interest on the treatment indicator, with standard errors robust to individual heteroskedasticity. 100,000 PKR ≈ 1,000 USD.

Table 2 presents baseline survey information split by whether subjects randomly received their baseline interview prior to or after the main onset of the harvest. This is not a test of

47 Balance tests were also performed with dummies for the activity order (receiving the time-preference activities rather than the control activities first) and for the frame order (receiving the near frame rather than the far frame first within the time preference activities); tables available on request. For the activity order, only "married" is imbalanced the 10% level. In line with the pre-analysis plan, "married" is therefore added to the controls in all estimations which control for the activity order. For the frame order, no variables show imbalance. The joint test of balance across variables is not significantly different from zero for either the activity order or the frame order.

pre-treatment balance, since subjects clearly knew when responding to the baseline survey whether their interview was taking place before or after the harvest. Thus rather than suggesting a failure of randomisation, the observed imbalances in ability to borrow informally, household decision-making, and trust in NRSP and oneself to keep future appointments likely reflect real differences before and after the main onset of the harvest. For example, the imbalance on "trust" may reflect the fact that women interviewed just before the harvest think of "future appointments" as taking place during the busy harvest time, whereas women interviewed during the harvest think of "future appointments" as taking place in a less busy time. An exception is highest education attained, whose imbalance is likely due to pure chance. Ability to borrow informally and formally and education are already included in the vector of pre-specified controls. The index of household decision-making and both measures of trust are added as controls in all regressions involving the survey timing treatment.

The key descriptive statistic for the quasi-experiment is whether liquidity is indeed increasing over the survey period. Figure 1 plots fractional polynomial fits of subjects' total income for the last two weeks, and expected income for the next two weeks, by the date on which they received the baseline survey. The first panel shows that income received in the last two weeks is smoothly increasing over the entire survey period. There is no sharp discontinuity around late April when the bulk of harvesting begins, but this likely reflects that some farmers start harvesting earlier and others later, and that other forms of economic activity already begin to take place in the run up to the harvest. The second panel shows that expected income is also increasing, except for the final weeks of the survey period when harvesting has ended. Overall, it therefore appears that subjects are on a path of increasing income, at least until the very end of the survey period.[48]

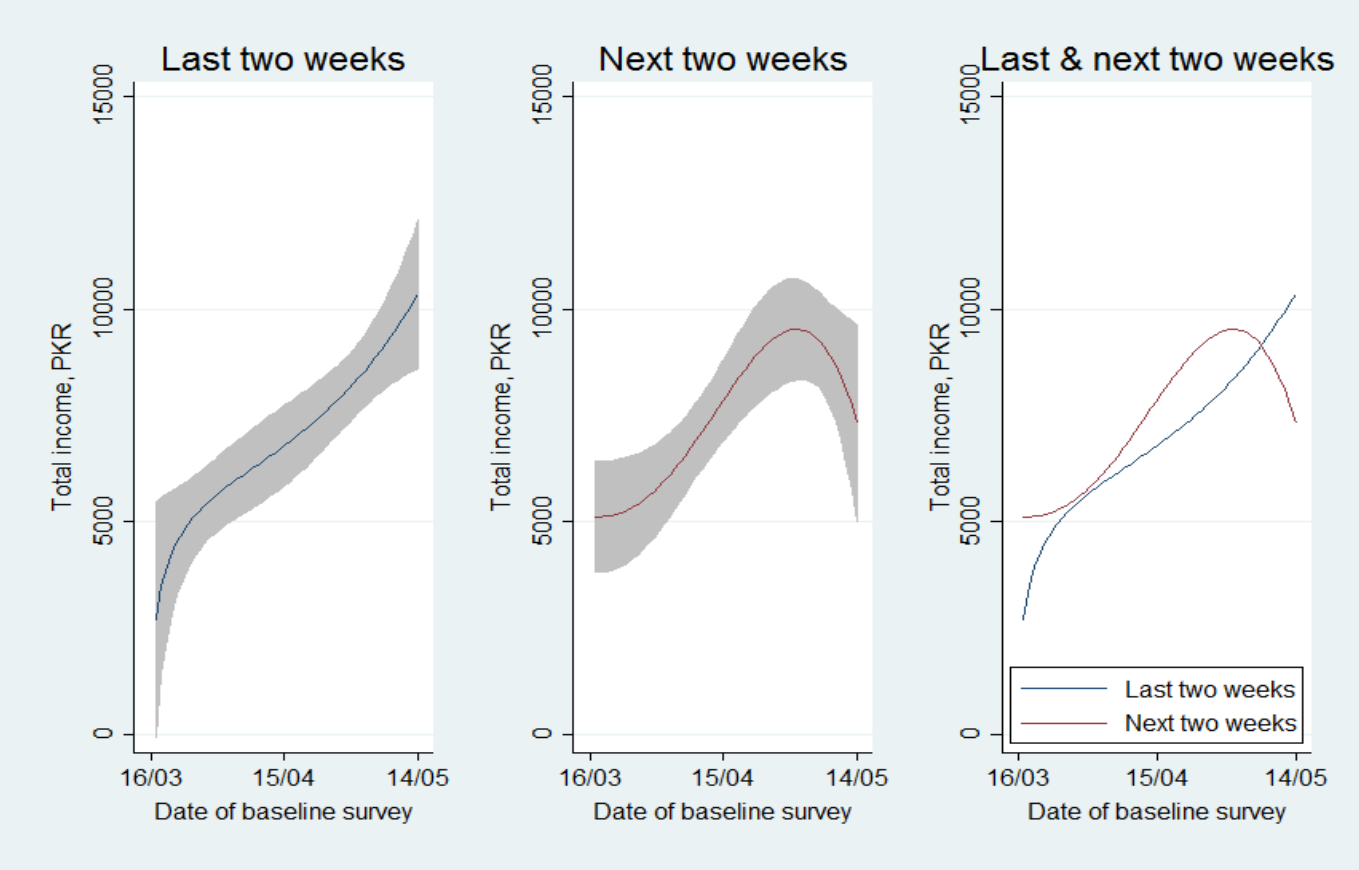
48 As the third panel shows, expected income in the next two weeks appears to be slightly larger than income in the past two weeks until the end of the survey period. However, given the noise in these measures their confidence intervals (not shown) overlap for virtually all survey dates.

Table 2: Descriptive statistics by survey timing

	Mean	Pre-harvest Mean	Post-harvest Mean	Mean Diff.	Total N	Pre-harvest N	Post-harvest N
Liquidity							
HH income (100,000 PKR)	2.70	2.64	2.77	-0.12	523	285	238
Savings (100,000 PKR)	0.44	0.42	0.47	-0.05	525	286	239
Bank account	0.23	0.23	0.23	-0.00	525	286	239
Could borrow	0.98	0.98	0.99	-0.01	525	286	239
Could borrow formal	0.97	0.95	0.99	-0.03	525	286	239
Could borrow informal	0.03	0.06	0.00	0.05***	525	286	239
Harvest							
Harvests wheat	0.37	0.40	0.33	0.07	525	286	239
Demographics							
Muslim	0.88	0.85	0.90	-0.05	525	286	239
Education (years)	2.24	2.55	1.87	0.68**	525	286	239
Housewife	0.75	0.72	0.78	-0.06*	525	286	239
Age	37.85	37.48	38.30	-0.83	525	286	239
Married	0.88	0.87	0.90	-0.03	525	286	239
HH size	6.30	6.29	6.31	-0.02	524	286	238
HH head	0.06	0.07	0.05	0.02	525	286	239
HH decisions (index 0-1)	0.11	0.13	0.08	0.04*	525	286	239
Trust							
Trust NRSP (1-5)	4.22	4.16	4.29	-0.12***	525	286	239
Trust self (1-5)	4.42	4.39	4.46	-0.08***	525	286	239

Notes: All variables are taken from the baseline survey, conducted at the start of the day one session. Treatment status — receiving the baseline interview before or after the onset of the harvest — is computer-randomised prior to the start of the entire survey. Mean diff. represents difference in means across the two treatment arms. In line with the pre-analysis plan, *, ** and *** indicate significance of this difference at the 10%, 5% and 1% levels respectively, as estimated from a regression of the variable of interest on the treatment indicator, with standard errors clustered at the village level. For time-varying controls, this is not a test of balance, but rather shows how these baseline covariates differ when the survey is conducted before or after the harvest. HH decisions, Trust NRSP and Trust self are added to the controls (the others imbalanced covariates are already included in the vector of controls).

Figure 1: Income in the last two & next two weeks, by baseline survey date



Notes: Plots display the predicted value and 95% confidence intervals for a fractional polynomial fit of the reported income values, after winsorizing at the 95th percentile. Confidence intervals are suppressed in the third panel to allow visibility of both lines simultaneously. Values are as measured on day one during the baseline survey. Respondents are first asked to list all household income sources, then to report income and expected income values separately by source. The survey tablet calculates the totals, which the enumerator verifies with the respondent.

2.4.3 *Measured "time-inconsistency"*

Figures 2 and 3 show the distributions of near-frame switch-points on day one ($x_{1.0}$) and far-frame switch-points on day one ($x_{2.0}$), by participation fee timing and whether the day one interview is pre- or post-harvest respectively. Across all frames it is clear that, whilst many subjects are prepared to wait for payments between 400 and 500 Rs two weeks later rather than 400 Rs at the earlier date, a substantial fraction of subjects are unwilling to wait until offered much larger amounts. Furthermore, in each frame 15-20% of subjects (labelled as ">1000") are unwilling to wait two weeks even when offered the maximum delayed payment amount of 1000 Rs. This may imply that their true marginal rate of intertemporal substitution is exceedingly high. However, it may also signal either lack of understanding or unwillingness to accept any deferred payments because of lack of trust. The results below are therefore presented with and without these subjects.

Table 3 displays the relationship between subjects' inconsistencies across the near and the far frame on day one ($x_{1.0} \neq x_{2.0}$) and across the near and the far frame on day fifteen ($x_{2.1} \neq x_{3.1}$). Several patterns emerge. First, the level of apparently "time-inconsistent" decision-making is high, with less than half of the sample making the same choice across the near and the far frames on day one or on day fifteen. These inconsistencies across frames are unlikely to be due to indifference or noise: for those subjects who are classified as statically "present-biased" on day one ($x_{1.0} > x_{2.0}$), the average number of questions that the subject differs by across the near and far frame is 4.1. For subjects who are classified as statically "future-biased" on day one ($x_{1.0} < x_{2.0}$) the average numbers of questions that subjects differ by across frames is 3.9. Only 24.1% (23.1%) of subjects classified as statically "present-biased" ("future-biased") on day one deviate by just one question across the near and the far frame.

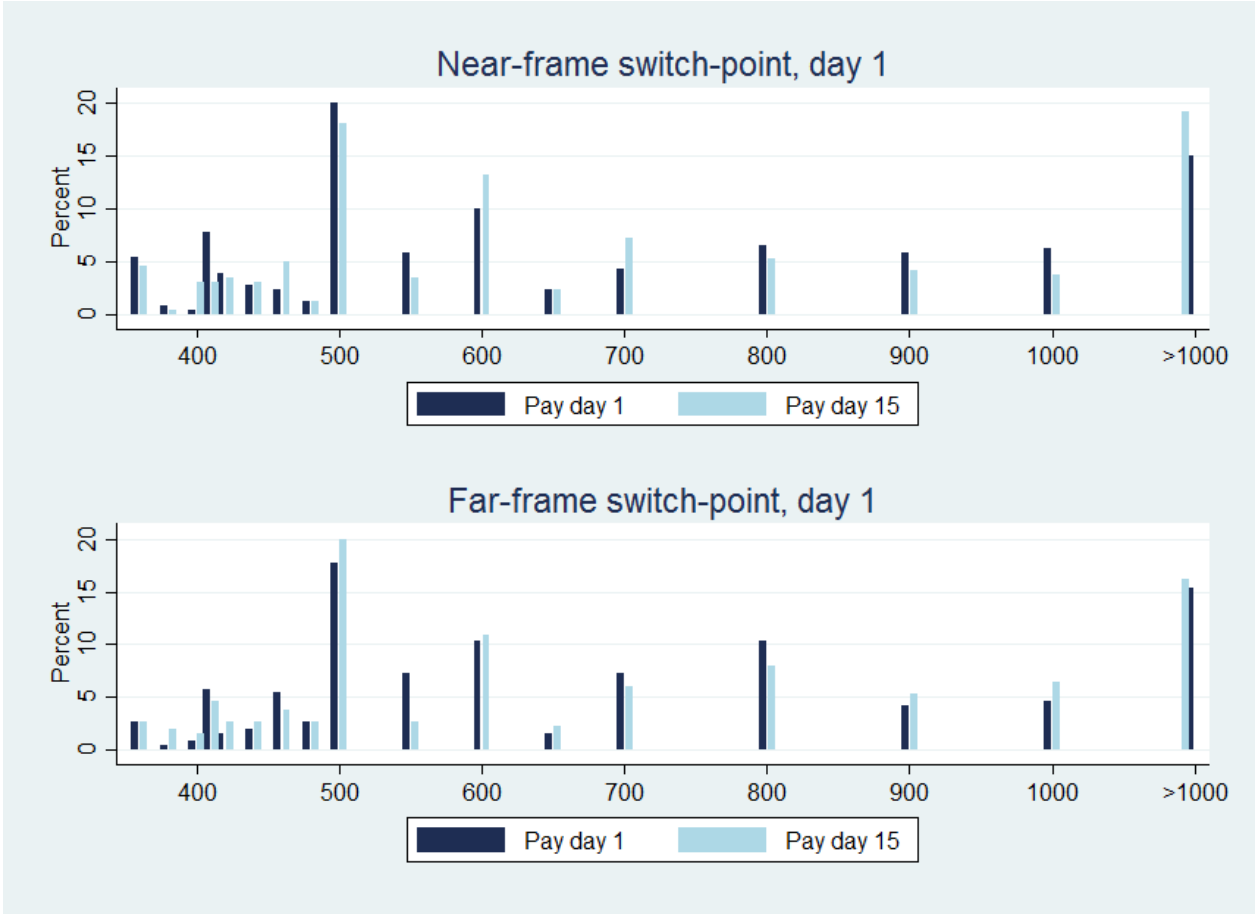
Second, in line with studies mentioned earlier, the proportions of subjects appearing "future-biased" is always similar in magnitude to the proportion appearing "present-biased". Specifically, 35.2% appear "future-biased" and 26.1% "present-biased" on day one, whilst 23.2% appear "future-biased" and 27.4% appear "present-biased" on day fifteen.

Table 3: Static inconsistency, day one & day fifteen

	Day 15 static choices			
Day 1 static choices	"Present-biased"	"Time-consistent"	"Future-biased"	Total
"Present-biased"	41	55	40	136
	(7.9%)	(10.5%)	(7.7%)	(26.1%)
"Time-consistent"	51	119	32	202
	(9.8%)	(22.8%)	(6.1%)	(38.7%)
"Future-biased"	51	84	49	184
	(9.8%)	(16.1%)	(9.4%)	(35.2%)
Total	143	258	121	522
	(27.4%)	(49.4%)	(23.2%)	(100.0%)

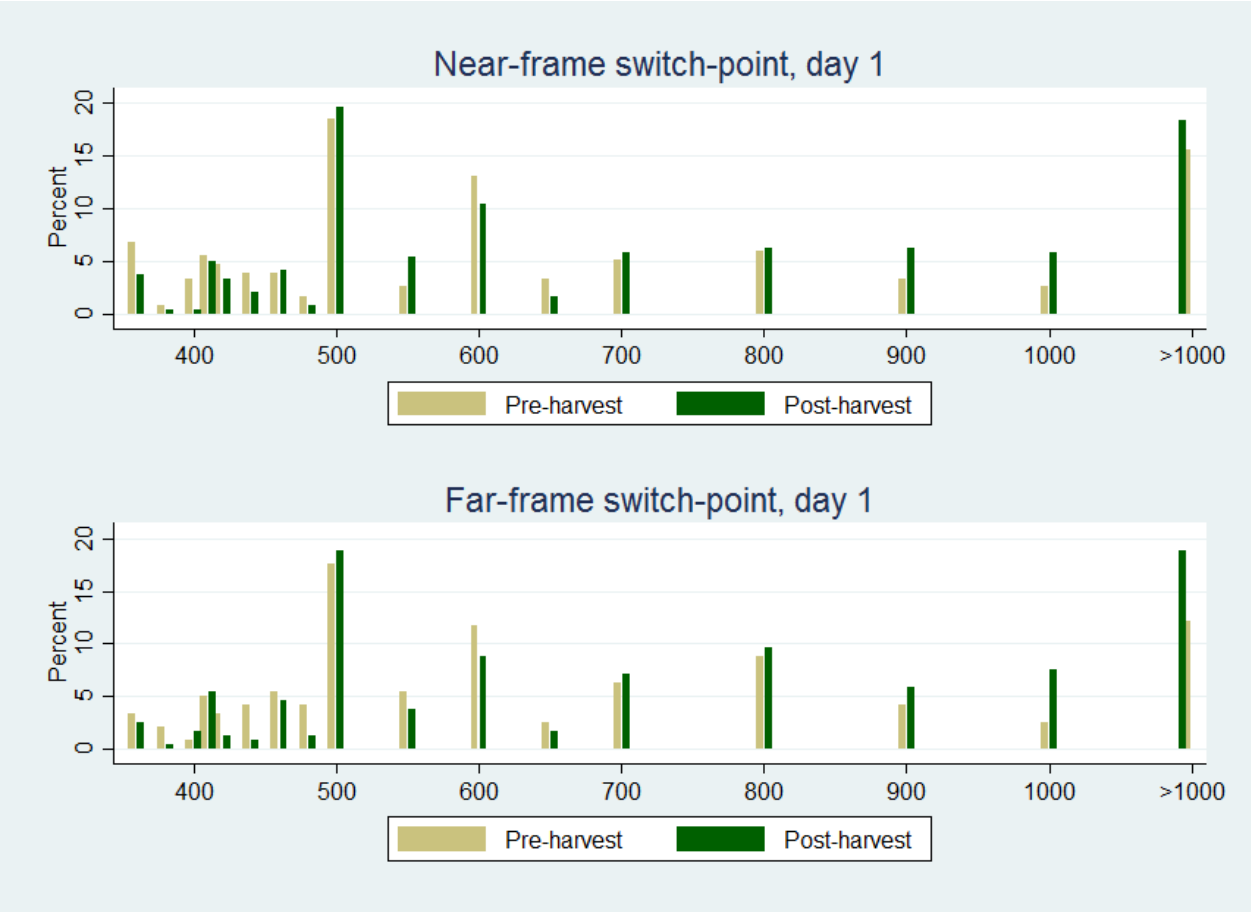
Notes: "Present-biased", "time-consistent" and "future-biased" are dummy variables indicating that a subject's near-frame switch-point in the multiple price list activity on that day was respectively greater than, less than or the same as her far-frame switch-point on that day. Subjects are classified as either "present-biased", "time-consistent" or "future-biased". Subjects received the same activity order and frame order on day fifteen as they had received on day one.

Figure 2: Effects of windfall timing on day one switch-points



Notes: All variables are measured on day one. Switch-points are in Pakistani Rupees (100 PK ≈ 10 USD). In each frame, subjects were asked to choose between 400 PKR on the earlier date or the amount shown on the later date. The switch-point in each frame is equal to the first value at which the subject chose to receive the payment on the later date. In the near frame, the earlier date was today and the later date was two weeks from today. In the far frame, the earlier date was two weeks from today and the later date was four weeks from today.

Figure 3: Effects of survey timing on day one switch-points



Notes: All variables are measured on day one. Switch-points are in Pakistani Rupees (100 PKR ≈ 1 USD). In each frame, subjects were asked to choose between 400 PKR on the earlier date or the amount shown on the later date. The switch-point in each frame is equal to the first value at which the subject chose to receive the payment on the later date. In the near frame, the earlier date was today and the later date was two weeks from today. In the far frame, the earlier date was two weeks from today and the later date was four weeks from today.

Third, "time-consistency" appears to increase from day one to day fifteen, from 38.7% to 49.4% of the sample. Notably, this increase in "time-consistency" comes from a reduction in the proportion of subjects making "future-biased" decisions, from 35.2% on day one to 23.2% on day fifteen. This may be indicative of a learning effect, whereby subjects made "future-biased" decisions through misunderstanding on day one but not on day fifteen. Alternatively, given that half of day one sessions take place before the harvest but (by construction) less than half of day fifteen sessions take place before the harvest, the increase in time-consistent decision-making on day fifteen could also reflect a mechanical effect of the sample being on average less liquidity-constrained on day fifteen than day one. Meanwhile there is no evidence of a decrease in the proportion of "present-biased" decisions on day fifteen. This goes against a hypothesis in which "present-bias" on day one is driven by lack of trust that enumerators will return with payments: on day fifteen subjects have already seen enumerators return once, and thus should trust them more, but subjects exhibit the same degree of "present-bias".

Lastly, just 7.9% appear statically "present-biased" on both dates, and just 9.4% appear statically "future-biased" on both dates. This suggests that a model attributing choices solely to unchanging, underlying β and δ is unlikely to be able to explain the majority of "time-inconsistent" choices.

Table 4 displays the relationship between static inconsistencies on day one ($x_{1.0} \neq x_{2.0}$) and dynamic "time-inconsistent" revisions on day fifteen ($x_{2.1} \neq x_{2.0}$). Again, several features of the data are striking. First, whilst most subjects (38.7%) appear "time-consistent" in their day one static choices ($x_{1.0} = x_{2.0}$), relatively few subjects (19%) appear dynamically "time-consistent" on day fifteen ($x_{2.1} = x_{2.0}$). Again, revisions to decisions on day fifteen are unlikely to be due to small mistakes or noise: for those subjects who are classified as dynamically "present-biased" i.e. $x_{2.1} > x_{2.0}$ on day fifteen, the average number of questions that the subject differs by across the near frame on day fifteen compared to the far frame on day one is 5.6. For those classified as dynamically "future-biased", i.e. $x_{2.1} > x_{2.0}$, the number of questions is 4.3. Only 9.1% (20.3%) of dynamically"present-biased" ("future-biased") subjects deviate

by just one question across the near frame on day fifteen compared to the far frame on day one.

Table 4: Static inconsistency on day one & revisions on day fifteen

	Day 15 dynamic revisions			
Day 1 static choices	"Present-biased"	"Time-consistent"	"Future-biased"	Total
"Present-biased"	95	11	30	136
	(18.2%)	(2.1%)	(5.7%)	(26.1%)
"Time-consistent"	81	69	52	202
	(15.5%)	(13.2%)	(10.0%)	(38.7%)
"Future-biased"	83	19	82	184
	(15.9%)	(3.6%)	(15.7%)	(35.2%)
Total	259	99	164	522
	(49.6%)	(19.0%)	(31.4%)	(100.0%)

Notes: For static reversals on day one, "present-biased", "time-consistent" and "future-biased" are dummy variables indicating that a subject's near-frame switch-point on day one was respectively greater than, less than or the same as her far-frame switch-point in the multiple price list activity on day one. Subjects are classified as either "present-biased", "time-consistent" or "future-biased". For dynamic revisions on day fifteen, "present-biased", "time-consistent" and "future-biased" are dummy variables indicating that a subject's near-frame switch-point in the multiple price list activity on day fifteen was respectively greater than, less than or the same as her far-frame switch-point on day one. Subjects are again classified as either "present-biased", "time-consistent" or "future-biased". Subjects received the same activity order and frame order on day fifteen as they had received on day one.

Second, more subjects make revisions in a "present-biased" direction (49.6%) than in a "future-biased" direction (31.4%), although the latter is still a large proportion. Similarly, a larger number of subjects make both static day one choices and dynamic day fifteen revisions in a "present-biased" direction (18.2%) than make both in a "future-biased" direction (15.7%). However, if making both decisions in the same direction is indicative of an underlying "type", then the latter statistic suggests that a large number of subjects could be classified as "future-biased".

Lastly, a substantial proportion of subjects exhibit one type of static "time-inconsistency" on day one but another type of dynamic "time-inconsistency" on day fifteen. Thus again, a model in which subjects' choices solely reflect their "deep" time preference parameters is unlikely to be able to explain a large proportion of subjects' choices.

Finally, to give a sense of whether apparent "present-bias" is associated with a demand for commitment savings, Tables 5 and 6 show the relationship between measured "time-inconsistency" and whether subjects report being interested in taking up any of a number of hard and soft commitment savings accounts described to them in the baseline survey.[49] While subjects who appear "present-biased", either statically or dynamically, are more likely than not to report being willing to take up such a product, the same holds for apparently "time-consistent" and apparently "future-biased" individuals. Overall, a large proportion of the reported demand for commitment savings products comes from individuals who, according to the tasks, are "time-consistent" or indeed "future-biased". Thus if present-biased preferences are the main driver of demand for commitment savings, then subjects' choices over time-dated monetary payments do not appear to be capturing such preferences well.[50] Alternatively (or additionally), it may be that a large part of the demand for commitment savings is instead driven by motives other than present-bias, for example to shield resources from one's spouse or relatives, or indeed because subjects have limited access to other types of savings accounts.

49 Individual subjects' responses across the hard and softer types of commitment account offered were highly correlated, hence Table 5 uses a simple indicator for whether a subject says yes to any of the types of commitment. Subjects' responses to these questions did not constitute actual take-up decisions, since the products are not currently offered by NRSP. However, since the survey was introduced as research to help NRSP understand the financial decision-making of its clients, and to determine what kinds of products it might offer in the future, subjects hopefully do perceive an incentive to answer truthfully. Interestingly, the reported demand for every type of commitment account was moderately higher in post-harvest rather than pre-harvest baseline interviews (results available on request). Thus perhaps unsurprisingly, it appears that subjects are more interested in taking up commitment savings when they have more cash-on-hand.

50 A slightly different explanation could be that some subjects who appear "time-consistent" or "future-biased" are in fact present-biased but already use costly commitment technologies. They may demand commitment savings accounts if such accounts offer a less costly way to commit. Toussaert (2015) makes a very similar point for the case in which subjects have Gul-Pesendorfer temptation preferences, and commitment removes the need for costly self-control. However, again this points to monetary tasks not identifying present-bias or other forms of self-control problems well, although for a different reason.

Table 5: Static inconsistency on day one & demand for commitment

	Day 1 demand		
Day 1 static choices	No	Yes	Total
"Present-biased"	66	71	137
	(12.6%)	(13.5%)	(26.1%)
"Time-consistent"	89	113	202
	(17.0%)	(21.5%)	(38.5%)
"Future-biased"	73	113	186
	(13.9%)	(21.5%)	(35.4%)
Total	228	297	525
	(43.4%)	(56.6%)	(100.0%)

Notes: For day one static choices, "present-biased", "time-consistent" and "future-biased" are dummy variables indicating that a subject's near-frame switch-point on day one was respectively greater than, less than or the same as her far-frame switch-point in the multiple price list activity on day one. Subjects are classified as either "present-biased", "time-consistent" or "future-biased". For day one reported demand for commitment savings, "yes" is a dummy variable equal to one if a subject reported that they would be interested in taking up at least one of the six commitment savings products asked about in the baseline survey.

Table 6: Revisions on day fifteen & demand for commitment

	Day 1 demand		
Day 15 dynamic revisions	No	Yes	Total
"Present-biased"	117	142	259
	(22.4%)	(27.2%)	(49.6%)
"Time-consistent"	42	57	99
	(8.0%)	(10.9%)	(19.0%)
"Future-biased"	67	97	164
	(12.8%)	(18.6%)	(31.4%)
Total	226	296	522
	(43.3%)	(56.7%)	(100.0)

Notes: For day fifteen dynamic choices, "present-biased", "time-consistent" and "future-biased" are dummy variables indicating that a subject's near-frame switch-point on day fifteen was respectively greater than, less than or the same as her far-frame switch-point in the multiple price list activity on day one. Subjects are classified as either "present-biased", "time-consistent" or "future-biased". For day one reported demand for commitment savings, "yes" is a dummy variable equal to one if a subject reported that they would be interested in taking up at least one of the six commitment savings products asked about in the baseline survey.

2.5 RESULTS

2.5.1 *Participation fee timing*

Figure 4 presents the results of the participation fee timing experiment graphically, while Table 7 presents the same results in regression form. The estimating equation is the following logit model:[51]

$$\Pr\left[Y_{idvt} = 1 | treat_{iv}, \mathbf{X_{iv}}, \eta_{vdt}\right] = \Lambda\left[\alpha_1 treat_{iv} + \boldsymbol{\gamma}\mathbf{X_{iv}} + \eta_{vdt}\right], \tag{10}$$

where Y_{ivtd} is a binary outcome variable for subject i in village v on session day $d \in \{1, 15\}$ at calendar date t. The dummy variable $treat_{iv}$ is an indicator equal to one if subject i was assigned to being paid the participation fee on day one rather than day fifteen.[52] $\mathbf{X_{iv}}$ is a vector of pre-specified time-invariant individual controls from the baseline survey: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). η_{vdt} is a village fixed effect, included because randomisation was stratified by village (Bruhn and McKenzie, 2009).[53] Standard errors are robust to individual-level heteroskedasticity. $\hat{\alpha}_1$ represents the average treatment effect, since assignment to treatment and actual treatment status coincided by construction.[54] Since village fixed effects are included, $\hat{\alpha}_1$ is identified by comparing within each

51 Logit models are presented here as the preferred specifications, given that these are binary dependent variables. However, in line with the pre-analysis plan, ordinary least squares specifications of the form $y_{ivdt} = \alpha_1 treat_{iv} + \gamma\mathbf{X_{iv}} + \eta_{vdt} + \epsilon_{ivdt}$ were also estimated. All results are virtually identical across logit and OLS models; available on request.

52 The pre-analysis plan specifies all regressions in terms of a treatment dummy which is equal to one if subjects are paid on day *fifteen*. However, subsequent presentations made clear that audiences were more comfortable interpreting payment on day *one* as the treatment and payment on day fifteen as the control, as paying the subjects at the end of a study is the more common practice. The signs of predicted treatment effects are thus are straightforwardly reversed from the pre-analysis plan.

53 These are denoted as village-day-date fixed effects, given that 10 is estimated separately for day one and day fifteen, and that all subjects in a village were surveyed on the same date for their day one and day fifteen sessions.

54 Treatment-arm-specific instructions displayed on the enumerator's survey tablet, as well as treatment-arm-specific payment and voucher procedures which had to be followed and verified by the enumerator before she could proceed.

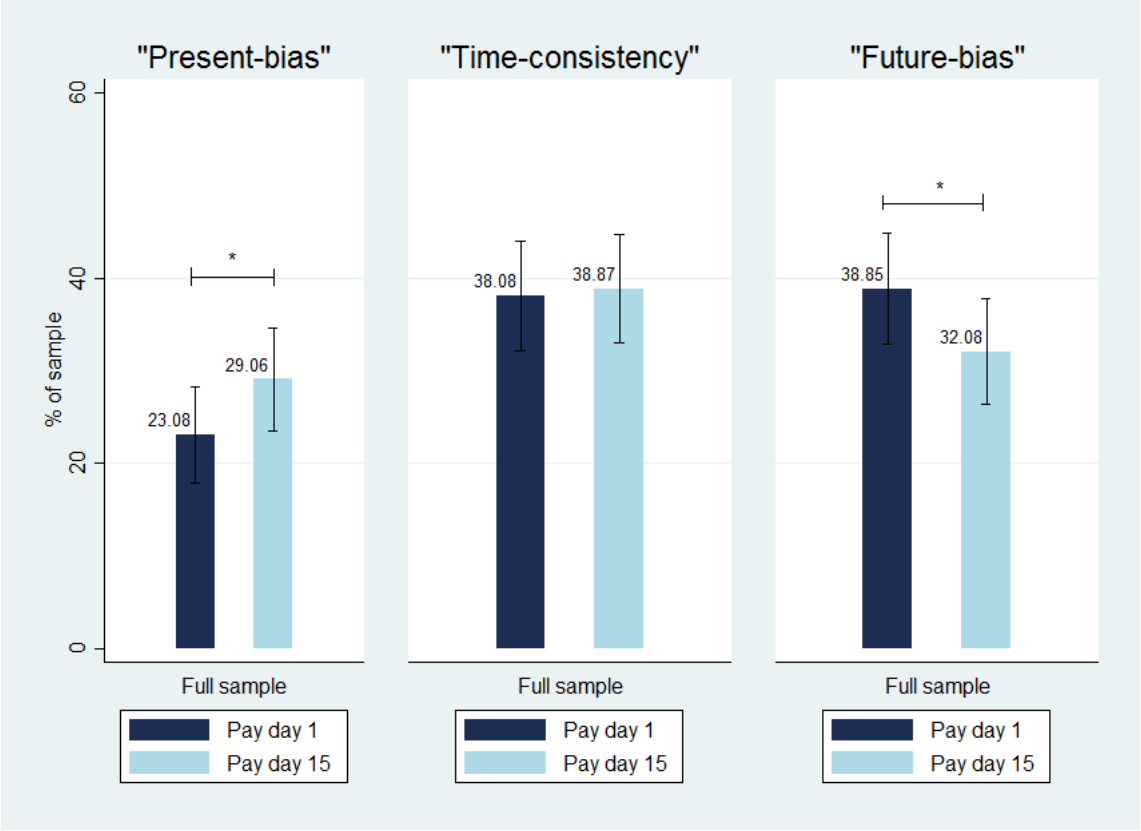
village the five subjects who were paid their participation fee on day one to the five subjects who were paid on day fifteen. As described above, a subject's day one choices are classified as "present-biased" ($x_{1.0} > x_{2.0}$), "time-consistent" ($x_{1.0} = x_{2.0}$) or "future-biased" ($x_{1.0} < x_{2.0}$) by comparing her choices across the near and the far frame on day one.

The key experimental prediction 1b from Section 2.3.5 is that $\hat{\alpha}_1$ should be negative when the dependent variable is appearing "present-biased" on day one and positive when the dependent variable is appearing "future-biased" on day one. Columns (1) and (3) of Table 7 show that the results are consistent with these predictions: receiving the participation fee on day one, as opposed to being told on day one that the participation fee will be paid on day fifteen, reduces the probability of a subject appearing "present-biased" by 7.1 percentage points (p-value 0.081) and increases the probability of her appearing "future-biased" by 8.0 percentage points (p-value 0.080). There is no effect on the probability of appearing "time-consistent", consistent with there being no clear theoretical prediction: the treatment should shift some subjects from appearing "present-biased" into appearing "time-consistent", but shift others out of appearing "time-consistent" and into appearing "future-biased".

Columns (4), (5) and (6) of Table 7 restrict the sample to the most conservative definition of "switchers": subjects who at some point switch to choosing a later payment in both the near frame and the far frame, on both day one and day fifteen. For "switchers", receiving the participation fee on day one reduces the probability of appearing "present-biased" by 11.1 percentage points (p-value 0.078) and increases the probability of appearing "future-biased" by 15.8 percentage points (p-value 0.012). Thus the point estimates are more significant for this subsample and the effect sizes appear larger, although not significantly so. The effects observed in the full sample in columns (1) and (3) are therefore clearly not driven by "non-switchers", i.e. subjects who may have had severe trust concerns, or who may have been trying to signal their poverty to NRSP.

For completeness, Table 13 in Appendix 2.8.3 shows OLS regressions estimating the average treatment effect on the switch-points in the near and the far frame respectively. These outcome variables do not directly translate into the percentage of individuals appearing "present-

Figure 4: Effects of windfall timing on day one static inconsistency



Notes: All dependent variables are measured on day one. "Present-biased" ("future-biased") is a dummy indicating a near-frame switch-point greater than (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "future-biased" or "time-consistent".

Table 7: Treatment effects on day one static inconsistency – windfall timing

	(1) **"Present-biased" Day one All** Mfx / (s.e.)	(2) **"Time-consistent" Day one All** Mfx / (s.e.)	(3) **"Future-biased" Day one All** Mfx / (s.e.)	(4) **"Present-biased" Day one Switchers** Mfx / (s.e.)	(5) **"Time-consistent" Day one Switchers** Mfx / (s.e.)	(6) **"Future-biased" Day one Switchers** Mfx / (s.e.)
Participation fee timing						
Pay on day 1	-0.071*	-0.009	0.080*	-0.111*	-0.050	0.158**
	(0.041)	(0.048)	(0.046)	(0.063)	(0.070)	(0.067)
Controls	✓	✓	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓	✓	✓
Observations	493	503	504	260	272	294
Control mean	0.291	0.389	0.321	0.297	0.354	0.348

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee treatment status. "Present-biased" ["time-consistent"]("future-biased") is a dummy indicating a near-frame switch-point greater than [equal to] (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "time-consistent" or "future-biased". "Switchers" are subjects who at some point switch to accepting a deferred payment in both the near frame and the far frame, on both day one and day fifteen. Pre-specified controls are: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). Reported effects represent the marginal effects at the mean. N=525 for the full sample, and N=294 for switchers. Sample sizes differ where village fixed effects perfectly predict the outcome variable.

biased" or "future-biased", since that depends on the relative movement in the near-frame and the far-frame switch-points for each individual. The treatment effect of *pay on day one* appears negative for all frames, i.e. making subjects appear more patient. The point estimate also appears larger for the near frame than the far frame on day one, which is in line with a reduction in appearing "present-biased". However, none of these estimates are significant.

Turning to day fifteen measures of time preferences, columns (1)-(3) of Table 14 in in Appendix 2.8.3 show that there is no significant treatment effect when the dependent variable is appearing statically "present-biased" ($x_{2.1} > x_{3.1}$), "time-consistent" ($x_{2.1} = x_{3.1}$) or "future-biased" ($x_{2.1} < x_{3.1}$) on day fifteen. This is at odds with the predictions of the model if $\kappa > 0$ and if subjects who receive the participation fee on day one do not save much of it: in that case we would expect to see the reverse pattern of the treatment effects observed on day one, and indeed this was the set of predictions registered in the pre-analysis plan. A possible explanation is that subjects who received the participation fee on day one did in fact save a substantial amount of it. If so, they would look very similar on day fifteen to subjects who only just received the participation fee, and thus we would not see a significant difference in subjects' choices across treatment arms.

In contrast, the model specifically predicts that there should be no effect on day fifteen revisions even if $\kappa > 0$. This is because revisions are driven by subjects reacting to new information about income or consumption, which they receive after the day one choices but before the day fifteen choices. The experiment does not generate any difference in this information, since the timing of the participation fee is already revealed prior to the day one choices. Columns (4)-(6) of Table 14 support this prediction of the model: there is no significant treatment effect when the dependent variable is making "present-biased" revisions on day fifteen ($x_{2.1} > x_{2.0}$), appearing dynamically "time-consistent" on day fifteen ($x_{2.1} = x_{2.0}$), or making "future-biased" revisions on day fifteen ($x_{2.1} < x_{2.0}$).

2.5.2 *Survey timing*

Figure 5 shows the results of the survey timing experiment graphically, while Table 8 presents them in regression form. The estimating equation is the following logit model:

$$\Pr[Y_{idvt} = 1 | \text{pre-harvest}_v, \mathbf{X_{iv}}, \eta_{vdt}] = \Lambda[\alpha_1 \text{pre-harvest}_v + \gamma \mathbf{X_{iv}}], \tag{11}$$

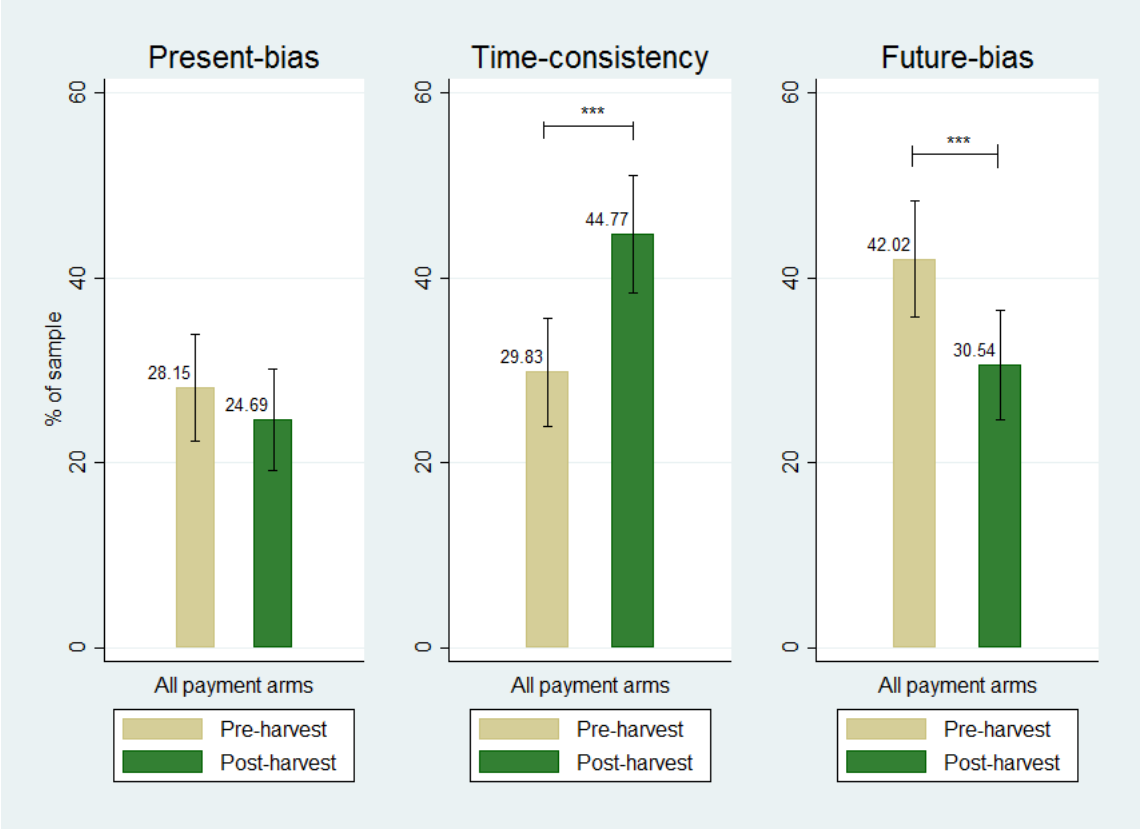
where pre-harvest_v is a dummy indicating that the subject's village received its baseline interviews prior to the main onset of the wheat harvest.[55] Standard errors are clustered at the village level, since survey dates were randomised at the village level. As explained above, as a conservative measure this equation is estimated on the first 48 villages only. Point estimates become slightly larger and even more significant if the final five villages are included in the estimating sample; results available on request.[56]

In terms of prediction 2b, columns (1)-(3) of Table 8 show that subjects randomly interviewed just prior to the harvest are not significantly more likely to appear "present-biased". However, there is a dramatic reduction of 14.6 percentage points in the proportion who appear "time-consistent". This is driven by the fact that, in contrast to prediction 2b, subjects interviewed prior to the main harvest are 12.9 percentage points *more* likely to appear "future-biased". The heightened level of "future-bias" cannot be straightforwardly explained by subjects anticipating the easing of liquidity constraints after the harvest, as this would make them appear "present-biased". In contrast, it appears that subjects interviewed prior to the harvest are more likely to use the experimental payments as a way to smooth consumption into the future, and thus appear "future-biased".

55 This "pre-post" specification was registered in the pre-analysis plan to allow comparability with the effects of the participation fee timing treatment. It was also not clear ex ante how much income would already increase in the run-up to the main onset of the harvest. Given the smoothness of the increase in income over the survey period shown in Figure 1, I also re-estimate Equation 11 with the day of the baseline survey period (1-48) on the right-hand side. Subjects are on average 2.0 pp more likely to appear "time-consistent" and 2.2pp less likely to appear "future-biased" for each day later into the baseline survey period.

56 The treatment effects also become larger if I drop individuals who never switch to the later payment in one or both frames; results available on request.

Figure 5: Effects of survey timing on day one static inconsistency



Notes: All dependent variables are measured on day one. "Present-biased" ("future-biased") is a dummy indicating a near-frame switch-point greater than (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "future-biased" or "time-consistent". "Pre-harvest" ("post-harvest") indicates a day one session prior to (later than) 25th April 2016.

Table 8: Treatment effects on day one static inconsistency – survey timing

	(1) **"Present-biased" Day one All** Mfx / (s.e.)	(2) **"Time-consistent" Day one All** Mfx / (s.e.)	(3) **"Future-biased" Day one All** Mfx / (s.e.)	(4) **"Present-biased" Day one All** Mfx / (s.e.)	(5) **"Time-consistent" Day one All** Mfx / (s.e.)	(6) **"Future-biased" Day one All** Mfx / (s.e.)
Survey timing						
Pre-harvest	0.016	-0.146***	0.129***	0.045	-0.147**	0.100*
	(0.037)	(0.055)	(0.049)	(0.046)	(0.064)	(0.055)
Pre-harvest*Harvests wheat				-0.083	0.004	0.082
				(0.087)	(0.097)	(0.094)
Harvests wheat	-0.004	-0.004	0.009	0.039	-0.006	-0.036
	(0.047)	(0.053)	(0.049)	(0.068)	(0.078)	(0.076)
Controls	✓	✓	✓	✓	✓	✓
Observations	476	476	476	476	476	476
Control mean	0.244	0.456	0.300	0.244	0.456	0.300

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee treatment status. "Present-biased" ["time-consistent"]("future-biased") is a dummy indicating a near-frame switch-point greater than [equal to] (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "time-consistent" or "future-biased". "Pre-harvest" indicates a day one session prior to 25th April 2016. "Harvests wheat" is a dummy variable equal to one if the subject's household will harvest wheat at this harvest. The sample excludes the five villages which were included at the end of the survey period to boost sample size. Pre-specified controls are: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). Additional controls due to imbalance on survey timing are: trust in NRSP to keep a future appointment; trust in oneself to keep future appointments; and decision-making power within the household. Reported effects represent the marginal effects at the mean. N=476.

A possible explanation, not predicted in the pre-analysis plan, is that subjects interviewed prior to the harvest face more uncertainty in their income from the harvest or other sources.[57] To check for evidence consistent with this story, I re-estimate Equation 11 with the difference between a subject's maximum and minimum expected income over the next four weeks as the dependent variable. Indeed, this measure of income uncertainty is strongly positively correlated with *pre-harvest*: whilst subjects interviewed after the main onset of the harvest report an average expectation range of 3441 PKR, those interviewed prior to the harvest report a range which is 1358 PKR larger.[58] Nonetheless, columns (4)-(6) of Table 8 show that the effects of the increase in "future-bias" is not concentrated among subjects whose households will actually harvest wheat: the point estimates on the interaction term *pre-harvest*harvests wheat* are small and never significant. This suggests that changes in uncertainty over the harvest period affect all members of a village, not just those directly harvesting wheat, for example because the harvest catalyses other forms of economic activity.[59]

Again for completeness, Table 13 in Appendix 2.8.3 shows the effects of the survey timing on individuals' switch-points in the near and the far frame separately. As columns (1) and (2) show, having a baseline interview prior to the harvest makes individuals appear more patient in both frames. Again, this is in contrast with the pre-registered predictions, but may reflect a desire to save until uncertainty about harvest yields is resolved. Columns (3) and (4) of Table 15 show that the pre-harvest treatment has no effect on the near-frame and far-frame switch-points on day fifteen; whilst Table 16 in Appendix 2.8.3 shows that it also does not affect measures of time preference in the day fifteen interviews. This is intuitive, since whilst the pre-harvest treatment guarantees that subjects received their day one interview prior to the harvest, many of these subjects still received their day fifteen interview after harvesting

57 Alternatively or additionally, it may be that the women surveyed prefer to receive experimental income after the harvest because at that time there is lots of cash around and thus plausible deniability about this income is higher.

58 A very similar point estimate is obtained when the dependent variable is the difference between a subject's maximum and minimum expected income from the wheat harvest itself. However, the estimate is not significant, as there are far fewer observations; results available on request.

59 In line with the pre-analysis plan, I also estimate fully saturated models which include both the *pay on day one* and the *pre-harvest* treatments, their interaction, and the interaction of each of these terms with whether the household harvests wheat. There is some evidence that the negative effect of *pay on day one* on measured "present-bias" is concentrated among households who harvest wheat, as the interaction term *pay on day one*harvests wheat* carries a large and negative coefficient. Otherwise the coefficients all become insignificant, likely because the inclusion of so many interaction terms severely weakens power. Results available on request.

had started, and so in the day fifteen interviews they faced similar conditions to subjects who received both interviews after harvesting had begun.[60]

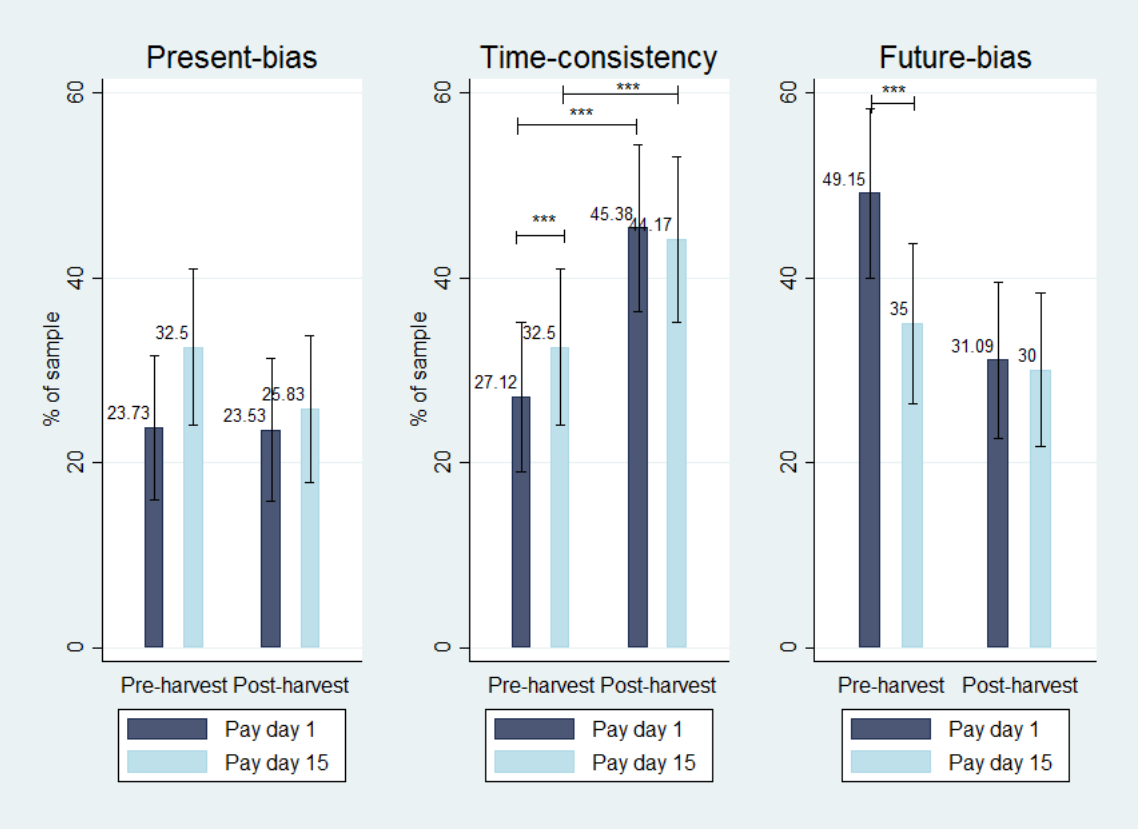
2.5.3 *Interaction of treatments*

Combining the two treatments, Figure 6 shows that the treatment effects of being paid on day one as compared to day fifteen appear to be concentrated among those who are interviewed before the harvest. To test this taking into account possible correlation of standard errors within villages, I re-estimate Equation 11 adding interaction terms between the survey timing treatment and the participation fee timing treatment. Table 9 presents the results. Column (1) in fact finds no evidence in support of experimental prediction 3b, that the effect of receiving the participation fee on day one on "present-bias" should be concentrated among subjects interviewed prior to the harvest: the interaction term is insignificant, and the main treatment effect is not significantly different from the estimate in column (1) of Table 7 (although it loses significance). In contrast, the effect of the participation fee timing on "future-bias" is entirely driven by subjects interviewed prior to the harvest: the coefficient on the interaction term in column (3) implies that subjects who are paid on day one and prior to the harvest are 19.2 percentage points more likely to appear "future-biased" (p-value <0.01). The almost identical but negative coefficient on the interaction term in column (2) suggests that these "future-biased" individuals are those who would otherwise have appeared "time-consistent".

Turning to columns (4)-(6), there is still no significant evidence that being interviewed pre-harvest increases measured "present-bias" (experimental prediction 2b), although again the point estimate carries the predicted sign. The coefficient on the interaction term also carries the predicted sign but is marginally insignificant (p-value 0.101). The large effect of *pre-harvest* on reducing "time-inconsistency" is unchanged compared to Table 8, and the interaction term is insignificant. However, column (6) shows that the spike in measured "future-bias" prior to

60 I also run OLS regressions where the continuous outcome variable is a subject's switch-point in the near and the far frame respectively, on day one and on day fifteen. The treatment effect of *pre-harvest* is never significant; results available on request.

Figure 6: Effects of windfall timing, by survey timing



Notes: All dependent variables are measured on day one. "Present-biased" ("future-biased") is a dummy indicating a near-frame switch-point greater than (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "future-biased" or "time-consistent". "Pre-harvest" ("post-harvest") indicates a day one session prior to (later than) 25th April 2016.

Table 9: Treatment effects – interaction of windfall & survey timing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	"Present-biased"	**"Time-consistent"**	**"Future-biased"**	**"Present-biased"**	**"Time-consistent"**	**"Future-biased"**	**"Present-biased"**	**"Time-consistent"**	**"Future-biased"**
	Day one	**Day one**	**Day one**	**Day one**	**Day one**	**Day one**	**Day one**	**Day one**	**Day one**
	All	**All**	**All**	**All**	**All**	**All**	**All**	**All**	**All**
	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)	Mfx / (s.e.)
Participation fee timing									
Pay on day 1	-0.044	0.065	-0.023				-0.020	0.015	0.007
	(0.046)	(0.066)	(0.059)				(0.051)	(0.072)	(0.070)
Survey timing									
Pre-harvest				0.058	-0.116*	0.058	0.048	-0.109	0.062
				(0.044)	(0.067)	(0.058)	(0.050)	(0.074)	(0.069)
Interaction of timings									
Pay on day 1*Pre-harvest	-0.021	-0.181**	0.192***	-0.088	-0.060	0.140**	-0.067	-0.075	0.133
	(0.055)	(0.075)	(0.068)	(0.054)	(0.069)	(0.064)	(0.075)	(0.101)	(0.096)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	476	476	476	476	476	476	476	476	476
Control mean	0.262	0.441	0.297	0.262	0.441	0.297	0.262	0.441	0.297

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee treatment status. "Present-biased" ("future-biased") is a dummy indicating a near-frame switch-point greater than (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "future-biased" or "time-consistent" (not shown). "Pre-harvest" ("post-harvest") indicates a day one session prior to (later than) 25th April 2016. "Harvests wheat" is a dummy variable equal to one if the subject's household will harvest wheat at this harvest. The sample excludes the five villages which were included at the end of the survey period to boost sample size. Pre-specified controls are: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). Additional controls due to imbalance on survey timing are: trust in NRSP to keep a future appointment; trust in oneself to keep future appointments; and decision-making power within the household. Reported effects represent the marginal effects at the mean. N=476.

the harvest comes entirely from individuals who receive the participation fee on day one. Again, this aligns with the earlier suggestion that "future-bias" here is in fact capturing subjects' attempts to undertake precautionary savings through the experiment — specifically, they are saving the participation fee.[61]

2.5.4 *Heterogeneity by existing liquidity constraints*

The same reasoning behind experimental prediction 3b implies that the effect of participation fee should be stronger for subjects who are in general more liquidity-constrained, as they are less able to smooth the timing of the participation fee themselves. To test for this, Table 10 shows estimations of Equation 10 interacting the participation fee timing treatment with household savings, whether the household would be able to borrow in the next two months, whether the household owns a bank account, and household income.

None of the interaction terms are significant, although lack of power means large effects cannot be ruled out.[62] Nonetheless, the direct effect of these variables is in line with what the model with $\kappa > 0$ would predict: that subjects who are less able to borrow against future income should also in general exhibit higher levels of apparent "present-bias", as they seek to "borrow" through the experiment; and that subjects who have excess liquidity to smooth into the future should exhibit higher levels of "future-bias", as they seek to "save" through the experiment. Specifically, the marginal effect of the control *could not borrow* on appearing "present-biased" is positive, large and significant (34.3 percentage points, p-value 0.033). In contrast, the marginal effect of the control *has a bank account* is negative (9.8 percentage points,

61 An alternative explanation is that subjects are truly present-biased but that receiving a windfall prior to the harvest affects their propensity to become sophisticated and use the experimental time-preference payments as a commitment device. However, descriptive evidence from the baseline survey goes against the idea that "future-bias" is capturing subjects who are present-biased but sophisticated. Neither the static nor the dynamic measures of "future-bias" are correlated with a subject's reports that she finds it hard to save, makes unnecessary purchases, finds it easy to keep track of her money, or uses one or more commitment devices to save. One exception is that both static and dynamic measures of "future-bias" are negatively correlated with subjects reporting that they attempt to use commitment devices to save but fail. Results available on request.

62 The pre-analysis plan specified correcting for multiple hypothesis testing across each of the four interaction terms, since each is a separate proxy for the participation fee treatment interacted with liquidity constraints. When corrected, the interaction terms naturally become even less significant; results available on request.

Table 10: Treatment effects of windfall timing – heterogeneity

	(1) "Present-biased" Day one Mfx / (s.e.)	(2) "Future-biased" Day one Mfx / (s.e.)	(3) "Present-biased" Day one Mfx / (s.e.)	(4) "Future-biased" Day one Mfx / (s.e.)	(5) "Present-biased" Day one Mfx / (s.e.)	(6) "Future-biased" Day one Mfx / (s.e.)	(7) "Present-biased" Day one Mfx / (s.e.)	(8) "Future-biased" Day one Mfx / (s.e.)
Participation fee timing								
Pay on day 1	-0.071	0.050	-0.074*	0.094**	-0.064	0.045	-0.146	0.114
	(0.046)	(0.051)	(0.041)	(0.046)	(0.045)	(0.053)	(0.089)	(0.102)
Liquidity								
Savings (100,000 PKR)	0.015	-0.088*	0.015	-0.051	0.015	-0.054	0.016	-0.053
	(0.036)	(0.049)	(0.030)	(0.034)	(0.030)	(0.034)	(0.030)	(0.034)
Could not borrow	0.343**	-0.282	0.242	0.213	0.341**	-0.276	0.344**	-0.285
	(0.160)	(0.311)	(0.297)	(0.434)	(0.160)	(0.304)	(0.164)	(0.313)
Bank account	-0.098*	-0.009	-0.098*	-0.007	-0.082	-0.092	-0.099*	-0.004
	(0.056)	(0.061)	(0.057)	(0.062)	(0.074)	(0.095)	(0.057)	(0.061)
HH income (100,000 PKR)	-0.016	0.044**	-0.017	0.045**	-0.016	0.043**	-0.030	0.049**
	(0.016)	(0.018)	(0.017)	(0.018)	(0.016)	(0.018)	(0.022)	(0.024)
Heterogeneity								
Pay on day one*Savings	0.000	0.072						
	(0.054)	(0.062)						
Pay on day one*Could not borrow			0.147	0.000				
			(0.332)					
Pay on day one*Bank account					-0.035	0.155		
					(0.109)	(0.118)		
Pay on day one*HH income							0.028	-0.012
							(0.030)	(0.034)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓	✓	✓	✓	✓
Observations	493	504	493	498	493	504	493	504
Control mean	0.291	0.321	0.291	0.321	0.291	0.321	0.291	0.321

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after treatment status is revealed. "Present-biased" ("future-biased") is a dummy indicating a near-frame switch-point greater than (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "future-biased" or "time-consistent" (not shown). Measures of liquidity constraints are taken from the baseline survey. "HH income" and "savings" are denominated in Pakistani Rupees (100 PKR = 1 USD) winsorized at the 95% level. "Could not borrow" is a dummy equal to one if the subject could not borrow from a formal or informal source in the next two months if they needed to. "Bank account" is a dummy variable equal to one if the household has a bank account. Reported effects represent the marginal effects at the mean. N=525. Sample sizes differ where liquidity variables are missing or village fixed effects perfectly predict the outcome variable.

p-value 0.083). Moreover, subjects with higher household income are significantly more likely to appear "future-biased" (4.3 percentage points for every $1,000 USD increase in household income, p-value 0.015).[63]

2.6 ROBUSTNESS

I now consider various alternative possible explanations for these findings, although I find little evidence to support each one. I then show that the effect of the participation fee timing appears to be present regardless of the order of the activities or of the frames within the time preference activities. Finally, I show that the results are robust to various different specifications and treatments of the data.

2.6.1 *Alternative explanations*

TRUST: In theory, receiving the participation fee during the first session might increase a subject's trust that the future-dated payments offered in the time-preference activities will be paid, since the subject sees confirmation that the enumerator is definitely delivering payments. Such an effect could reduce these subjects' propensity to prefer the sooner option in the near frame out of lack of trust that later options will be paid, leaving these subjects who receive the participation fee on day one less likely to appear "present-biased".[64] As outlined in Section 2.3.2, practical steps were taken to minimise trust issues as far as possible. Furthermore, as mentioned in Section 2.4.3, the descriptive statistics suggest that measured "present-bias" is

63 Interestingly, of the other controls being Muslim is by far the biggest predictor of being "future-biased". This might be because Ramadan — a time of increased expenditure on special foods used to break the fast — occurred just after the end of the survey period, leading Muslims in particular to allocate liquidity towards the future. It is also possible that certain teachings or norms in Islam could lead Muslims to make choices that are more patient. However, it is harder to imagine why this would hold only for the near frame and thus generate "future-bias".

64 The reverse confound is also possible — that subjects who are promised that the participation fee will be delivered at the second session become more trusting that the second session is an appointment which is likely to be kept, and thus that payments dated around that time are likely to be delivered.

unlikely to reflect lack of trust in enumerators turning up, since the proportion of subjects who appear statically "present-biased" on day fifteen (after enumerators have already come back once) is if anything slightly *greater* than on day one. Moreover, as discussed earlier, columns (4)-(6) in Table 7 show that the treatment effects appear even stronger when subjects who never switched to a later payment in one or more frames are dropped. One way to interpret this is as a robustness check which drops all subjects who were so distrusting that they refused to choose a later payment no matter the amount.

However, to test directly whether a subject's level of trust that future meetings and payments will take place affects measured "present-bias" in this context, I re-estimate Equation 10 controlling for the two different proxies of trust from the baseline survey: a subject's reported trust that an NRSP representative would keep a future appointment, and trust that she herself would keep a future appointment. The measure of trust in oneself may capture a subject's belief about her own level of personal organisation and reliability. Indeed, it is highly correlated with self-reported measures from the baseline survey that the subject is good at keeping track of her money, and that she uses one or more types of commitment device to save. Alternatively, it may be a euphemistic way for her to report whether she trusts NRSP staff to come back. Indeed, it also has a highly significant (although small) correlation with the number of years that the subject has been an NRSP client.[65] Either way, it should proxy the subject's belief about the probability that a future-dated payment would be successfully delivered to her by an NRSP representative. Yet as columns (1) and (2) in Table 11 show, neither *trust NRSP* nor *trust self* is a significant predictor of appearing "present-biased" or indeed "future-biased".[66]

The lack of evidence that these proxies of trust affect measured "present-bias" does not indicate that they are simply a poor proxy of subjects' beliefs that future payments will take place. To see this, columns (3) and (4) of Table 11 show that an increase in *trust self* significantly reduces a subject's switch-point in both the near and the far frame, thus making her appear

65 *Trust NRSP* is also correlated with the same survey measures of reliability, although less strongly, and its correlation with years as an NRSP client is marginally insignificant.

66 To test whether the effect of the participation fee timing is different across those who report high or low trust at baseline, I also estimate alternative specifications including interaction terms between each of the proxies of trust and the participation fee timing treatment. The point estimates on these interaction terms are insignificant, although power is severely weakened by inclusion of the interaction terms; results available on request.

Table 11: Treatment effects of windfall timing – trust controls

	(1)	(2)	(3)	(4)
	"Present-biased"	**"Future-biased"**	**Near-frame**	**Far-frame**
	Day one	**Day one**	**switch-point**	**switch-point**
	All	**All**	**Day one**	**Day one**
			All	**All**
	Mfx / (s.e.)	Mfx / (s.e.)	OLS	OLS
Participation fee timing				
Pay on day 1	-0.070*	0.078*	-12.810	-3.444
	(0.041)	(0.046)	(20.069)	(19.483)
Trust				
Trust NRSP (1-5)	0.011	0.022	-5.811	-8.068
	(0.037)	(0.039)	(16.506)	(17.946)
Trust self (1-5)	-0.022	0.030	-79.506***	-72.607***
	(0.048)	(0.053)	(23.872)	(22.849)
Controls	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓
Observations	493	504	523	523
Control mean	0.291	0.321	680.189	684.792

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee timing treatment status. "Present-biased" ("future-biased") is a dummy indicating a near-frame switch-point greater than (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "future-biased" or "time-consistent" (not shown). Switch-points are in Pakistani Rupees (100 PKR $\approx$ 1 USD). "Trust NRSP" and "Trust self" are 1-5 Likert-scale responses ("how strongly do you agree or disagree with the following statement?") to the following statements: "if a female representative of NRSP made an appointment to see me about a different study, they would be unlikely to cancel or change that appointment"; "if I made an appointment to see someone, for example a female representative of NRSP involved in a different study, I would be unlikely to cancel or change that appointment". Pre-specified controls are: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). Reported effects in columns (1) and (2) represent the marginal effects at the mean. N=525. Sample sizes in columns (1) and (2) differ where village fixed effects perfectly predict the outcome variable.

significantly more "patient" in the near and the far frame.[67] The results are therefore in line with concerns that subjects who perceive a lower probability of future payments taking place may more "impatient".[68] However, the reason that this has no net effect on "present-bias" or "future-bias" is that the effect of trust on increasing "patience" is virtually identical across the near and the far frame.[69]

ARBITRAGE: Another possible explanation for why giving subjects a windfall or interviewing them after the harvest decreases measured "present-bias" is that under these treatment conditions subjects already have enough liquidity to fund their desired level of current consumption. If so, experimental payments dated "now" do not translate into consumption "now", and so the monetary tasks become unable to measure true present-bias (O'Donoghue and Rabin, 2015). Instead, individuals should chose time-dated monetary payments to maximise experimental earnings, which in this context means that they should choose any later payment greater than 400 Rs (i.e. any positive return) and should do so across both frames.[70] However, the data do not support this: the median subject who receives the participation fee on day one still prefers an earlier payment of 400 Rs rather than a later payment of up to 550 Rs in the near frame, and up to 600 Rs in the far frame. Similarly, the median subject interviewed after the harvest prefers an earlier payment of 400 Rs rather than a later payment of up to 600 Rs in the near frame, and up to 650 Rs in the far frame. Thus even those subjects with liquidity from the participation fee and the harvest still behave very impatiently and as if their consumption must track the timing of experimental payments – whether rationally,

67 Near-frame coefficient -79.5 PKR p-value 0.001, FDR-adjusted q-value 0.002 taken across the two proxies of trust (Anderson, 2012). Far-frame coefficient -72.6 PKR, p-value 0.01, FDR-adjusted q-value 0.002 taken across the two proxies of trust.

68 Going beyond the pre-analysis plan, I also re-estimate Equation 10 controlling for the number of years that a subject has been an NRSP microfinance client, as an alternative proxy of trust. Just like *trust NRSP* and *trust self*, the number of years as an NRSP client is not significantly correlated with "present-bias" or "future-bias", nor is its interaction with the treatment significant. Its correlation with the near-frame and far-frame switch-points is positive but small and marginally insignificant. Results available on request.

69 This suggests that subjects may weight the subjective probability of experimenter default similarly across the near and far frame, rather than exhibiting a "certainty effect" or broader convex probability weighting (i.e. overweighting the decline between today and the later date in the near frame compared to the decline between the earlier date and the later date in the far frame).

70 As mentioned earlier, no subject reports receiving a financial return on any form of savings other than livestock. It is therefore safe to assume that saving a marginal 400 Rupess outside the experiment always produces a return inferior to positive returns offered within the experiment.

because their demands for liquidity now are still not met; or less rationally, because the asset integration in their decision-making is only partial ($0 < \kappa < 1$).[71]

COGNITIVE FUNCTIONING: Another possible concern is that receiving the participation fee on day one, or being interviewed after the harvest, might lead subjects to make more consistent choices by reducing financial stress, thereby freeing up mental bandwidth and leading to fewer errors (Banerjee and Mullainathan, 2008; Mani et al., 2013). To test for this, I re-estimate Equations 10 and 11 controlling for a subject's score on the digit span test, maths test, and time taken to complete the numerical Stroop test. Like the time preference activities, these cognitive tests were elicited after the revelation of the participation fee timing, thus including these controls captures any simultaneous effects that the treatments may have had on subjects' cognitive functioning. Columns (1) and (2) of Table 17 in Appendix 2.8.4 report the results for the participation fee treatment, and columns (1) and (2) of Table 18 in Appendix 2.8.4 report the results for the survey timing treatment. The coefficients on *pay on day one* and *pre-harvest* appear unchanged and are never significantly different from the main specifications; although the coefficient on *pay on day one* at times becomes marginally insignificant since the inclusion of additional controls again weaken power.[72]

RISK PREFERENCES: Receiving the participation fee on day one could also in theory move a subject to a less curved part of her within-period utility function. If she projects this onto her future utility function (via so-called "projection bias") this might increase her propensity to take the later payment in the time-preference activities if the later payment is seen as risky. To check for this, I re-estimate Equations 10 and 11 adding a proxy for subjects' risk preferences, which were also elicited after the revelation of the participation fee timing. Specifically, the included covariate is a measure of a subject's certainty premium

71 O'Donoghue and Rabin (2015) claim that, by calibration arguments, evidence of very high impatience over the short term should itself by taken as evidence of present-bias. I remain agnostic on this point.

72 Interestingly, mathematical performance itself is strongly negatively correlated with "present-bias" and strongly positively correlated with "future-bias". This may be in line with Dohmen et al. (2010), who find that subjects with higher cognitive ability are more patient. It may also reflect higher-ability subjects better understanding the opportunity to save through the experiment.

(denominated in Rupees), aggregated across five sets of certainty-equivalent questions which involved different probabilities. Columns (3) and (4) of Tables 17 and 18 shows that the relationship between the risk premium and measured time preferences is zero or very small, and that including risk controls does not significantly change the main treatment effects.

OPTIMISM: Similarly, it could be that receiving the participation fee on day one makes subjects feel more optimistic that uncertain future events will go in their favour, which may change their responses to time-preference questions aside from the pure effect of easing liquidity constraints.[73] To test for this, I also re-estimate Equations 10 and 11 adding an extra control proxying subjects' optimism. Optimism is measured via a probability task, which was also administered after subjects learned whether they would receive the participation fee on day one or day fifteen. Specifically, *optimism* measures the difference between a subject's stated belief of her own probability of winning a draw and the objective, given probability of winning that draw, aggregated across five draws with different probabilities. As columns (5) and (6) or Tables 17 and 18 show, the coefficient on *optimism* is a tightly-estimated zero, and including this measure does not significantly change the main treatment effects.

2.6.2 *Salience and frame order effects*

To test whether the salience of liquidity constraints matters for subjects' responses to the time preference activities, I re-estimate Equation 10 adding in a dummy for whether subjects were subject to the salience treatment (*time preferences first*). Similarly, to test whether subjects' responses are sensitive to whether the near or the farm frame is elicited first, I also add controls for the randomised frame order within the MPL (*near frame first*) and indeed the interaction the frame order with the salience treatment. Table 12 displays the results. The coefficients on *pay on day 1* are not significantly different from columns (1)-(3) of Table 7 and retain their

73 Subjects' mood could also impact their discount rates directly, for example Haushofer and Fehr (2014) show that experimentally-induced negative affect states increase discount rates.

significance. The salience treatment has no effect on measures of "time-inconsistency". In contrast, frame order effects are large: subjects are 17.5 percentage points less likely to appear "present-biased" if the near frame is elicited first (p-value 0.008, FDR-adjusted q-value 0.012 taken across the three outcome variables in Table 12) and 12.3 percentage points more likely to appear "future-biased", although this is only borderline significant (p-value 0.059, FDR-adjusted q-value 0.128). It is possible that this is capturing a learning effect, or an effect in which subjects become more patient as tasks progress (Andersen et al., 2006).[74] However, these effects also persist in the multiple price lists elicited on day fifteen (results available on request), making a learning story less plausible. An alternative explanation is that subjects become more impatient about payments in the far frame when answering the far frame second. Evidence on the switch-points from the subject frames is inconclusive: point estimates suggest that on day one, answering the near frame first leads to both a small reduction in the near-frame switch-point — i.e. increased patience in the near frame — and an increase in the far-frame switch-point — i.e. decreased patience in the far frame — but neither estimate is significant; results available on request.

2.6.3 *Alternative specifications*

I also estimate various alternative specifications of Equation 10, to test for the sensitivity of the results to treatment of the data. First, to exclude any possible effects of making responses unincentivized, I re-estimate Equation 10 but this time drop any subjects who had already randomly drawn to be paid for their answers to the activities on day one. Results are identical to those for the main sample in columns (1)-(3) of Table 7; tables available on request. Second, as a maximally conservative way of excluding subjects who may have misunderstood the task or made a small mistake at some point, I re-estimate Equation 10 but this time drop all subjects who exhibit multiple-switching in any of the frames on either date. Point estimates

74 Andersen et al. (2006) randomise frame order and find that discount rates increase modestly by 3.5 pp between the first and the second frame but not between the second and third frame, although they are considering frames of different time lengths rather than frames covering near and far dates.

Table 12: Estimated treatment effects – activity & frame order

	(1) **"Present-biased"** **Day one** **All** Mfx / (s.e.)	(2) **"Time-consistent"** **Day one** **All** Mfx / (s.e.)	(3) **"Future-biased"** **Day one** **All** Mfx / (s.e.)
Participation fee timing			
Pay on day 1	-0.070*	-0.009	0.080*
	(0.041)	(0.048)	(0.046)
Activity & frame order			
Time prefs first	-0.051	0.080	-0.010
	(0.058)	(0.067)	(0.070)
Near frame first	-0.175***	0.072	0.123*
	(0.062)	(0.068)	(0.065)
Time prefs first*Near frame first	0.125	-0.146	-0.000
	(0.088)	(0.099)	(0.097)
Controls	✓	✓	✓
Village f.e.'s	✓	✓	✓
Observations	493	503	504
Control mean	0.321	0.410	0.269

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee, activity order and frame order treatment status. "Present-biased" ["time-consistent"]("future-biased") is a dummy indicating a near-frame switch-point greater than [equal to] (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "time-consistent" or "future-biased". Whether the subject is married is added as an additional control, due to imbalance on the salience treatment. Reported effects represent the marginal effects at the mean. N=525. Sample sizes differ where village fixed effects perfectly predict the outcome variable.

for the treatment effects are virtually unchanged, but become marginally insignificant due to the reduction in sample size; tables available on request. Third, as well as dropping those subjects who never switch to the later payment in one or more frames, I also drop those who always choose the later payment, i.e. who already "switch" in the first question. Point estimates for the effect of the participation fee timing are virtually identical to the full sample, and remain highly significant for "future-bias" but become insignificant for "present-bias"; results available on request. Fourth, to allow for the possibility that subjects may be indifferent rather than having a strict preference at the point where they "switch", I re-set the dummies for "present-bias" and "future-bias" to equal to one only if the subject's responses differ

by more than one question across the two frames concerned. I then re-run the estimations of 10 with these more conservative dummy variables as the dependent variables. The point estimates for the treatment effects of the participation fee timing are very similar to the initial results. However, they are no longer significant, since multiple villages are now dropped from the logit estimations with village fixed effects due to lack of variation in the new dependent variable; results available on request.

2.7 CONCLUSION

This paper provides the first causal, experimental evidence that monetary measures of poor individuals' apparent "present-bias" — and indeed "future-bias" — are highly sensitive to their liquidity constraints. Thus an assumption of "narrow framing", required for choices over time-dated monetary payments to identify subjects' time preferences when subjects' ability to smooth consumption varies over time, appears to be violated. The findings therefore add to the growing scepticism over the use of monetary tasks to elicit time preferences.

Recent literature has proposed "real effort tasks" as an alternative time-preference elicitation method, in order to mitigate concerns about both arbitrage and trust (Augenblick et al., 2015).[75] However, such tasks may still suffer from a bias similar to that described in this paper, if subjects face time constraints which vary over time.[76] For example, an individual may be very busy now and thus prefer to defer the completion of effort tasks to some other point in the near frame; but she may anticipate being much less time-constrained in the more distant future, and so may be less inclined to defer effort tasks in the far frame. Such an individual would therefore spuriously appear "present-biased" in effort tasks. Individuals may also be optimistic about their future time-constraints, just as they may be optimistic about future liquidity constraints, and this would again bias real effort tasks in the direction of finding

75 Tasks using consumption goods have also been proposed (Read and Van Leeuwen, 1998). However, as Ubfal (2016) shows in Uganda, subjects may also able to arbitrage these tasks if they have access to trading opportunities or stockpiles of the good at home.

76 Real effort tasks may suffer from issues of generalisability, for example if preferences over real-world leisure or other forms of real-world consumption are different to preferences over consumption of "leisure time" in the lab.

"present-bias". It therefore remains an empirical question as to whether any bias in real effort tasks caused by variation in subjects' time constraints is greater than the bias in monetary tasks caused by variation in subjects' liquidity constraints.[77] More research is needed on this matter.

If researchers continue to use monetary tasks to elicit time preferences, then the results presented here suggest that researchers should exercise particular caution in certain contexts. In particular, the common practices of conducting developing-country surveys in the lean season, and paying out participation fees at the end of the study, are both likely to bias results towards finding less "time-consistency" and more "present-bias". Researchers should therefore be conscious of these issues when planning the timing of future surveys.

In terms of policy implications, the findings suggest a need for caution in offering commitment savings products if their purported goal is to help individuals to overcome their present-bias. Individual targeting on the basis of measured time preferences may be highly inaccurate, and the true extent of present-biased financial decision-making in the population may be lower than estimates suggest. Of course, lack of present-bias in financial decision-making at the individual level does not mean that individuals may not demand commitment for other reasons. Most saliently, individuals may be willing to pay for commitment, or to forego interest, in order to hide money from their spouse or relatives (Brune et al., 2011; Schaner, 2011; Schaner, 2015). Indeed, a household may still act as if it is present-biased even if its individual members are time-consistent, as long as members' discount rates differ (Adams et al., 2014). However, if intra-household bargaining or more general "other-control" issues are the source of demand for commitment, then a very different type of financial product might provide the solution. For example, products may be designed to create privacy, or to label funds for specific purposes such as school fees, without the need for hard commitments that tie up liquidity. This remains a key area for future research.

77 Carvalho et al. (2016) find that the timing of the interview in relation to pay-day does not affect responses in real effort tasks. However, whilst in their context time constraints are unlikely to vary much before and after pay-day, time constraints may vary much more starkly in contexts such as before and after the onset of the harvest.

2.8 APPENDICES

2.8.1 *Proof of predicted effects of windfalls*

Recall that a static choice reversal in the direction of "present-bias" ($x_{1.0} > x_{2.0}$) will occur if Equation 7 holds, i.e. if:

$$\frac{u(\kappa c_0 + x) - u(\kappa c_0)}{\beta E_0 \left[V_1(\kappa w_1 + \mathbf{x_{2.0}}) - V_1(\kappa w_1)\right]} > E_0 \left[\frac{V_1(\kappa w_1 + x) - V_1(\kappa w_1)}{V_2(w_2 + x_{2.0}) - V_2(w_2)}\right].$$

For x and $x_{2.0}$ sufficiently small, for $\kappa > 0$ this reduces to the condition that:

$$\frac{u'(\kappa c_0)}{\beta E_0 V_1'(\kappa w_1)} > E_0 \left[\frac{V_1'(\kappa w_1)}{V_2'(\kappa w_2)}\right]. \tag{12}$$

Conversely, a static choice reversal in the direction of "future-bias" will occur if:

$$\frac{u'(\kappa c_0)}{\beta E_0 V_1'(\kappa w_1)} < E_0 \left[\frac{V_1'(\kappa w_1)}{V_2'(\kappa w_2)}\right]. \tag{13}$$

PROOF OF PROPOSITION 1. If a subject receives positive and unanticipated windfall income Z at $t = 0$, a static choice reversal in the direction of "present-bias" will now hold on the condition that:

$$\frac{u'(\kappa(c_0 + Z))}{\beta E_0 V_1'(\kappa w_1)} > E_0 \left[\frac{V_1'(\kappa w_1)}{V_2'(\kappa w_2)}\right]. \tag{12'}$$

By the concavity of $u(.)$, condition 12′ is less likely to hold than condition 12 for $Z > 0$ since the numerator of the left-hand side has decreased. Thus the subject is less likely to exhibit static choice reversals in the direction of "present-bias".

On the other hand, a static reversal in the direction of "future-bias" will now occur on the condition that:

$$\frac{u'(\kappa(c_0 + Z))}{\beta E_0 V_1'(\kappa w_1)} < E_0 \left[\frac{V_1'(\kappa w_1)}{V_2'(\kappa w_2)} \right]. \tag{13'}$$

Again, by the concavity of $u(.)$, Equation 13' is more likely to hold than Equation 13 for $Z > 0$, since the numerator of the left-hand side has decreased. *QED.*

PROOF OF PROPOSITION 2. If a subject is told at $t = 0$ that she will receive positive windfall income Z at $t = 1$, a static reversal in the direction of "present-bias" will now occur if:

$$\frac{u'(\kappa c_0)}{\beta E_0 V_1'(\kappa(w_1 + Z))} > E_0 \left[\frac{V_1'(\kappa(w_1 + Z))}{V_2'(\kappa w_2')} \right], \tag{12''}$$

where w_2' may differ from w_2 in Equation 12 if the subject chooses to adjust period-1 savings in response to the windfall.[78] By the concavity of $u(.)$, condition 12'' is more likely to hold than Equation 12 since the denominator of the left-hand side has decreased.[79]

Conversely, a static reversal in the direction of "future-bias" will occur if:

$$\frac{u'(\kappa c_0)}{\beta E_0 V_1'(\kappa(w_1 + Z))} < E_0 \left[\frac{V_1'(\kappa(w_1 + Z))}{V_2'(\kappa w_2')} \right]. \tag{13''}$$

This condition is less likely to hold than condition 13, since again by the concavity of $u(.)$, the denominator of the left-hand side has decreased. *QED.*

78 The assumption is only that current $t = 0$ savings are predetermined with respect to experimental income.

79 The numerator of the right-hand side has also increased; although the denominator of the right-hand side could also have increased identically to offset this, if the subject perfectly smooths the windfall across periods 1 and 2 via saving.

2.8.2 *Experimental materials*

Figure 7: Structure of experimental sessions

<table>
<tr><td colspan="2">1. Baseline survey, day one session only
2. Income and expenditure survey</td></tr>
<tr><td colspan="2">3. Participation fee explanation & payment (payment on day 1 vs. day 15)
4. Explanation of activity incentivization</td></tr>
<tr><td>Activities – “salience” order

5. Time preferences (near vs. far first)

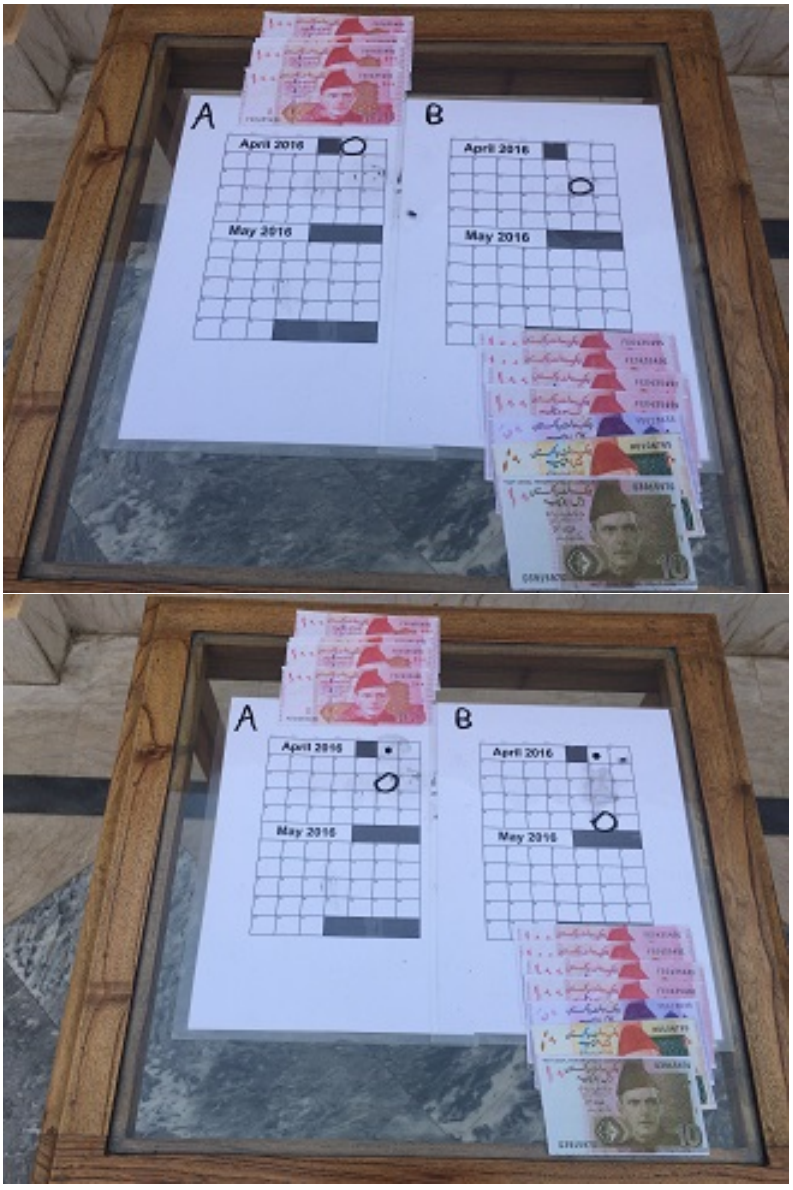
6. Control activities</td><td>Activities – “control” order

5. Control activities

6. Time preferences (near vs. far first)</td></tr>
<tr><td colspan="2">7. Activity payments</td></tr>
</table>

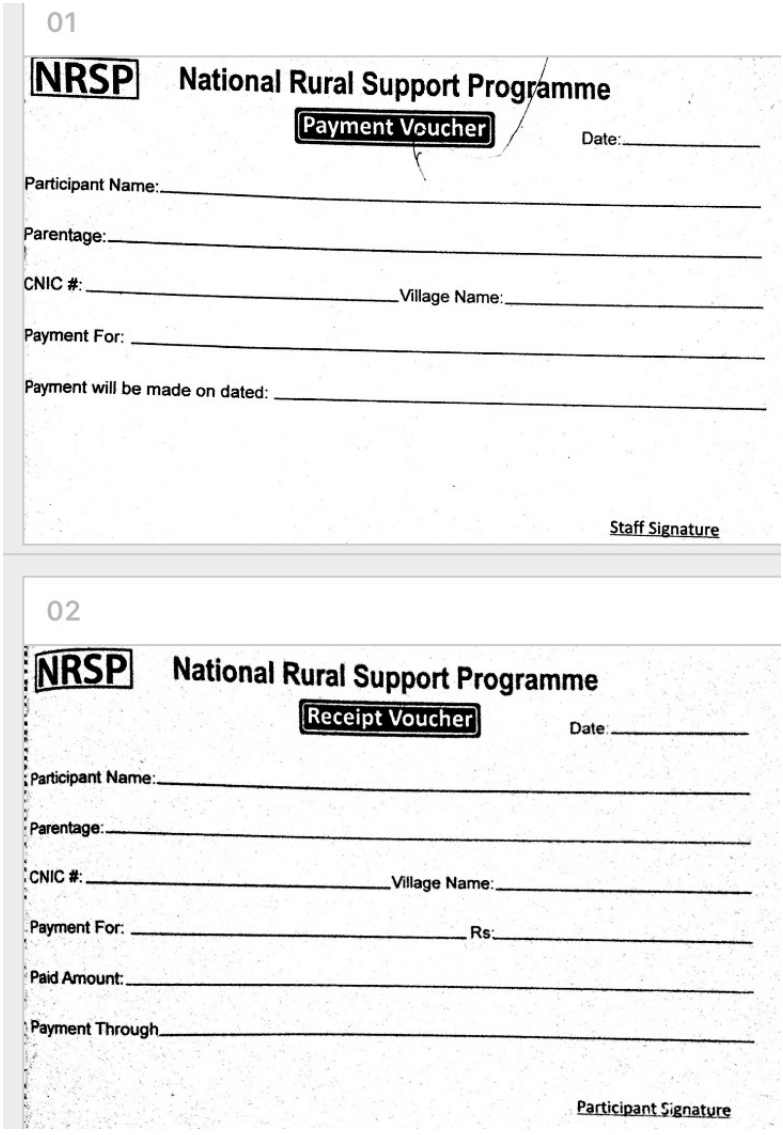
Notes: The ordering of items in italics was subject to randomisation. Time preference activities consisted of multiple price lists. Cognitive tasks comprised a maths test, digit span test, and numerical Stroop test. Risk preference activities consisted of certainty equivalents. Optimism was measured by a novel task eliciting subjective probabilities over the probability of winning a given lottery with set objective probabilities. The respondent was asked how likely she thought it was that an average person would win that lottery, and how likely it was that she herself would win.

Figure 8: Multiple price list demonstrations



Notes: The top panel illustrates the near frame, and the bottom panel illustrates the far frame.

Figure 9: Payment vouchers



Notes: The top panel displays the vouchers for promise of future payment. The bottom panel shows the vouchers for receipt of payment. These vouchers were used for participation fees and for payments in the time preference and risk preference activities.

Table 13: Treatment effects on switch-points – windfall timing

	(1) **Near frame Day one All** OLS	(2) **Far frame Day one All** OLS	(3) **Near frame Day fifteen All** OLS	(4) **Far frame Day fifteen All** OLS
Participation fee timing				
Pay on day 1	-16.86	-7.29	-5.44	-12.45
	(20.34)	(19.70)	(20.98)	(20.36)
Controls	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓
Observations	523	523	520	519
Control mean	680.2	684.8	753.9	738.2

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee treatment status. Switch-points are in Pakistani Rupees (100 PKR ≈ 1 USD). In each frame, subjects were asked to choose between 400 PKR on the earlier date or the amount shown on the later date. The switch-point in each frame is equal to the first value at which the subject chose to receive the payment on the later date. In the near frame, the earlier date was today and the later date was two weeks from today. In the far frame, the earlier date was two weeks from today and the later date was four weeks from today. Pre-specified controls are: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). N=525. Sample sizes differ where village fixed effects perfectly predict the outcome variable.

Table 14: Treatment effects on day fifteen inconsistencies – windfall timing

	(1) **"Present-biased" revision All** Mfx / (s.e.)	(2) **"Time-consistent" no revision All** Mfx / (s.e.)	(3) **"Future-biased" revision All** Mfx / (s.e.)	(4) **"Present-biased" Day fifteen All** Mfx / (s.e.)	(5) **"Time-consistent" Day fifteen All** Mfx / (s.e.)	(6) **"Future-biased" Day fifteen All** Mfx / (s.e.)
Participation fee timing						
Pay on day 1	-0.002	-0.039	0.039	-0.001	-0.046	0.052
	(0.048)	(0.039)	(0.042)	(0.043)	(0.048)	(0.041)
Controls	✓	✓	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓	✓	✓
Observations	510	435	510	480	520	461
Control mean	0.494	0.205	0.300	0.274	0.513	0.213

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee treatment status. In columns (1)-(3), "present-biased" ["time-consistent" ("future-biased") is a dummy indicating a near-frame switch-point on day fifteen greater than [equal to] (less than) the far-frame switch-point on day one. In columns (4)-(6), "present-biased" ["time-consistent"]("future-biased") is a dummy indicating a near-frame switch-point on day fifteen greater than [equal to] (less than) the far-frame switch-point on day fifteen. In each case, subjects are classified as either "present-biased", "time-consistent" or "future-biased". Pre-specified controls are: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). Reported effects represent the marginal effects at the mean. N=525. Sample sizes differ where village fixed effects perfectly predict the outcome variable.

Table 15: Treatment effects on switch-points – survey timing

	(1)	(2)	(3)	(4)
	Near frame	**Far frame**	**Near frame**	**Far frame**
	Day one	**Day one**	**Day fifteen**	**Day fifteen**
	All	**All**	**All**	**All**
	OLS	OLS	OLS	OLS
Survey timing				
Pre-harvest	-69.48**	-84.40***	-16.37	-35.88
	(27.07)	(26.03)	(24.11)	(23.15)
Controls	✓	✓	✓	✓
Observations	476	476	474	473
Control mean	694.6	716.2	761.3	751.1

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee treatment status. Switch-points are in Pakistani Rupees (100PKR $\approx$ 1 USD). In each frame, subjects were asked to choose between 400 PKR on the earlier date or the amount shown on the later date. The switch-point in each frame is equal to the first value at which the subject chose to receive the payment on the later date. In the near frame, the earlier date was today and the later date was two weeks from today. In the far frame, the earlier date was two weeks from today and the later date was four weeks from today. Pre-specified controls are: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). Additional controls due to imbalance on survey timing are: trust in NRSP to keep a future appointment; trust in oneself to keep future appointments; and decision-making power within the household. The sample excludes the five additional villages which were included to boost power in the windfall timing experiment. N=476. Sample sizes differ where village fixed effects perfectly predict the outcome variable.

Table 16: Treatment effects on day fifteen inconsistencies – survey timing

	(1) **"Present-biased" revision All** Mfx / (s.e.)	(2) **"Time-consistent" no revision All** Mfx / (s.e.)	(3) **"Future-biased" revision All** Mfx / (s.e.)	(4) **"Present-biased" Day fifteen All** Mfx / (s.e.)	(5) **"Time-consistent" Day fifteen All** Mfx / (s.e.)	(6) **"Future-biased" Day fifteen All** Mfx / (s.e.)
Survey timing						
Pre-harvest	0.068	-0.035	-0.032	0.061	0.012	-0.070
	(0.056)	(0.037)	(0.047)	(0.052)	(0.052)	(0.046)
Controls	✓	✓	✓	✓	✓	✓
Observations	474	474	474	474	474	474
Control mean	0.462	0.220	0.318	0.241	0.514	0.245

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee treatment status. In columns (1)-(3), "present-biased" ["time-consistent" ("future-biased") is a dummy indicating a near-frame switch-point on day fifteen greater than [equal to] (less than) the far-frame switch-point on day one. In columns (4)-(6), "present-biased" ["time-consistent"]("future-biased") is a dummy indicating a near-frame switch-point on day fifteen greater than [equal to] (less than) the far-frame switch-point on day fifteen. In each case, subjects are classified as either "present-biased", "time-consistent" or "future-biased". "Pre-harvest" indicates a day one session prior to 25th April 2016. The sample excludes the five additional villages which were included at the end of the survey period to boost sample size. The sign and significance of estimated coefficients remains unchanged when these villages are dropped. Pre-specified controls are: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). Additional controls due to imbalance on survey timing are: trust in NRSP to keep a future appointment; trust in oneself to keep future appointments; and decision-making power within the household. Reported effects represent the marginal effects at the mean. N=476. Sample sizes differ where village fixed effects perfectly predict the outcome variable.

2.8.4 *Additional robustness checks*

Table 17: Treatment effects of windfall timing – additional controls

	(1) **"Present-biased" Day one** Mfx / (s.e.)	(2) **"Future-biased" Day one** Mfx / (s.e.)	(3) **"Present-biased" Day one** Mfx / (s.e.)	(4) **"Future-biased" Day one** Mfx / (s.e.)	(5) **"Present-biased" Day one** Mfx / (s.e.)	(6) **"Future-biased" Day one** Mfx / (s.e.)
Participation fee timing						
Pay on day 1	-0.058	0.076	-0.071*	0.081*	-0.071*	0.081*
	(0.041)	(0.047)	(0.041)	(0.046)	(0.041)	(0.046)
Cognitive functioning						
Digit span 1 (score 1-7)	0.021	0.003				
	(0.025)	(0.026)				
Maths 1 (score 1-8)	-0.023***	0.030***				
	(0.008)	(0.009)				
Stroop time 1	-0.001	0.001				
	(0.001)	(0.001)				
Risk preferences						
Certainty premium 1 (100 PKR)			-0.005	0.002		
			(0.004)	(0.004)		
Optimism						
Probablity optimism 1					-0.000	0.000
					(0.000)	(0.000)
Controls	✓	✓	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓	✓	✓
Observations	488	499	493	504	493	504
Control mean	0.291	0.321	0.291	0.321	0.291	0.321

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee timing treatment status, as are cognitive functioning, risk preferences and optimism. "Present-biased" ("future-biased") is a dummy indicating a near-frame switch-point greater than (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "future-biased" or "time-consistent" (not shown). "Certainty premium" is a PKR value aggregated across five sets of certainty-equivalent questions which involved different probabilities. "Optimism" is the difference between a subject's subjective belief of her own probability of winning a draw and the objective, given probability of winning that draw, aggregated across five draws with different probabilities. Pre-specified controls are: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). Reported effects represent the marginal effects at the mean. N=525. Sample sizes differ where village fixed effects perfectly predict the outcome variable.

Table 18: Treatment effects of survey timing – additional controls

	(1) **"Present-biased" Day one** Mfx / (s.e.)	(2) **"Future-biased" Day one** Mfx / (s.e.)	(3) **"Present-biased" Day one** Mfx / (s.e.)	(4) **"Future-biased" Day one** Mfx / (s.e.)	(5) **"Present-biased" Day one** Mfx / (s.e.)	(6) **"Future-biased" Day one** Mfx / (s.e.)
Survey timing						
Pre-harvest	0.020	0.126	0.013	0.145	0.018	0.129
	(0.036)	(0.051)	(0.038)	(0.050)	(0.038)	(0.049)
Cognitive functioning						
Digit span 1 (score 1-7)	0.008	0.008				
	(0.023)	(0.022)				
Maths 1 (score 1-8)	-0.017	0.017				
	(0.009)	(0.009)				
Stroop time 1	-0.002	0.001				
	(0.001)	(0.001)				
Risk preferences						
Certainty premium 1 (100 PKR)			-0.002	0.007		
			(0.003)	(0.004)		
Optimism						
Probablity optimism 1					-0.000	0.000
					(0.000)	(0.000)
Controls	✓	✓	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓	✓	✓
Observations	471	471	476	476	476	476
Control mean	0.244	0.300	0.244	0.300	0.244	0.300

Notes: Standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All dependent variables are measured after revelation of participation fee timing treatment status, as are cognitive functioning, risk preferences and optimism. "Present-biased" ("future-biased") is a dummy indicating a near-frame switch-point greater than (less than) the far-frame switch-point in the multiple price list activity. Subjects are classified as either "present-biased", "future-biased" or "time-consistent" (not shown). "Pre-harvest" indicates a day one session prior to 25th April 2016. "Certainty premium" is a PKR value aggregated across five sets of certainty-equivalent questions which involved different probabilities. "Optimism" is the difference between a subject's subjective belief of her own probability of winning a draw and the objective, given probability of winning that draw, aggregated across five draws with different probabilities. Pre-specified controls are: household income, savings, and possession of a bank account; ability to borrow in the next two months (formally or informally) if needed; whether the household harvests wheat; household religion; respondent's education; and her occupation (housewife or other). Reported effects represent the marginal effects at the mean. N=476. Sample sizes differ where village fixed effects perfectly predict the outcome variable.

3

THE POWER TO PROTECT: HOUSEHOLD BARGAINING AND FEMALE CONDOM USE

with Marije Groot Bruinderink, Wendy Janssens, and Karlijn Morsink

3.1 INTRODUCTION

Partners may differ in their willingness to adopt technologies that increase household welfare. In particular, there is evidence that women have a stronger preference for risk reduction (Schubert et al., 1999; Agnew et al., 2008), and for household public goods such as children's health and education (Duflo, 2003). Women may also bear more of the costs of non-adoption, through responsibility for domestic chores and caring duties, or greater exposure to certain health risks (Miller and Mobarak, 2013). The extent to which households invest in welfare and adopt such technologies may therefore depend on women's bargaining power and control over household resources (Attanasio and Lechene, 2014; Duflo and Udry, 2004), as well as whether technologies can be adopted unilaterally or require the agreement of both partners (Ashraf et al., 2014a).

Contraceptive technologies are a key example. The only well-established dual protection technologies — i.e. contraceptives that also protect against HIV/AIDS and other STIs — are condoms. By their nature condoms are not concealable, and thus their use requires the agreement of both partners. Thus if women have a stronger demand for condom adoption than men, then women with low bargaining power may struggle to convince their partners to

adopt. In contrast, contraceptive technologies that protect solely against pregnancy — such as the pill and injectables — can be used by women alone, and are often sufficiently concealable such that women may adopt them unilaterally.[1] A large economic literature has documented the consequent impact of the pill on women's economic empowerment (Bailey, 2006; Chiappori and Oreffice, 2008; Goldin and Katz, 2002).

There are several reasons to believe that women have a higher willingness than men to adopt condoms. First and foremost, women are more vulnerable to HIV infection: in 2015, women accounted for 59% of all individuals aged 15 and over living with HIV, and the rate of new infections among young women aged 15-24 was double that among young men (UNAIDS, 2016b). The reason for this gender disparity is that women tend to have older partners, lower access to sexual and reproductive health services, and a higher biological risk than men of becoming infected from heterosexual intercourse (UNAIDS, 2016a). Furthermore, women may also be more risk-averse than men (Eckel and Grossman, 2008), and disproportionately bear the costs of unwanted pregnancy, again increasing their demand for dual protection. Meanwhile, the sexual health literature has highlighted how men in particular find male condoms uncomfortable (Crosby et al., 2004), associate them with stigma or loss of "masculinity" (Koster et al., 2015), or perceive them as unnecessary out of a belief that female partners are less risky (Flood, 2003).[2] Women with low bargaining power may therefore face pressure or even violence from their male partners if they do not agree to unprotected sex (Holland et al., 1991; Kalichman et al., 1998; MacPhail and Campbell, 2001; Go et al., 2003; Dunkle et al., 2004; Silverman et al., 2008). This may partly explain why condoms are subject to chronic under-adoption from the perspective of policymakers: in 2015 alone, an estimated 3.3 billion risky sex acts took place without condoms in Sub-Saharan Africa, leading to 910,000 new HIV infections (UNAIDS, 2016b).[3]

1 Similar points have been made about women adopting ROSCAs (Anderson and Baland, 2002), microfinance (Van Tassel, 2004), and individual savings accounts (Schaner, 2015).

2 Both men and women may also dislike male condoms insofar as they carry connotations of mistrust, or of the relationship not being serious (Tavory and Swidler, 2009).

3 The total estimated need is an estimate of the number of sex acts between sex workers and their clients (38%), non-regular partners (20%), partners requiring condoms as a method of family planning (15%), HIV-affected couples (14%), and men who have sex with men (8%). Of the 2.7 billion condoms which were used, female condoms accounted for just 1.6%.

Female condoms are an alternative dual protection technology to male condoms, with a similar design. Like male condoms, female condoms are also observable and require agreement by both partners for use.[4] Yet qualitative evidence suggests that men associate female condoms with lower stigma and disutility than male condoms (Wanyenze et al., 2011; Koster et al., 2015). However, female condoms have slightly lower efficacy than male condoms when perfectly used and worse effectiveness when imperfectly used (Farr et al., 1994; Trussell, 2011), especially when first adopted (Beksinska et al., 2012).[5] If the idea that female condoms yield lower disutility to men means that women with low bargaining power can convince their partners to use female condoms, but not male condoms, then making female condoms available to such women may decrease rates of unprotected sex. This is likely to lead to substantial welfare gains, given the negative externalities of risky sexual behaviour in terms of transmission of HIV/AIDS and other STIs.

In this paper, we therefore test whether providing access to and information about female condoms can increase condom use, particularly among women with low household bargaining power. To do so, we evaluate a condom programme in the slums of Matola, Mozambique. The female HIV prevalence rate in Matola has been estimated at 29.6% (Saúde, 2015), and thus condom adoption is a key public health concern. In the programme, women attend a series of six fortnightly group sessions, in which a local health worker provides information about female condoms as well as other contraceptives, HIV/AIDS and STIs, and broader sexual and reproductive health issues. Participants are also given free and discreet access to both male and female condoms at the end of each session.[6] The intervention thus allows us to study which women, if any, adopt female condoms when informational and supply constraints are alleviated.

4 Female condoms can also be inserted by women prior to intercourse. This may allow women with low bargaining power to change the default from unprotected sex to female condom use as partners enter into bargaining over condom use. It could also potentially act as a commitment device, for example if arousal leads to increases in irrational behaviour (Ariely and Loewenstein, 2006). In principle, both of these mechanisms could increase the likelihood that women with a given bargaining power are able to convince their partners to use female condoms.

5 Trussell (2011) finds in the US that first-generation female condoms are 95% effective at preventing pregnancy in the first year of use if perfectly used, but 79% effective in typical use. Meanwhile male condoms are 98% effective if perfectly used, and 85% effective in typical use. On the other hand, Macaluso et al. (2007) find in a 20-week RCT that female condoms are associated with a much higher rate of failure or incorrect use than male condoms (34% versus 9%), but that this does not lead to a significant difference in efficacy between male and female condoms.

6 The female condoms distributed are of the FC2 variety, which are made of synthetic nitrile and have been found to have greater acceptability among users. See https://fc2femalecondom.com/fc2-global-home/ for details.

To formalise our predictions, we introduce a model of intra-household bargaining over dual protection methods. Both men and women value the levels of pleasure and of health protection associated with different contraceptive technologies. However, for the reasons outlined above, we argue that women place a lower relative weight than men on the pleasure loss associated with condom use, and a larger relative weight on the level of health protection provided. When the only dual protection options available are male condoms or sex unprotected by condoms, the model therefore predicts that many women prefer to use male condoms, but that those with low bargaining power are unable to convince their partners to do so. When female condoms are introduced, some of these women with low bargaining power are now able to convince their partners to adopt female condoms (but not male condoms). Meanwhile, some women who were previously using male condoms also substitute into using female condoms. However, the marginal loss in protection for these women, due to the slightly lower efficacy of female condoms, is far outweighed by the benefits from the reduced rates of unprotected sex.

To test the predictions, we use a phased-in experimental design to compare women who attended the condom training to women who were randomised to wait an additional six months before beginning the course. Our primary outcome variable is the use of female and male condoms at endline, and how this varies by baseline bargaining power. Since measuring condom use inevitably relies on self-reporting, we also collect weekly coital diary data for a random subsample of the women. These diaries allow us to re-estimate the treatment effects using a more complex and granular survey instrument — which by its nature is less vulnerable to misreporting — in order to minimise concerns about recall and experimenter demand effects. Furthermore, the diaries allow us to investigate impacts at the sex-act level. To measure bargaining power, we collect information about assets brought by the woman to the relationship, and also enumerate three different survey modules covering decision-making, dynamics of the relationship, and violence.[7]

7 Whilst we also sought to include incentivized measures of remaining STI-free and of bargaining power, unfortunately the National Research Ethics Committee of Mozambique's procedures do not allow any financial remuneration of survey participants.

The results show a large impact of treatment on female condom use: an increase of 18.4 percentage points in the proportion of women who have ever used female condoms (equivalent to 209% of the endline mean in the control group) and an increase of 7.7 percentage points (385%) in the proportion of those currently using female condoms. We see no significant evidence of substitution away from male condoms. Moreover, the diary data show that treatment leads to an increase of 0.29 sex acts per person per week (37% of the endline mean in the control group), and an increase of 9.7 percentage points (970%) in the proportion of sex acts per person per week protected by female condoms. Among women in a stable relationship, take-up of female condoms is driven by women with lower baseline bargaining power. This is in line with the predictions of the model, and with the assumption that female condoms are more acceptable to men than male condoms. Adoption of female condoms is also particularly strong among women who are not in a stable relationship at baseline and women who are HIV-positive, suggesting increased coverage of the sex acts with the highest risk of HIV transmission.

To our knowledge, this is the first experimental study to consider intra-household bargaining over condom use, and how the introduction of female condoms might interact with the bargaining process.[8] Gertler et al. (2005) model bargaining over male condom use between female sex workers and male clients in Mexico, and compare the model's predictions to observational data on sex workers' recent transactions. They establish that unprotected sex is associated with a higher price. Our contribution is distinct insofar as we model the repeated household bargaining process that takes place within couples, rather than a one-shot or finite-horizon interaction like that between sex workers and clients, or casual partners. We also consider the introduction of an alternative technology — namely female condoms. The existing literature on bargaining within couples focuses on fertility. Eswaran (2002) models how increasing women's bargaining power within the household may decrease fertility and also child mortality. As mentioned above, Ashraf et al. (2014a) consider "moral hazard" in the use of concealable contraceptives. Similarly, Rasul (2008) shows that couples in Malaysia appear to bargain without commitment over fertility. In contrast, we consider bargaining over

8 For qualitative studies in the sexual health literature, see Choi et al. (2004).

dual protection methods that also guard against HIV/AIDS and other STIs, which by their nature are not concealable.

A number of medical studies have examined the effects of introducing female condoms alongside male condoms (Vijayakumar et al., 2006). In particular, Fontanet et al. (1998) evaluate a randomised provision of either male condoms or male and female condoms to sex workers in Thailand. They find an insignificant decrease in the incidence of unprotected sex and STI prevalence when both types of condoms are offered, rather than just male condoms. More recently, Mantell et al. (2015) compare a shorter and more intensive course of information about female condoms among university students in KwaZulu Natal. They find that both treatments are effective in reducing incidence of unprotected sex, but that there is no significant difference across the two course types (although this may reflect a lack of power). Coman et al. (2013) use propensity score matching to study the effects of female condom availability via a long-term programme in the USA, again finding positive impacts on coverage. Meanwhile Ashraf et al. (2014b) examine the effect of incentives to sell female condoms on sales by agents, but do not study impacts on the end users.

We also add to a large literature on interventions designed to reduce transmission of HIV. A number of studies have focused on providing teenage girls with HIV risk information (Dupas, 2011), educational subsidies and abstinence curricula (Duflo et al., 2015), education-linked conditional cash transfers (Baird et al., 2013), and vocational and "life skills" training (Bandiera et al., 2015). Others have experimented with giving high-risk populations financial incentives to stay free from sexually transmitted infections (De Walque et al., 2012; Galárraga et al., 2014; Bjorkman Nyqvist et al., 2015), to take HIV tests (Thornton, 2008; Baird et al., 2014), and to undertake voluntary adult male circumcision (Thirumurthy et al., 2014; Godlonton et al., 2016).

In terms of policy implications, our results show that expanding access to and information about female condoms can expand coverage especially among women with low household bargaining power. This is particularly important given that women with low bargaining power are a high-risk group: for example, violent male partners are more likely to have extrama-

rital or multiple sexual relationships, and as a result to be HIV-positive (Dunkle et al., 2004; Silverman et al., 2008; Decker et al., 2009). Furthermore, women with low bargaining power are less likely to have access to antiretroviral therapy (Ruger, 2004). The fact that use of female condoms does not simply substitute for use of male condoms also suggests that the correct cost comparison for female condom programmes is not male condom programmes, but rather the costs of antiretroviral therapies, unwanted pregnancies, and other costs associated with unprotected sex.[9] Finally, our results reinforce findings in the medical literature that female condoms can be introduced with relatively simple information and access programmes (Mantell et al., 2015). Similar introductions and demonstrations of male condoms are a common component of sex education programmes across the world, for example in high schools and youth groups. Our results highlight the potential benefits of adding female condoms to such programmes.

The rest of this paper is organised as follows. Section 3.2 describes the context and intervention. Section 3.3 outlines the model and predictions for the effects of the intervention. Section 3.4 describes the study design and data sources. Section 3.5 details the results and robustness checks. Section 3.6 concludes.

3.2 CONTEXT AND INTERVENTION

3.2.1 *HIV and condom use in Maputo*

Our study took place in Matola, the capital of Maputo Province, Mozambique. Matola lies approximately 10km west of Maputo City, and with 893,000 inhabitants is the second most

9 One argument against the widespread introduction of female condoms has been their higher unit cost compared to male condoms: specifically, $0.57 compared to $0.03 (Mantell et al., 2015). This is due to a monopoly on the production of WHO-approved female condoms, and consequent low production volumes (Peters et al., 2010). Lower-cost female condoms have been developed in India and approved by the EU, although have been waiting for several years for WHO approval (*ibid.*). Estimates suggest that increases in the scale of female condom production could lead to a cost-neutral increase in coverage relative to male condoms (Dowdy et al., 2006).

populated area of Mozambique (Instituto Nacional de Estatistica Moçambique, 2014). HIV prevalence in Maputo Province is high and disproportionately affects women, at an estimated 29.6% for women and 15.8% for men (Saúde, 2015). Concurrency has been identified a contributing factor: like elsewhere, having multiple concurrent sexual partners is regarded as acceptable and even desirable for men, although this practice is more frowned-upon for women (Macia et al., 2011). This is true both inside and outside of stable relationships. 85% of the women in our sample are in stable relationships, with an average duration of nine years and no break-ups during the length of our study. Nonetheless, 36% of the women in stable relationships report that they believe their partner is involved with other people. In such a climate, dual protection methods which protect against transmission of HIV and other STIs are not close substitutes for pure contraceptive methods such as the pill. As a result, dual protection methods may be used in addition to pure contraceptive methods: in our baseline sample, 39% of respondents are currently using pure contraceptive methods (mainly the pill or injectables), and of those 40% are also currently using male condoms.

Both male and female condoms are available in Matola, but male condoms are far more accessible, especially for men. Female condoms are generally only available at health facilities, which are often difficult to reach: subjects in our sample report taking on average 60 minutes to get to the nearest health facility. Moreover, even at these health facilities, female condoms are subject to frequent stock-outs (Pilz, 2014). Meanwhile, male condoms are readily available, both for free at health facilities and cheaply on the private market. Nonetheless, women report feeling uncomfortable purchasing male condoms, and are also less able than men to access places where male condoms are sold more discreetly, such as bars and nightclubs.

Despite the widespread availability of male condoms, at least for men, there is evidence that men's preferences constrain adoption.[10] Of the women in our study who are currently sexually active but not using any form of protection at baseline, by far the most common reason is that their partner does not like using male condoms or simply refuses to use them

10 In addition to the fact that women in the study area are more vulnerable to HIV/AIDS, there is also evidence that women have lower desired current fertility than their male partners: 12% of women in stable relationships say that they want another child now, whereas 23% claim that their partner does. Men may also have higher desired total fertility, as 68% of women claim their partner wants another child whereas only 55% of the women say they want another child.

(45% of responses).[11] This appears to be less of a problem for women with higher bargaining power: the probability of a woman reporting that her partner obstructs use of male condoms in this way is negatively and highly significantly correlated with four of our six measures of women's bargaining power; see Section 3.4.4 for details.[12]

3.2.2 *Female condom intervention*

The intervention we study is run by Pathfinder International, and is aimed at women in populations with high HIV risk. The programme consists of six group sessions lasting ninety minutes each, held every two weeks over a three month period. The facilitators are trained female health workers from the local area, and thus are socially proximal to the participants. The sessions cover information on female condoms and demonstration of their use on pelvic models; information about other contraceptive methods; information on HIV/AIDS and other STIs; and discussions around issues of consent, negotiation of contraceptive use, intimate partner violence, and women's rights. These discussions are a key ethical component of the programme design, in order to mitigate the risk of these women facing increased violence when introducing new contraceptives into the home. Group sizes range from a minimum of five to a maximum of twelve women per facilitator — thresholds set by the NGO for creating an environment conducive to discussion. Male and female condoms are made freely and discreetly available to participants at the end of each session.

11 The other main reasons are having given up condoms when they wanted to fall pregnant (27%), and discomfort (12%).

12 All bargaining power variables are constructed from principal component analysis, and ordered such that a higher value implies higher bargaining power. The correlation between "partner refuses male condoms =1" and a one standard deviation increase in the principal component of bargaining power in the following domains are: negative relation -0.11 (p-value 0.09); violence -0.13 (p-value 0.03); assets 2 -0.10 (p-value 0.002); assets 3 -0.18 (p-value $<$0.01). We also observe a positive correlation between assets brought by the woman to the relationship and baseline use of male condoms in the last thirty days. Specifically, the correlation is a 7.3 percentage point increase in the likelihood of use per one standard deviation in the assets 1 (p-value 0.035). In contrast, women in more violent relationships are *more* likely to be currently using male condoms at baseline (8.1 percentage points increase in the likelihood of use per one standard deviation in the first principal component of violence, p-value 0.022). However, the relationship between bargaining power and actual use of male condoms may be confounded by other factors, for example if violent male partners are more likely to be HIV-positive (Decker et al., 2009).

The intervention thus allows us to study which women, if any, adopt female condoms when both informational and supply constraints are alleviated. We do not attempt to disentangle these two channels. Moreover, the estimated treatment effect may also include the effect of simply coming together in a group with other women to discuss personal issues. However, since any standard sex education programme would likely involve all of these components, the effect of this whole package is arguably of most interest to policymakers.

3.3 MODEL

We first introduce a simple household bargaining model, to formalise predictions about which women might adopt female condoms in terms of their bargaining power. We focus on a couple's decision over dual protection technologies, and abstract from pure contraceptive technologies such as the pill which only protect against pregnancy. This is in light of the evidence discussed above, that pure contraceptives are not close substitutes for dual protection methods in contexts such as our study setting.

We first consider the couple's ideal choice of dual protection technology if a whole continuum of technologies were available. We then consider the case in which male condoms are the only dual protection technology on offer, and thus the couple must choose between male condoms or unprotected sex. The latter includes sex which is protected by contraceptives such as the pill but not by a dual method, and thus which is protected against pregnancy but not against HIV/AIDS and other STIs. This binary-choice setting is arguably a close approximation to our study setting prior to the introduction of female condoms. We then consider what happens when female condoms are introduced, expanding the couple's choice set to male condoms, female condoms, or unprotected sex. As a simplification, we assume that the financial and opportunity costs of acquiring any of the technologies is zero, since our experiment makes both male and female condoms available for free. Nonetheless, the same

qualitative predictions hold for any intervention which significantly expands the supply of female condoms, and thus vastly reduces the direct and opportunity costs of acquiring them.

3.3.1 *Preferences*

Consider a population of heterosexual couples $i \in I$, each consisting of a male m_i and a female f_i. Let $k \in \{m, f\}$ denote gender. When considering the choice of contraceptive technology, individuals have preferences over the pleasure P_i and health H_i offered by each method. Health represents being free of HIV/AIDS, other STIs, and associated health problems such as transmission of HIV from mother to child during pregnancy or childbirth. For simplicity, we model P_i and H_i as the levels of pleasure and health that a technology yields to the couple jointly. For example, H_i may include a certain average level of risk to the couple of becoming infected with HIV. However, individual partners may differ in the relative weight that they place on P_i and H_i. For example, the woman may place a higher weight on health if she is disproportionately likely to become infected with HIV due to social or biological factors. Specifically, let the preferences of the spouse of gender k in couple i be given by:

$$u_i^k(P_i, H_i) = \beta_i^k \ln P_i + (1 - \beta_i^k) \ln H_i, \tag{14}$$

where β_i^k denotes an individual's relative preference for pleasure over health.

Assumption 1. $\beta^f < \beta^m$. We assume that women place a higher relative weight than men on the health protection offered by different dual protection technologies, and consequently a lower relative weight on the pleasure offered. This assumption is motivated by the discussions in Sections 3.1 and 3.2.1, given women's disproportionate risk of being infected by their male partner, and men's stronger reported displeasure from male condom use.

Assumption 2. Decisions over dual protection technologies are taken co-operatively. The decision to use a dual protection method must be taken jointly, since use of both male and female

condoms is observable. Moreover, the decision must be taken each time a couple has sex, in contrast to decisions over longer-acting contraceptives. Thus in stable couples, contraception can be thought of as a repeated game with a long time horizon. As such, it is natural to model the choice of dual protection method within a cooperative model, resulting in choices that are Pareto efficient.

Efficiency implies that couples choose a contraceptive technology to maximise:

$$V_i = \alpha_i u_i^f (P_i, H_i) + (1 - \alpha_i) u_i^m (P_i, H_i), \tag{15}$$

where $\alpha_i \in [0, 1]$ is the woman's Pareto weight in the couple's sharing rule (Browning and Chiappori, 1998). We interpret this weight as her bargaining power, and thus Equation 15 represents a collective model of the household (Chiappori, 1992). The weight α may depend on factors such as her relative contribution to the couple's income and housework, and her options outside of the relationship.

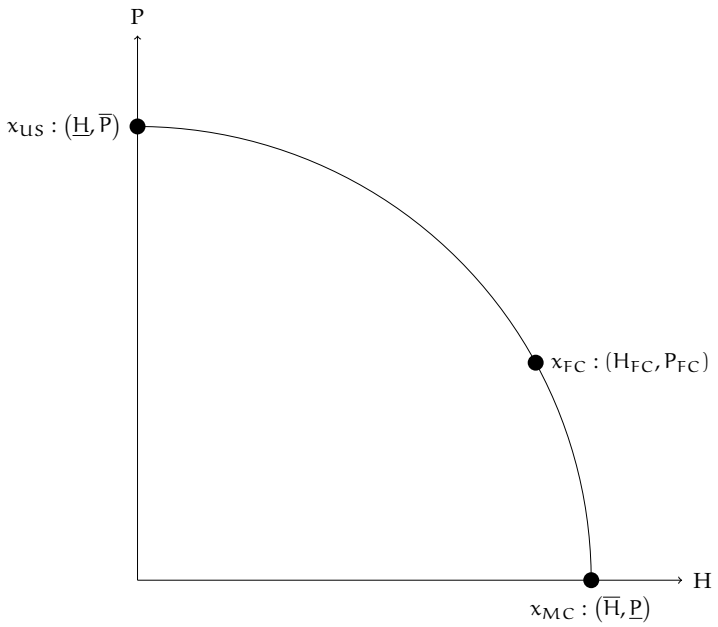
3.3.2 *Optimal contraceptive choice and bargaining power*

In general, let the dual protection technology frontier be represented by a continuum $x \in [\underline{x}, \overline{x}]$, which is a segment of the real line. We adopt the convention that H is increasing in x and P is decreasing in x. By definition of being on the frontier, $P'(H) < 0$. Let $P''(H) \leqslant 0$, such that the frontier is weakly concave.[13] We also define $\overline{P} = P(\underline{x})$, $\underline{P} = P(\overline{x}) > 0$, $\underline{H} = H(\underline{x}) > 0$ and $\overline{H} = H(\overline{x})$. Figure 10 illustrates this contraceptive technology frontier.

We first consider the couple's ideal choice of H — with P being implied by the technological frontier — in this general, hypothetical scenario where the whole continuum of dual protection technologies is available. To simplify matters, we assume that the couple first decides whether or not to have sex, and then bargains over the choice of dual protection technology

13 Strictly speaking, we further assume that there is a continuously differentiable vector-valued function $(P(x), H(x))$ which satisfies $P'(x) < 0$ and $H'(x) > 0$.

Figure 10: Dual-protection technology frontier



(including unprotected sex). This abstracts from simultaneous bargaining over whether to have sex and which type of contraception to use.[14] Dropping the i subscripts for convenience, the couple's utility maximisation problem written more generically is:

$$\max_{H} V\left(P\left(H\right), H; \alpha\right) \tag{16}$$

As long as Assumption 1. holds and thus $\beta^f < \beta^m$, then it is straightforward to show that the optimal choice of health is increasing in α. This holds for Cobb-Douglas utility, but also more generally for strictly quasi-convex preferences; see Appendix 3.7.1 for proof. The logic of this is intuitive — if the female places relatively greater weight on health than the male does, then the more bargaining power she holds, the more the couple's choice of dual protection technology will be tilted towards health (and consequently away from pleasure).

14 We plan to address this in future work, see Section 3.5.2.

3.3.3 *Choice between male condoms and unprotected sex*

We next consider the couple's choice when the only dual protection technology available is male condoms — thus the couple faces a choice between male condoms or sex that is unprotected from HIV/AIDS and other STIs. Let unprotected sex and male condoms represent the end-points of the technology frontier: unprotected sex offers the greatest pleasure but the lowest health; while male condoms offer the greatest health but the lowest pleasure. Formally:

$$x_{US} = \underline{x} \text{ and } P(x_{US}) = \overline{P},\ H(x_{US}) = \underline{H} \tag{17}$$

$$x_{MC} = \overline{x} \text{ and } P(x_{MC}) = \underline{P},\ H(x_{MC}) = \overline{H} \tag{18}$$

The couple must therefore choose between these two corner solutions.[15] If the couple opts to have unprotected sex then its utility is:

$$V(x_{US};\alpha) = \left(\alpha\beta^f + (1-\alpha)\,\beta^m\right)\ln\overline{P} + \left(\alpha\left(1-\beta^f\right) + (1-\alpha)\left(1-\beta^m\right)\right)\ln\underline{H} \tag{19}$$

Meanwhile, if the couple chooses to use a male condom, its utility is:

$$V(x_{MC};\alpha) = \left(\alpha\beta^f + (1-\alpha)\,\beta^m\right)\ln\underline{P} + \left(\alpha\left(1-\beta^f\right) + (1-\alpha)\left(1-\beta^m\right)\right)\ln\overline{H} \tag{20}$$

The couple will choose the male condom over unprotected sex if and only if:

$$V(x_{MC};\alpha) \geqslant V(x_{US};\alpha), \tag{21}$$

15 Given that couples cannot commit to condom use, but rather have to agree each time and thus may renegotiate, it is infeasible for the couple to convexify the technology frontier by committing to use male condoms on a certain fraction of occasions and have unprotected sex on the remainder.

or equivalently if:

$$\frac{\alpha\left(1-\beta^{f}\right)+(1-\alpha)\left(1-\beta^{m}\right)}{\alpha\beta^{f}+(1-\alpha)\beta^{m}} \gtrless \frac{\ln\overline{P}-\ln\underline{P}}{\ln\overline{H}-\ln\underline{H}} \tag{22}$$

Re-arranging Equation 22, we can define a cut-off level of female bargaining power, $\alpha^{*}_{MC,US}$, above which the couple will use male condoms rather than engaging in unprotected sex:

$$\alpha^{*}_{MC,US} = \frac{\beta^{m}}{\beta^{m}-\beta^{f}} - \frac{1}{(\beta^{m}-\beta^{f})}\frac{\ln\overline{H}-\ln\underline{H}}{(\ln\overline{H}-\ln\underline{H}+\ln\overline{P}-\ln\underline{P})} \tag{23}$$

Intuitively, the first term of Equation 23 says that if the male places a higher relative weight on pleasure (β^{m}), then higher female bargaining power is required to convince him to use male condoms. Meanwhile, the second term shows that if male condoms offer a relatively large health improvement over unprotected sex, then lower female bargaining power is required to secure their use. As long as:

$$\beta^{f} < \frac{\ln\overline{H}-\ln\underline{H}}{\ln\overline{H}-\ln\underline{H}+\ln\overline{P}-\ln\underline{P}} < \beta^{m}, \tag{24}$$

then $\alpha^{*}_{MC,US} \in (0,1)$. That is, if the health benefits provided by the male condom are large enough to outweigh the female's relative emphasis on pleasure, but not large enough to outweigh the male's, then adoption of the female condom will depend on the female's bargaining power.

3.3.4 *Introduction of female condoms*

Our treatment can be thought of as expanding the set of readily-available technologies from the binary set $x \in \{MC, US\}$ to the ternary set $x \in \{MC, FC, US\}$. As discussed in the introduction, the introduction of female condoms provides an intermediate technology option, since there is evidence that female condoms provide lower efficacy but greater pleasure (especially to men). This is illustrated in Figure 10.

Intuitively, the introduction of this interior option may allow some couples to get closer to their optimal choice. Couples who are currently engaging in unprotected sex and couples who are currently using male condoms both decide whether to stick with their original choice or to switch to using female condoms. If couples have uncertainty about the pleasure and health associated with female condoms, then many couples may try female condoms but then switch back to their original technology choice. For simplicity, in what follows we considered only the permanent adoption decision, once any uncertainty about female condoms has been resolved.

Comparing the utility of sex protected by a female condom to the utility of unprotected sex, a couple will switch from unprotected sex to using the female condom if the following condition holds:

$$\alpha \geqslant \alpha^*_{FC,US} = \frac{\beta^m}{\beta^m - \beta^f} - \frac{1}{(\beta^m - \beta^f)} \frac{\ln H_{FC} - \ln \underline{H}}{(\ln H_{FC} - \ln \underline{H} + \ln \overline{P} - \ln P_{FC})} \tag{25}$$

Meanwhile, couples who were using male condoms at baseline will switch to using female condoms if the following condition holds:

$$\alpha \leqslant \alpha^*_{MC,FC} = \frac{\beta^m}{\beta^m - \beta^f} - \frac{1}{(\beta^m - \beta^f)} \frac{\ln \overline{H} - \ln H_{FC}}{(\ln \overline{H} - \ln H_{FC} + \ln P_{FC} - \underline{P})} \tag{26}$$

The range of values of α which satisfy Equations 25 and 26 becomes larger — and thus the likelihood of female condom use becomes greater — as H_{FC} increases and as P_{FC} increases, *ceteris paribus*. Moreover, as long as the following condition is satisfied:

$$\beta^f < \frac{\ln H_{FC} - \ln \underline{H}}{\ln H_{FC} - \ln \underline{H} + \ln \overline{P} - \ln P_{FC}} \leqslant \frac{\ln \overline{H} - \ln H_{FC}}{\ln \overline{H} - \ln H_{FC} + \ln P_{FC} - \ln \underline{P}} < \beta^m, \tag{27}$$

then the range of values of alpha which satisfy Equations 25 and 26 will fall in the unit interval and use of female condoms will depend on female bargaining power. Intuitively, adoption of female condoms will depend on female bargaining power as long as the health benefits of female condoms compared to unprotected sex outweigh the loss of pleasure compared to unprotected sex from the perspective of the female, but not the male (the first inequality in

Equation 27); and as long as the health benefits of the male condom compared to the female condom outweigh the pleasure benefits of the female condom compared to the male condom from the perspective of the female, but not the male (the final inequality in Equation 27).

3.3.5 *Model predictions*

Equations 25 and 26 imply that, as long as Equation 27 holds, women with intermediate values of α are those who will take up female condoms. However, this does not mean that we will necessarily observe an "inverse-U" relationship between bargaining power and female condom adoption in our experimental sample. This will depend on the extent to which women from the full distribution of bargaining power sign up for the condom programme. Women with the lowest bargaining power may not sign up if their partners disapprove of them participating, or indeed if they predict that they will not be able to convince their partners to use female condoms. Meanwhile, women with the highest levels of bargaining power may not sign up if they are already able to persuade their partners to use male condoms. If the women with lowest bargaining power do not sign up, then we may not observe the upward-sloping portion of the inverse-U relationship between bargaining power and female condom adoption, and instead observe a straightforward negative relationship. Vice versa, if women with the highest bargaining power do not sign up, then we may observe a straightforward positive relationship between bargaining power and female condom adoption.

The model also predicts that of the women who are engaging in unprotected sex at baseline, it is the women with relatively higher bargaining power — although still relatively low bargaining power compared to the whole distribution — who will take up female condoms. Importantly, this implies increased coverage among such women. Meanwhile, of those women using male condoms at baseline, it is the women with relatively lower bargaining power who will switch from male to female condoms. Empirically, it is important to see which of these effects dominate — if take-up of female condoms largely comes from women who were

engaging in unprotected sex at baseline, then female condoms represent a key addition to the set of contraceptive technologies used to prevent HIV/AIDS and other STIs. On the other hand, if female condoms are mainly used as substitutes for male condoms, then offering female condoms will not lead to an increase in condom coverage; although couples who switch must be better off, as they move closer to their optimal technology. Moreover, the marginal loss of efficacy for those couples who switch from male to female condoms may be cause for concern.

Finally, the model predicts that women who place a higher weight on health, and thus have a lower β^f, will have a higher demand for both male and female condoms, *ceteris paribus*. Particular groups of women who are likely to have a lower β^f are those who are HIV-positive;[16] those who have high perceptions of HIV or STI risk; and those who are not in a stable relationship, since their HIV or STI risk may be higher and they may be more concerned about the risk of pregnancy. These groups of women should therefore be more likely to use male condoms at baseline, and more likely to switch from unprotected sex to female condoms when female condoms are introduced.[17]

3.4 STUDY DESIGN AND DATA SOURCES

3.4.1 *Study design*

Pathfinder International began its female condom programme in Matola in 2011, and in 2014 sought to expand this project to four additional neighbourhoods. We exploited this expansion via a phased-in experimental design. The healthcare workers who facilitate the programme

16 In principal, individuals who they are already HIV-positive could either increase or decrease their preference for protection, depending on their level of altruism, their expectations of guilt or punishment from infecting another person, and the status of their stable partner if they have one. However, existing empirical evidence points towards decreased risk-taking after learning a HIV-positive status (Paula et al., 2014).

17 This prediction is reinforced if lower values of β^f are correlated with lower values of β^m, for example if the male partners of HIV-positive women also care more about protection. However, we do not observe proxies of β^m.

first conducted door-to-door recruitment to identify women willing to participate. The eligibility criteria were that women needed to be between 18 and 49 years of age, sexually active, and not pregnant. To avoid spillovers, if a woman signed up with a friend or relative then they were both included in the programme but were automatically assigned to a separate set of groups which would not form part of the study. Otherwise, the average population of these neighbourhoods was 20,000 inhabitants, thus spillovers between unconnected individuals are unlikely to be a cause for concern. Overall 317 women were recruited into the study, and were subsequently administered a baseline survey in August 2014.[18] Shortly after the baseline, one facilitator fell severely ill, and there was nobody sufficiently trained to replace her. Thus all 35 individuals recruited by her (treatment and control) were dropped from the programme and the study. This leaves a baseline sample of 298 respondents.

At the end of the baseline survey, it was explained to each participant that assignment to the first or second wave of the programme would be determined by computer randomisation. Once the entire sample had responded to the baseline survey, the research team randomly allocated half of the respondents recruited by each facilitator to the treatment group and half to the control group.[19] The reason for stratifying on facilitator was that each facilitator recruited in a separate area of the neighbourhoods, and we wanted to achieve balance within each area. This also ensured that there would be enough space for treatment and control participants to attend sessions close to their home.

The treatment group then received the intervention in September-December 2014. Following this, the endline survey was conducted in February-March 2015. A total of 232 respondents were traced and administered the endline survey. The retention rate was thus 78%, which although imperfect is reasonably high for a study in an urban slum area. Whilst attrition is higher in the control group (9.0 percentage points, p-value 0.049), Table 19 shows that the predictors of attrition are not differential across treatment and control. Following the endline survey, the control group then received the intervention from March-May 2015. The control

18 Respondents signed an informed consent form prior to the data collection, and it was explained that they could leave the study at any time.

19 The randomisation was done in private, given the sensitive nature of participating in our intervention. Specifically, a member of the research team took the list of respondents for each facilitator, assigned each a random number in Excel, sorted respondents on this number, and assigned the first half to treatment and the second half to control.

group for one facilitator had already begun their sessions by the time they were administered the endline survey, hence these five observations are dropped from all estimations of treatment effects, leaving a final estimating sample of 227.

Table 19: Predictors of attrition – treatment & control

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Treatment		**Control**		**Test $\beta_1 = \beta_2$**			**N**	
	Mfx	**p-val**	**Mfx**	**p-val**	**χ^2**	**p-val**	**T**	**C**	**All**
Demographics									
Age	-0.01	0.12	-0.01	0.18	0.01	0.91	152	146	298
Years of education	-0.01	0.41	-0.01	0.52	0.03	0.87	149	146	295
Literate	-0.09	0.27	-0.09	0.32	0.02	0.88	151	144	295
HH head	-0.05	0.49	-0.01	0.90	0.19	0.66	152	146	298
Income									
Has job	-0.03	0.67	0.04	0.57	0.47	0.50	151	144	295
Personal income last 30 days	-0.00	0.08	-0.00	0.26	1.13	0.29	152	146	298
Relationships									
In a stable relationship (incl. married)	-0.08	0.32	-0.01	0.95	0.45	0.50	152	146	298
Married (officially or unofficially)	-0.02	0.78	0.07	0.40	0.59	0.44	151	146	297
Years relation	-0.01	0.21	-0.01	0.11	0.04	0.84	121	114	235
Sexual behaviour									
HIV positive (self-report)	0.12	0.07	0.04	0.63	1.00	0.32	131	129	260
STI last 3 months (self-report)	0.06	0.47	-0.19	0.20	2.06	0.15	135	124	259
Names FC as contraceptive	-0.04	0.53	-0.06	0.43	0.00	0.97	150	146	296
Baseline use									
Ever used FC	0.05	0.60	0.12	0.32	0.07	0.80	152	146	298
Ever used MC	0.08	0.28	-0.02	0.78	1.04	0.31	152	146	298
Ever used other	-0.07	0.27	0.03	0.69	1.15	0.28	152	146	298
Used MC last 30 days	-0.04	0.53	-0.07	0.42	0.01	0.92	152	146	298
Current use MC	0.07	0.23	0.02	0.82	0.55	0.46	152	146	298
Current use other	-0.03	0.68	0.08	0.26	1.09	0.30	152	146	298

Notes: N=298 prior to attrition. Lower sample sizes reflect observations that are missing or not applicable. "Treatment" contains all individuals assigned to the treatment group, whether or not they attended the sessions. Columns 1-4 show marginal effects and p-values for logit regressions of the probability of attritting on each covariate, in the treatment and control group respectively. Columns 5 and 6 show the χ^2 statistic and p-value for the test that the marginal effects are equal across the treatment and control groups. Columns 7-9 show sample sizes. "Ever used other" and "Current use other" refer to use of any other modern contraceptive method apart from condoms, e.g. the pill, injectables, or an IUD.

3.4.2 *Survey data*

Table 20 shows measures of key covariates and contraceptive use elicited for the full baseline sample.[20] All are balanced across treatment and control. 85% of respondents report being in a stable relationship, of whom 63% are married.[21] The rest of the sample (15%) are sexually active but not in a stable relationship. The vast majority of respondents report having had just one sexual partner in the last twelve months, with 10% reporting zero partners and 3% reporting two partners. 33% of respondents report being HIV-positive. This may be an underestimate given its self-reported nature, although the figure is close to the average figure of 29.4% for Maputo Province (Saúde, 2015). 13% report having had an STI in the last month; but again this may be under-reported, especially since some further respondents report having had symptoms associated with STIs (not shown).[22] Fewer than half — 41% — are able to name the female condom when asked to list contraceptive methods.[23]

Our primary outcome variables are the use of contraceptive methods, disaggregated by female condoms, male condoms and other modern contraceptive methods (which, as discussed, are mainly the pill and injectables). For each method, we ask respondents whether they have ever used that method, and whether they are currently using it.[24] For male and female condoms, we also ask whether they have used that method in the last thirty days. These different measures allow us to distinguish whether respondents have tried a method at some point (ever use), have used recently (last 30 days) or consider it to be a part of their current portfolio

20 Table 37 in Appendix 3.7.2 shows that all baseline covariates are also balanced in the balanced panel excluding attritters.

21 The latter includes traditional marriages and respondents who describe themselves as "living as married" but not legally married, which is common in this region due to the high bride price and costs of obtaining a marriage certificate.

22 We do not test for HIV, since we would have low power to measure an impact over the length of our study: official guidelines require a minimum of four weeks after exposure before the first test, and a second test three months later in the case of a negative result for high-risk cases (https://www.gov.uk/government/publications/time-period-for-hiv-testing-position-statement). We also opted not to test for STIs such as chlamydia, given the already sensitive nature of participation in the study and the budgetary implications of providing treatment to those who test positive.

23 At endline this increases to 71% for the treatment group. However, it also increases to 65% for the control group, since they hear about female condoms by virtue of being enrolled in the programme. Hence we do not focus on this as an outcome variable.

24 This is in line with the Demographic and Health Survey questions on contraceptive use.

Table 20: Baseline balance of covariates & use – full sample

	Mean	Control Mean	Treatment Mean	t-test	Total N	Control N	Treatment N
Demographics							
Age	30.32	30.12	30.52	-0.42	298	146	152
Years of education	6.21	6.26	6.17	0.27	295	146	149
Literate	0.84	0.84	0.85	-0.17	295	144	151
HH head	0.22	0.21	0.24	-0.51	298	146	152
Income							
Has job	0.38	0.42	0.33	1.64	295	144	151
Personal income last 30 days	745.85	854.52	641.46	1.41	298	146	152
Relationships							
In a stable relationship (incl. married)	0.85	0.85	0.84	0.17	298	146	152
Married (officially or unofficially)	0.63	0.64	0.62	0.37	297	146	151
Years relation	8.66	8.62	8.70	-0.08	235	114	121
# Partners last 12 months	0.92	0.92	0.93	-0.23	298	146	152
Sexual behaviour							
Pregnant	0.05	0.05	0.06	-0.42	297	145	152
HIV positive (self-report)	0.33	0.35	0.31	0.75	260	129	131
STI last 3 months (self-report)	0.13	0.13	0.13	-0.10	259	124	135
Names FC as contraceptive	0.41	0.44	0.39	0.90	296	146	150
Baseline use							
Ever used FC	0.09	0.09	0.09	0.11	298	146	152
Ever used MC	0.74	0.76	0.73	0.59	298	146	152
Ever used other	0.72	0.72	0.72	0.04	298	146	152
Used FC last 30 days	0.03	0.01	0.04	-1.39	298	146	152
Used MC last 30 days	0.32	0.28	0.35	-1.26	298	146	152
Current use FC	0.02	0.02	0.03	-0.33	298	146	152
Current use MC	0.39	0.37	0.41	-0.79	298	146	152
Current use other	0.39	0.41	0.37	0.75	298	146	152

Notes: N=298 in the baseline sample. Lower sample sizes reflect observations that are missing or not applicable. "Treatment Mean" contains all individuals assigned to the treatment group, whether or not they attended the sessions. Column 4 presents the test statistic for the null hypothesis that the mean in the treatment group is equal to the mean in the control group. "Ever used other" and "Current use other" refer to use of any other modern contraceptive method apart from condoms, e.g. the pill, injectables, or an IUD.

(current use). Table 20 describes the baseline values of each of these measures. Baseline use of female condoms is low: nine percent of the respondents have ever used a female condom, three percent have used one in the last 30 days, and two percent are currently using female condoms. Male condom use is substantially higher: around three quarters of women have ever used a male condom, 32% have used one in the last 30 days, and 39% percent say they are currently using male condoms. Altogether 39% are currently using non-condom methods at baseline, comprising 20% using the pill and 14% using injectables, with a small number using intrauterine devices (IUDs), the diaphragm, and sterilisation. These women may have signed up to the female condom programme as a way to switch out of these methods, or because they are seeking additional protection against HIV/AIDS and other STIs.

Finally, Table 21 compares our sample to a representative urban sample of women from Maputo Province, as described in the 2011 Demographic Health Survey (DHS) data. Our sample is close to the overall adult female population of Maputo Province in terms of demographic characteristics such as age, years of education, marital status, pregnancy, and desired fertility. One exception is that the women in our sample are much less likely to have a job, which makes sense if women with a lower opportunity cost of time are more willing to participate in a time-intensive programme. On the other hand, the women in our study appear to have greater bargaining power than the representative sample: they are more literate, began to have sex at a later age, are more likely to have used a condom the last time they had sex, and report greater decision-making power. This is in line with the suggestion in Section 3.3.5, that women with very low bargaining power might not select into our study. That said, the women in our sample report higher rates of emotional, physical and sexual violence than the representative average.

Table 21: Comparison of sample to DHS representative sample

	Study Mean	DHS Mean	t-test	Study N	DHS N
Age	30.18	29.47	1.28	276	1007
Years of education	6.35	6.72	-1.82*	273	1007
Literate	0.85	0.76	3.82***	273	1007
Income					
Has job	0.37	0.58	-6.33***	273	1007
Relationships					
Married (officially or unofficially)	0.63	0.61	0.58	275	871
Pregnant	0.05	0.07	-0.76	275	1007
Wants another child in future	0.57	0.57	0.14	260	961
Sexual Behaviour					
Age of sexual debut	16.62	16.16	3.26***	273	955
Used condom during last time sex	0.54	0.31	6.32***	243	871
Decision-making visiting family	0.62	0.39	6.43***	272	580
Decision-making spending earnings	0.59	0.21	10.85***	275	569
Decision-making her health	0.53	0.39	4.04***	275	580
Experienced emotional violence	0.47	0.37	2.50**	212	372
Experienced physical or sexual violence	0.22	0.13	3.34***	276	1007

Notes: Column 1 displays the mean from our study, N=298. Column 2 shows the 2011 DHS mean for women in urban areas of Maputo Province, N=1007. Lower sample sizes reflect observations that are missing or not applicable. Variables selected for comparison are those that appear in both our study and the DHS, with similar or identical wording.

3.4.3 *Diary data*

A random subsample of the survey respondents also took part in weekly coital diaries. The diaries recorded detailed information on all of the respondents' sexual encounters in the seven days prior to each interview, including: type of sexual activity; relationship to the partner; whether any contraceptives were used, and if so, which ones; whether the two partners discussed the use of contraceptives prior to sex; and if so, who initiated this discussion. Diary interviews took place in a private place chosen by the respondent.[25] The same enumerator interviewed a given respondent each week, to maximise trust and confidentiality.

The diary sampling took place as follows. At baseline, respondents were asked about their willingness to participate in the diaries. This gave a diary sample of 57 respondents: 28 in the treatment group and 29 in the control group. Tables 38 and 39 in Appendix 3.7.2 show that there is balance in baseline covariates and use across the treatment and control groups in the diary sample. Diary interviews took place over a period of 17 weeks, beginning four weeks prior to the first group receiving its first session and ending one week after the last group received its last session.[26] Each week on average 75% of the diary sample participated. Individual respondents took part in the diaries an average of 15 times, with a minimum of three weeks and a maximum of 17 weeks. There are no significant differences in participation between the treatment and control group. Tables 40 and 41 in Appendix 3.7.2 also show that the final sample of diary participants are representative of the balanced panel of all survey participants in terms of baseline covariates and bargaining power, with a few exceptions: the diary participants have been in a relationship for longer than the average study participant,

25 Enumerators were carefully trained, including on first responder procedures if respondents reported experiences of sexual or physical violence.

26 For the diary data, the baseline period for each respondent is taken to run from the start of the diary data collection until one week after the facilitator that the respondent was assigned to began her first meeting for her treatment-group participants. The length of the baseline and endline period therefore varies depending on when a respondent's group held its first meeting. The average in the baseline period is 5.6 weeks. For robustness, we run all estimations setting the end of the baseline period to one week before this, i.e. the exact week of the first treatment-group meeting, and to two weeks after the first treatment-group meeting. Results are similar; tables available on request. Meanwhile the endline period is taken to run from the end of the baseline period (i.e. one week after the first treatment group meeting) until the end of the diary data collection, comprising 8.9 weeks on average.

and no diary respondents are pregnant. All of the results from the diary subsample presented below are robust to re-weighting to make the diary subsample representative of the full sample; tables available on request.

The diaries are a more complex and granular instrument than the baseline and endline surveys, and are also administered much closer in time to the events that they ask about. They are therefore less likely to suffer from reporting and recall bias. Indeed, medical literature has shown that subjects may over-report incidence of sex and use of condoms in surveys compared to diaries (Stalgaitis and Glick, 2014), although this varies across survey populations (Hoppe et al., 2008). Although the time period covered by the baseline and endline surveys was slightly different to that covered by the diary interviews, to the extent that we can cross-check responses we observe consistency in reporting. As far as there are inconsistencies, there is some evidence of under-reporting in the surveys: five out of 57 diary participants report never having used a female condom during the endline survey but do report using them in the diaries; whilst for male condoms the figure is four out of 57 respondents.[27] This strengthens our confidence that there is no systematic over-reporting of condom use in the survey data compared to the diary data. Nevertheless, we run the main impact analyses using both the survey data and the diary data, to show that the results hold across these two different survey instruments; see Section 3.5.2 for details.

As well as constructing variables for contraceptive use at the respondent level, the diary data allow us to analyse the impact of the intervention at the level of the sex act. Altogether respondents report a total of 349 sex acts during the endline period. This represents an average of 6.1 sex acts per respondent, with a minimum of 0 and a maximum of 30. For each sex act and each contraceptive method we set "sex act full endline period" equal to one if that method was used and the sex act took place at any point during the endline period. Similarly, we set "sex act use last 30 days" equal to one if that method was used and the sex act occurred in the last four weeks prior to a respondent's last diary observation. Finally, we set "sex act

27 We cannot make the opposite comparison given that the endline survey took place two months after the end of the diaries, so if a respondent reports using condoms in the survey but not the diaries it may be that they adopted them during those two months.

last 14 days" equal to one if the method was used and the sex act occurred 14 days prior to a respondent's last diary observation.[28]

3.4.4 *Bargaining power*

The women who are in a stable relationship at baseline are on average 5.7 years younger than their partner and earned 3,360 MZN (approximately 110 USD) less than him in the last 30 days. This may point towards these women having lower bargaining power than their male partners, but may also proxy a host of other features of the marriage and labour markets. To obtain a more direct proxy of the woman's outside option, we include a survey module about assets brought by the woman to the relationship. To proxy other features of her bargaining power, we also include three further sets of survey questions, covering decision-making, power dynamics in the relationship, and whether the woman has experienced emotional and physical violence. Table 22 provides summary statistics for each of the individual questions.

Since the assets module contains multiple questions, we perform a principal component analysis to determine which questions are proxying the same facet of women's bargaining power.[29] Table 23 shows that the seven questions load three distinct principal components. Similarly, to check which of our survey measures are proxying the same aspect of bargaining power, we run a principal component analysis including the questions from all three modules. Table 24 shows that the different survey modules each load a different factor. Thus altogether we take six principal components as our preferred measures of bargaining power: three from the assets module, and one each from the decision-making, negative relation and violence modules. For ease of comparability, we scale the components so that the woman with least bargaining power on that measure has a score of zero, and normalise them such that a one point increase in each measure represents an increase of one standard deviation. Table 42

28 When the sex act is the unit of observation we do not have repeat observations per unit, hence we construct the sex-act level variables during the endline period only.

29 This approach is preferable to testing for heterogeneity by each question separately and then correcting p-values for multiple hypothesis testing, since the principal component approach takes into account which measures are proxying the same concept as opposed to different concepts.

Table 22: Bargaining power – summary statistics

	Mean	sd	Min	Max	Total
Did you bring the following assets to your relationship...					
...jewellery?	0.08	0.27	0.00	1.00	264
...animals?	0.00	0.06	0.00	1.00	264
...land?	0.01	0.11	0.00	1.00	264
...electronics?	0.02	0.14	0.00	1.00	264
...money?	0.07	0.25	0.00	1.00	264
...mobile phone?	0.07	0.25	0.00	1.00	264
...kitchen utensils?	0.10	0.30	0.00	1.00	263
Who decides about...					
...buying clothes for you?	0.80	0.40	0.00	1.00	297
...buying phone credit?	0.76	0.43	0.00	1.00	297
...education for the children?	0.49	0.50	0.00	1.00	288
...health expenses for you?	0.55	0.50	0.00	1.00	297
...health expenses for the children?	0.41	0.49	0.00	1.00	291
...if you are allowed to work?	0.59	0.49	0.00	1.00	296
...how earnings are used?	0.60	0.49	0.00	1.00	297
...visits to friends?	0.64	0.48	0.00	1.00	296
...visits to family?	0.64	0.48	0.00	1.00	294
Who usually has more say when you talk about serious things	0.47	0.50	0.00	1.00	250
In general, who do you think has more power in your relationship	0.39	0.49	0.00	1.00	249
Negative relationship indicators					
Most of the time, we do what my partner wants to do	2.33	1.08	1.00	4.00	250
My partner won't let me wear certain things	2.61	1.11	1.00	4.00	250
When my partner and I are together, I'm pretty quiet	3.07	0.96	1.00	4.00	250
My partner has more say about important decisions that affect us	2.39	1.09	1.00	4.00	250
My partner tells me who I can spend time with	2.79	1.09	1.00	4.00	249
I feel trapped or stuck in our relationship	3.20	0.86	1.00	4.00	250
My partner does what he wants, even if I do not want him to	2.86	1.00	1.00	4.00	249
I am more committed to our relationship than my partner is	2.74	1.08	1.00	4.00	250
My partner is involved with other people apart from me	2.77	1.02	1.00	4.00	249
My partner always wants to know where I am	2.16	1.10	1.00	4.00	250
When my partner and I disagree, he gets his way most of the time	2.73	1.06	1.00	4.00	248
Emotional and physical violence					
...your partner hurt your feelings (no=1)?	0.71	0.45	0.00	1.00	229
...you have many fights with your partner (no=1)?	0.79	0.41	0.00	1.00	229
...your partner hit or physically hurt you (no=1)?	0.88	0.33	0.00	1.00	229
...your partner became angry with you without good reason (no=1)?	0.68	0.47	0.00	1.00	229
...your partner ignore you for several days (no=1)?	0.86	0.35	0.00	1.00	229

Notes: All values taken from the baseline survey (N=298). All variables are coded such that a higher value proxies greater bargaining power for the respondent. The assets module was enumerated to all women who lived with their partner at baseline, including a few who did not claim to be in a stable relationship (N=264). The decision-making module was enumerated to all respondents (N=298), except the questions "who has more influence" and "who has more power" which were asked only of women in a stable relationship at baseline (N=250). "Negative relationship" and "emotional and physical violence" questions were also only asked to women who were in a stable relationship at baseline (N=250). Lower observation numbers reflect values missing or unwillingness to answer.

in Appendix 3.7.3 shows that these principal components are balanced across treatment and control.

Table 23: Assets brought to the relationship – principal component analysis

	(1)	(2)	(3)	Unexplained
Assets brought to the relationship				
Jewellery			0.4091	0.2922
Livestock			0.7958	0.2609
Land		0.5964		0.4774
Electronic appliances		0.7373		0.3262
Money	0.6782			0.2543
Mobile phone	0.6587			0.2568
Kitchenware			0.4131	0.3372

Factor loadings from a principal component analysis of all asset questions simultaneously. Only loadings greater than or equal to 0.30 are displayed. All variables are coded such that a higher value proxies greater bargaining power for the respondent. As shown, the asset questions load three separate components, all of which are used in the heterogeneity analysis.

3.5 RESULTS

3.5.1 *Main impact results*

Our preferred estimations are ANCOVA linear probability models of the following form:[30]

$$\Pr[Y_{if1} = 1 | Y_{if0}, treat_{if}, \eta_f] = \alpha + \delta Y_{if0} + \beta treat_{if} + \eta_f, \tag{28}$$

where Y_{if1} is the outcome variable of interest at endline, and Y_{if0} is its value at baseline. $treat_{if}$ is a dummy for being assigned to the treatment group, i.e. to receiving the pro-

30 All of the results from the full sample are robust to using a logit specification; tables available on request. We employ a linear probability model to allow comparison with the treatment effects estimated for specific subsamples, such as the diary respondents. In such subsamples, appealing to the large-sample properties of maximum likelihood estimators becomes questionable, especially conditional on seventeen facilitator fixed effects.

Table 24: Bargaining power – principal component analysis

	(1)	(2)	(3)	Unexplained
Decision-making (1= she is involved)				
Clothes				0.6113
Phone credit				0.6633
Children's education	0.3543			0.3912
Her health	0.3516			0.4025
Children's health	0.3277			0.4628
Her employment	0.3397			0.4127
Spending earnings	0.3137			0.4806
Visiting friends	0.3590			0.3661
Visiting family	0.3494			0.3993
In general, more influence				0.7911
In general, more power				0.7707
Negative relationship (4=completely disagree)				
We do what he wants		0.3306		0.7046
He won't let me wear certain things				0.8640
I'm quiet around him				0.6304
He has more say about joint decisions				0.7509
He controls who I spend time with				0.7319
I feel trapped or stuck				0.7297
He does what he wants				0.7316
I'm more committed		0.3206		0.6817
He sees other people				0.7384
He wants to know where I am		0.3608		0.7433
He gets his way when we disagree				0.8898
Emotional and physical violence (1=no)				
He hurts my feelings			0.4252	0.5982
We have many fights			0.4319	0.5665
He hit or physically hurt me			0.3511	0.7303
He gets angry at me for no reason			0.3966	0.6417
He ignores me for several days			0.3795	0.6879

Notes: Factor loadings from a principal component analysis of all bargaining power variables simultaneously. Only loadings greater than or equal to 0.30 are displayed. All variables are coded such that a higher value proxies greater bargaining power for the respondent.

gramme in the first rather than the second phase. β represents the intent-to-treat effect, since not all individuals assigned to treatment attended the programme: the participation rate was around 65% for each individual session, with 20 women (17.7% of the control group) not attending any of the six sessions. η_f is a facilitator fixed effect, which is included for inference since randomisation was blocked on the seventeen facilitators (Bruhn and McKenzie, 2009). Standard errors are robust to individual-level heteroskedasticity, as this was the level of randomisation.

Table 25 displays the main impacts of the intervention. The programme has a substantial and highly significant effect on the use of female condoms. Specifically, we observe an 18.4 percentage point increase in the proportion of women who have ever used female condoms (equivalent to 209% of the endline mean in the control group); a 4.7 percentage point (470%) increase in the proportion who have used a female condom in the last thirty days; and a 7.7 percentage point (385%) increase in the proportion who are currently using female condoms. We see no evidence of changes in behaviour in the control group, as the control group endline means of female condom use are identical to those at baseline (see Table 20).[31] This reinforces the idea that our treatment expanded access to female condoms which are otherwise difficult to obtain in the study area. Indeed, reported use of female condoms in the treatment group at endline is highly correlated with the number of free female condoms that a respondent took from the sessions, which was discreetly recorded by facilitators.[32] This also weighs against concerns that reported use of female condoms among the treatment group might represent response bias. The fact that the treatment effect on ever use is higher than that on use in the last thirty days and current use suggests that many women in the treatment group try female condoms at the start of the intervention, then a smaller although sizeable fraction continue to use them. This is a natural adoption pattern if couples experiment with female condoms and thereby learn more about their costs and benefits, then some return to their original contraceptive method while others adopt female condoms more permanently.

31 There is thus no evidence that the control group experienced anticipation effects, disappointment effects, or spillovers.

32 The correlation between the total number of condoms taken and endline measures of female condom use are: ever use 0.38 (p-value < 0.01); use last 30 days 0.21 (p-value 0.02); current use 0.29 (p-value <0.01).

Table 25: Treatment effects – primary outcome variables

	(1) **Ever FC**	(2) **Ever MC**	(3) **Ever other**	(4) **Last 30 days FC**	(5) **Last 30 days MC**	(6) **Current FC**	(7) **Current MC**	(8) **Current other**
Treatment	0.184***	-0.012	0.020	0.047**	-0.052	0.077**	0.060	0.030
	(0.042)	(0.041)	(0.042)	(0.023)	(0.057)	(0.030)	(0.058)	(0.053)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓	✓	✓
Observations	227	227	227	227	227	227	227	227
Control mean endline	0.088	0.824	0.735	0.010	0.363	0.020	0.353	0.412

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms, male condoms and other modern contraceptive methods. Columns 1-3 refer to whether the respondent has ever used the method, columns 4 and 5 to whether she has used it in the last 30 days (this was only asked for condoms, not for other contraceptive methods), and columns 6 to 8 to whether she is currently using it. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

We do not observe any significant evidence that respondents substitute away from or increase their use of male condoms. However, our power to detect impacts on male condom use is lower given the high baseline rates of male condom use.[33] Table 26 shows that when we split the sample by women who are using or not using male condoms at baseline, in line with the model we see take-up from both groups. The strong take-up of female condoms among women not using male condoms at baseline (columns 1, 3 and 5) is particularly important from a policy perspective. It implies that the treatment decreases the number of women having sex unprotected from HIV/AIDS and other STIs, rather than generating pure substitution away from other dual contraceptive methods. In the full sample, the estimated treatment effect on current use of either male or female condoms is an increase of 7.9 percentage points; but we have lower power to detect changes in this aggregate outcome, and the point estimate is marginally insignificant with a p-value of 0.176 (not shown). Table 43 in Appendix 3.7.3 shows that again we do not observe any significant impacts on male condom use when we split the sample by those using and not using male condoms at baseline.

Table 25 shows that we also see no increase in or substitution away from other contraceptive methods such as the pill and injectables. This suggests that women who adopt female condoms were either previously using no contraceptives, or use female condoms in addition to other methods in order to protect against HIV/AIDS and other STIs. Indeed, of the women who are currently using female condoms at endline, 42% are also using other contraceptive methods (mainly the pill or injectables). Nevertheless, Table 44 in Appendix 3.7.3 shows that when we split the sample by women who are using or not using any method of contraception at baseline, the observed impacts on the use of female condoms are particularly strong among women who are not using any method at baseline (although they are not significantly higher than the impacts in the rest of the sample). We again do not observe any significant impacts on male condom use, either among those not using or using any contraceptive method at baseline; tables available on request.

33 We have 80% power to detect the following minimum detectable effect sizes at the 5% level in a two-tailed test: ever use – female condoms 9.5 percentage points, male condoms 15.6 percentage points; use last 30 days – female condoms 5.2 percentage points, male condoms 17.0 percentage points; current use – female condoms 6.3 percentage points, male condoms 18.1 percentage points.

Table 26: Treatment effects on female condom use, by baseline male condom use

	(1) **Ever FC No MC at baseline**	(2) **Ever FC Current MC at baseline**	(3) **Last 30 days FC No MC at baseline**	(4) **Last 30 days FC Current MC at baseline**	(5) **Current FC No MC at baseline**	(6) **Current FC Current MC at baseline**
Treatment	0.169***	0.232***	0.073**	0.030	0.085***	0.049
	(0.047)	(0.074)	(0.030)	(0.034)	(0.031)	(0.057)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	141	86	141	86	141	86
Control mean endline	0.092	0.081	0.000	0.027	0.000	0.054
$\chi^2(1) : (a) = (b)$		0.52		0.91		0.30
$Pr > \chi^2$		0.471		0.340		0.583

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms. Columns 1-2 refer to whether the respondent has ever used that method, columns 3-4 to whether she has used it in the last 30 days, and columns 5-6 to whether she is currently using it. Odd-numbered columns present results for the subsample of individuals not currently using male condoms at baseline; even-numbered columns present results for the subsample of individuals who are currently using male condoms method at baseline. The bottom two rows present chi-squared statistics and their p-values for the test that the treatment effect is the same across the two subsamples. These are obtained from seemingly unrelated estimations. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

As discussed in section 3.3.5, we would expect women who are not in a stable relationship to place a larger weight on the health offered by dual protection technologies, and so to have a higher demand for condoms. Table 45 in Appendix 3.7.3 shows that in line with this, the treatment effect on ever use of female condoms is stronger for women who are not in a stable relationship at baseline.[34] Nonetheless, when we restrict the sample to just those women in a stable relationship the estimated treatment effect is a 16.4 percentage point increase in ever use of female condoms (p-value < 0.01), a 5.6 percentage point increase in use in the last 30 days (p-value 0.042), and a 7.9 percentage point increase in current use (p-value 0.019); tables available on request. Thus the treatment does also lead to take-up of female condoms among women in stable relationships. This may be rational in particular if the partners are serodiscordant (i.e. one partner is HIV-positive while the other is HIV-negative) or if one or both partners have relations with others, or at least suspect their partner to.[35]

3.5.2 *Diary results*

3.5.2.1 *Impacts at the sex-act level*

We can use the diary data to measure treatment effects at the sex act level. In particular, we examine impacts on the proportion of total sex acts protected by male and female condoms, since these are key variables of interest for policymakers and for cost-benefit analyses. Table 27 shows the results. There is a significant increase in the proportion of sex acts protected by a female condom in the full endline period, the last 30 days, and the last 14 days (5.1, 8.5 and 6.7 percentage points respectively, all significant at the 5% level). Consistent with the respondent-level results from the survey data, there is no significant effect on the proportion of sex acts protected by male condoms, and the point estimates are small.

34 This is perfectly correlated with being in a stable relationship at endline, as we do not observe any break-ups of stable relationships over the study period.

35 In serodiscordant couples, use of condoms is still rational even if the partners do not expect to follow through on using them all of the time, since the chance of becoming infected from any one unprotected act is approximately 0.001 for female to male transmission and 0.002 for male to female transmission (Dowdy et al., 2006).

Table 27: Impacts on proportion of all sex acts protected – diary subsample

	(1) **% sex acts with FC full endline period** β / (s.e.)	(2) **% sex acts with MC full endline period** β / (s.e.)	(3) **% sex acts with FC last 30 days** β / (s.e.)	(4) **% sex acts with MC last 30 days** β / (s.e.)	(5) **% sex acts with FC last 14 days** β / (s.e.)	(6) **% sex acts with MC last 14 days** β / (s.e.)
Treat*endline	0.051**	-0.005	0.085**	0.015	0.067**	-0.004
	(0.023)	(0.087)	(0.034)	(0.119)	(0.031)	(0.135)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	349	349	204	204	143	143
Control mean	0.010	0.330	0.009	0.374	0.013	0.387

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are the number of sex acts. "Treat*endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat*endline" is the intent-to-treat effect. All regressions are linear probability models with the sex act as the unit of observation. Given that by definition there are no repeated baseline and endline observations at the level of the sex act, baseline values do not exist and so are not included as regressors. All specifications include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation.

3.5.2.2 *Individual-level diary results*

To further understand the precise mechanisms for the observed treatment effects, we use the diaries to examine the total number of sex acts per respondent per week, the proportion of protected sex acts per respondent per week, and the proportion of sex acts per respondent per week which involve discussions or arguments over condom use. Taking advantage of the weekly nature of the diaries, we estimate the following fixed effects panel specification:

$$Pr[Y_{ift} = 1|treat_{if}, \gamma_i, \eta_f] = \beta treat_{if} * endline + \gamma_i + \eta_f * endline, \quad t = 1, 2, ...T \quad (29)$$

where Y_{ift} is the outcome variable of interest for individual i assigned to facilitator f in week t. The unit of observation is thus the respondent-week. $endline$ is an indicator equal to one if the week falls in the endline period, i.e. one week or more after programme sessions led by i's facilitator have begun for the treatment group. γ_i is an individual fixed effect. The rest of the terms are defined as in Equation 28. Standard errors are again clustered at the individual level, since this was the level of randomisation.

Table 28 shows that the treatment leads to a significant increase in the number of sex acts per respondent per week. In the full endline period, respondents in the treatment group had sex 0.293 times per week more on average, compared to a control group mean of 0.793 times per week. In the last 30 and 14 days, the treatment effect on the number of sex acts per week was 0.415 and 0.521 respectively, (compared to 0.699 and 0.708 in the control group). All of these differences are significant at the 1% level. The fact that we observe an increase in the number of sex acts per week in the treatment group may suggest that couples bargain simultaneously over whether to have sex and which type of contraception to use. We plan to explore this in future work.

Table 29 shows the results for the proportion of sex acts protected per individual per week. In the full endline period, the treatment led to an increase of 9.7 percentage points in the proportion of sex acts per week protected by a female condom (over a control group mean of 1.0%). In the last 30 days the figure is 8.4 percentage points (compared to 1.0% in the control group), while the estimate for the last 14 days is not significant. Consistent with the

Table 28: Impacts on number of sex acts per respondent per week – diary subsample

	(1) **# sex acts per week full endline period** β / (s.e.)	(2) **# sex acts per week last 30 days** β / (s.e.)	(3) **# sex acts per week last 14 days** β / (s.e.)
Treat*endline	0.293**	0.415***	0.521***
	(0.128)	(0.133)	(0.171)
Individual f.e.'s	✓	✓	✓
Facilitator*endline f.e.'s	✓	✓	✓
Observations	850	531	364
Control mean	0.793	0.699	0.708

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are the number of sex acts per respondent per week. "Treat*endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat*endline" is the intent-to-treat effect. All regressions are respondent-level OLS fixed effects models with the respondent-week as the unit of observation. All specifications include individual fixed effects (N=56), and facilitator*endline fixed effects (N=17) for inference since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation.

Table 29: Impacts on % sex acts protected per respondent per week – diary subsample

	(1) **% fc-protected full endline period** β / (s.e.)	(2) **% mc-protected full endline period** β / (s.e.)	(3) **% fc-protected last 30 days** β / (s.e.)	(4) **% mc-protected last 30 days** β / (s.e.)	(5) **% fc-protected last 14 days** β / (s.e.)	(6) **% mc-protected last 14 days** β / (s.e.)
Treat*endline	0.097**	-0.095	0.084**	-0.017	0.034	-0.046
	(0.044)	(0.092)	(0.039)	(0.094)	(0.036)	(0.121)
Individual f.e.'s	✓	✓	✓	✓	✓	✓
Facilitator*endline f.e.'s	✓	✓	✓	✓	✓	✓
Observations	398	398	259	259	179	179
Control mean	0.010	0.392	0.010	0.419	0.012	0.443

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are the percentage of sex acts protected by female or male condoms per respondent per week. "Treat*endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat*endline" is the intent-to-treat effect. All regressions are respondent-level OLS fixed effects models with the respondent-week as the unit of observation. All specifications include individual fixed effects (N=56), and facilitator*endline fixed effects (N=17) for inference since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation.

respondent-level results from the survey data and with the overall sex-act-level results, there is no significant effect on the proportion of sex acts protected by male condoms.

Table 30 shows that we observe a large and highly significant reduction in the proportion of sex acts per week in which a discussion about condoms takes place: 19.4% of sex acts during the last 30 days and 35.6% of sex acts in the last 14 days respectively (compared to a control mean of 27.5% and 41.1% respectively). These results suggests that the availability of female condoms reduces the number of disagreements about the use of dual protection methods between partners. This is in line with our model, if the expansion from a binary to a ternary choice allows the couple to choose a dual protection technology that is closer to their preferred choice on the technological frontier.

3.5.3 *Heterogeneity by bargaining power*

We next examine which women adopt female condoms in terms of their bargaining power. Recall that the model predicts an inverse-U relationship between bargaining power and adoption of female condoms in the population. However, the relationship in our sample may be negative if women with the lowest bargaining power do not select into the condom programme; or positive if women with the highest bargaining power do not select in.

Table 31 displays the relationship between endline current use of female condoms and each of the baseline principal component measures of bargaining power and their square.[36] We control for baseline female condom use and for demographic factors which are significantly correlated with baseline bargaining power: namely age, education, income in the last thirty days, whether the respondent is in a stable relationship and whether she is the household head. This is to prevent the bargaining power measures from spuriously capturing the effect

36 Similar results to all of those results presented in this section are obtained for use in the last thirty days; available on request. We restrict attention to current use and use in the last 30 days, since ever use does not proxy sustained adoption.

Table 30: Impacts on % sex acts with discussions per respondent per week – diary subsample

	(1) **% sex acts with discussion full endline period** β / (s.e.)	(2) **% sex acts with discussion last 30 days** β / (s.e.)	(3) **% sex acts with discussion last 14 days** β / (s.e.)
Treat*endline	-0.100	-0.194*	-0.356***
	(0.108)	(0.109)	(0.120)
Individual f.e.'s	✓	✓	✓
Facilitator*endline f.e.'s	✓	✓	✓
Observations	398	259	179
Control mean	0.227	0.275	0.411

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are binary indicators for discussions and female-initiated discussions about condom use at the sex act level. "Treat*endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat*endline" is the intent-to-treat effect. All regressions are respondent level OLS fixed effects models with the respondent-week as the unit of observation. All specifications include individual fixed effects (N=56), and facilitator*endline fixed effects (N=17) for inference since randomisation was stratified on facilitator. Standard errors are clustered at the level of the individual, since this was the level of randomisation.

of variables which may be correlated with them and with condom use. For example, younger women may be less empowered but also more aware of the risks of unprotected sex, and use condoms for that reason. However, results are virtually unchanged if these controls are excluded (available on request). The overall pattern is clear: we observe a negative correlation between baseline bargaining power and endline use of female condoms, but not an inverse-U relationship.[37] Thus within our sample, it is the women with lower bargaining power who adopt female condoms.[38] This is in line with the evidence from Section 3.4.2 that women with low bargaining power do not select into our sample, and thus we may be capturing the downward-sloping portion of the inverse-U relationship predicted for the whole population. An exception is our measure of violence, which shows a positive correlation with use of female condoms and possibly an inverse-U shape, although the squared term is not significant. This again makes sense, given that according to table 21, the women in our sample are in the lower part of the distribution of bargaining power in terms of violence. Thus we may be capturing the upward-sloping portion of the inverse-U relationship between violence and female condom adoption predicted for the whole population.

We next test the model's prediction that among those women using male condoms at baseline, women with lower bargaining power will adopt female condoms. Table 32 reports the correlation (conditional on the same set of controls) between the different measures of bargaining power and current use of female condoms at endline, among those women who were using male condoms at baseline. We see consistent evidence of the negative correlation predicted by the model. On the other hand, as discussed in Section 3.5.1, we do not see evidence of a large degree of substitution away from male condoms. A possible explanation is if women with higher bargaining power who take up female condoms also intersperse their use with use of male condoms. Indeed, 81% of women who are currently using female condoms at endline

37 Table 46 in Appendix 3.7.3 shows similar results when each measure of bargaining power is interacted with the treatment variable. We lose considerable power when also interacting the square of each measure of bargaining power with the treatment variable (not shown). Facilitator fixed effects are not included, for comparability with Table 31 and since we lose power due to the low number of observations per cell if the sample is split along seventeen facilitators, treatment status, initial use, and bargaining power.

38 Similar results hold if instead of using the principal components we use a score for each module, summing an individual's responses; results available on request. Negative effects are also estimated for many of the individual questions about bargaining power taken separately, although only some are significant. However, neither of these approaches take into account the fact that different questions may be almost a repeat measure of the same construct.

Table 31: Current use of female condoms at endline, by baseline bargaining power

	(1) β / (s.e.)	(2) β / (s.e.)	(3) β / (s.e.)	(4) β / (s.e.)	(5) β / (s.e.)	(6) β / (s.e.)
Assets 1	-0.103***					
	(0.030)					
Assets 1 squared	0.016***					
	(0.005)					
Assets 2		-0.045***				
		(0.016)				
Assets 2 squared		0.004**				
		(0.002)				
Assets 3			-0.052***			
			(0.019)			
Assets 3 squared			0.004**			
			(0.002)			
Violence				0.112*		
				(0.060)		
Violence squared				-0.011		
				(0.009)		
Negative relation					-0.068	
					(0.117)	
Negative relation squared					0.007	
					(0.019)	
Decision-making						-0.192*
						(0.101)
Decision-making squared						0.038
						(0.025)
Controls	✓	✓	✓	✓	✓	✓
Observations	198	198	198	157	157	157
Control mean endline	0.020	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Excluding attritters, N=198 women are in a stable relationship at baseline. N=206 answer the assets module, including some women who live with their partner but do not report being in a stable relationship. N=160 have non-missing values for all of the decision-making, negative relation and violence questions (which is required for the principal components to be calculated) and all of the controls. "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table 23. "Decision-making", "Negative relation" and "Violence" are the first principal component from each of these modules, as identified in Table 24. Dependent variables are binary indicators for whether the respondent is currently using a female condom at endline. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Standard errors are robust to individual-level heteroskedasticity. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

also report currently using male condoms. This is a typical pattern of adoption observed in the medical literature, and is found to be associated with a large increase in the number of protected sex acts (Vijayakumar et al., 2006).

The model also predicts that among those women not using condoms at baseline, women with higher bargaining power will adopt. Section 3.5.1 showed evidence of strong adoption of female condoms among women who were not currently using male condoms at baseline. However, Table 33 shows that (conditional on the same set of controls) there is no evidence of a positive correlation between adoption and baseline bargaining power among these women. In contrast, the correlation between bargaining power and adoption is either insignificant or in a few cases negative. Again, this may suggest that the women with the lowest bargaining power do not select into our sample. Nonetheless, the fact that we observe women switching from sex unprotected by condoms to using female condoms remains promising for welfare.

3.5.4 *Heterogeneity by HIV status and risk perceptions*

As discussed in section 3.3.5, women who are HIV-positive and women who believe they face a high risk of infection are likely to place a stronger weight on the protection offered by contraceptive technologies, and thus to have higher demand for condoms. Indeed, a one point increase in the "risk perception for self" scale at baseline is correlated with an increase of 8.3 percentage points in the probability of ever having used male condoms at baseline (p-value 0.035).[39] Columns (1) and (2) of Table 34 show that women who report being HIV-positive at baseline indeed have a significantly larger treatment effect for "ever use of female condoms". Columns (3)-(6) show that HIV-positive women also have larger point estimates for the treatment effect on last thirty days and current female condom use, although neither are significantly different from the treatment effect for HIV-negative women. Table 47 in

39 The correlation with current use of male condoms at baseline is 4.7 percentage points, but this is marginally insignificant (p-value 0.146).

Table 32: Current use of female condoms at endline – women using male condoms at baseline

	(1) β / (s.e.)	(2) β / (s.e.)	(3) β / (s.e.)	(4) β / (s.e.)	(5) β / (s.e.)	(6) β / (s.e.)
Assets 1	-0.088					
	(0.055)					
Assets 1 squared	0.010					
	(0.010)					
Assets 2		-0.100*				
		(0.053)				
Assets 2 squared		0.009*				
		(0.005)				
Assets 3			-0.363**			
			(0.151)			
Assets 3 squared			0.072**			
			(0.034)			
Violence				0.160*		
				(0.096)		
Violence squared				-0.015		
				(0.015)		
Negative relation					-0.136	
					(0.169)	
Negative relation squared					0.013	
					(0.026)	
Decision-making						-0.255
						(0.208)
Decision-making squared						0.035
						(0.054)
Controls	✓	✓	✓	✓	✓	✓
Observations	75	75	75	60	60	60
Control mean endline	0.020	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^*$, $p < 0.05^{**}$, $p < 0.01^{***}$. The sample is restricted to women who report currently using male condoms at baseline, and who answer the assets module (columns 1-3) or all three survey modules on bargaining power (columns 4-6). "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table 23. "Decision-making", "Negative relation" and "Violence" are the first principal component from each of these modules, as identified in Table 24. Dependent variables are binary indicators for whether the respondent is currently using a female condom at endline. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table 33: Current use of female condoms at endline – women not using male condoms at baseline

	(1) β / (s.e.)	(2) β / (s.e.)	(3) β / (s.e.)	(4) β / (s.e.)	(5) β / (s.e.)	(6) β / (s.e.)
Assets 1	-0.082**					
	(0.035)					
Assets 1 squared	0.013**					
	(0.006)					
Assets 2		-0.063**				
		(0.029)				
Assets 2 squared		0.011**				
		(0.005)				
Assets 3			-0.031			
			(0.019)			
Assets 3 squared			0.002			
			(0.002)			
Violence				0.025		
				(0.090)		
Violence squared				-0.000		
				(0.013)		
Negative relation					0.010	
					(0.084)	
Negative relation squared					-0.002	
					(0.015)	
Decision-making						-0.138
						(0.098)
Decision-making squared						0.033
						(0.024)
Controls	✓	✓	✓	✓	✓	✓
Observations	123	123	123	97	97	97
Control mean endline	0.020	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^*$, $p < 0.05^{**}$, $p < 0.01^{***}$. The sample is restricted to women who report currently using male condoms at baseline, and who answer the assets module (columns 1-3) or all three survey modules on bargaining power (columns 4-6). "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table 23. "Decision-making", "Negative relation" and "Violence" are the first principal component from each of these modules, as identified in Table 24. Dependent variables are binary indicators for whether the respondent is currently using a female condom at endline. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table 34: Treatment effects, by baseline HIV status

	(1) **Ever FC HIV positive at baseline**	(2) **Ever FC HIV negative at baseline**	(3) **Last 30 days FC HIV positive at baseline**	(4) **Last 30 days FC HIV negative at baseline**	(5) **Current FC HIV positive at baseline**	(6) **Current FC HIV negative at baseline**
Treatment	0.392***	0.144***	0.120	0.047	0.200**	0.058
	(0.107)	(0.045)	(0.075)	(0.031)	(0.095)	(0.036)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	59	138	59	138	59	138
Control mean endline	0.133	0.067	0.000	0.017	0.033	0.017
$\chi^2(1) : (a) = (b)$		4.58		0.79		1.95
$Pr > \chi^2$		0.037		0.358		0.162

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms. Columns 1-2 refer to whether the respondent has used ever used female condoms, columns 3-4 to whether she has used them in the last 30 days, and columns 5-6 to whether she is currently them. Odd-numbered columns present results for the subsample of individuals who report being HIV-positive at baseline; even-numbered columns present results for the subsample who report being HIV-negative at baseline. The bottom two rows present chi-squared statistics and their p-values for the test that the treatment effect is the same across the two subsamples. These are obtained from seemingly unrelated estimations. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Appendix 3.7.3 shows that we do not observe significant treatment effects on male condom use either among HIV-positive or HIV-negative women.

As a proxy of women's beliefs about risk, respondents were asked to respond on a five-point scale to the question "what is the risk to you personally of contracting HIV/AIDS in a case of unprotected sex?".[40] We code a dummy variable "believes high risk to self" equal to one if the respondent's answer was above the median at baseline. Table 35 shows the results of re-estimating Equation 28 including this dummy and its interaction with treatment status. Columns (5) and (6) show that women with below-median perceptions of HIV risk at baseline increase their current use of both male and female condoms as a result of the treatment. Meanwhile women with above-median perceptions of HIV risk at baseline experience no treatment effect, as the negative interaction term completely offsets the main treatment effect for them.

A possible explanation is if women with low priors about the risk of unprotected sex are the ones who update their beliefs as a result of the programme, and thus change their behaviour accordingly. Column (7) of Table 35 provides suggestive evidence that this is the case. The treatment has a positive and significant effect on risk perceptions for women with below-median baseline risk perceptions. This is again completely offset by a negative interaction term for women whose baseline risk beliefs were above the median, suggesting the programme did not shift their beliefs at all. However, we caution against over-interpretation of this result, since the median baseline risk perception was "risky" (four) and so women who had above-median perceptions already answered the maximum of "very risky" (five). Thus even if the programme did increase these women's perceptions of risk, we would not capture this in our measure.[41]

40 As a framing question, respondents are first asked about the probability of a woman in general contracting HIV from unprotected sex. The pattern of results is similar when using this variable; tables available on request.

41 We also see a significant treatment effect of 14.9 percentage points (p-value 0.022) on the proportion of women agreeing or strongly agreeing with the statement "my partner is involved with other people apart from me". This may also proxy increased risk perceptions. When we analyse the continuous version of this variable, we cannot reject that women with high and low baseline beliefs increase their belief equally in response to treatment.

Table 35: Treatment effects – heterogeneity by baseline HIV risk beliefs

	(1) **Ever FC**	(2) **Ever MC**	(3) **Last 30 days FC**	(4) **Last 30 days MC**	(5) **Current FC**	(6) **Current MC**	(7) **Risk perception for self Score 1-5**
Treatment	0.208**	-0.002	0.072	0.078	0.159***	0.392***	0.427**
	(0.085)	(0.079)	(0.065)	(0.112)	(0.061)	(0.108)	(0.212)
Believes high risk to self	-0.085	0.019	-0.045	0.067	0.010	0.208**	
	(0.060)	(0.063)	(0.039)	(0.086)	(0.024)	(0.085)	
Treatment*Believes high risk to self	-0.032	-0.014	-0.034	-0.177	-0.111	-0.454***	-0.483**
	(0.097)	(0.094)	(0.068)	(0.129)	(0.068)	(0.126)	(0.236)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓	✓
Observations	227	227	227	227	227	227	226
Control mean endline	0.088	0.824	0.010	0.363	0.020	0.353	4.471

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female and male condoms. "Believes high risk to self" is an indicator equal to one if the respondent's answer to the question "what is your risk of contracting HIV/AIDS in a case of unprotected sex?" was above the median on a 1-5 scale. In practice this corresponds to an answer of "5, very risky" since the median response was "4, risky". All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17), since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Finally, Table 48 in Appendix 3.7.3 shows that heterogeneity by bargaining power is not proxying the heterogeneity by risk beliefs: both effects remain strong and significant when controlled for simultaneously.

3.5.5 *Effects on knowledge, STIs and well-being*

As described in Section 3.5.2, we observe a significant increase in the number of sex acts per week and a significant decrease in the number of discussions or arguments. Both of these results suggest positive impacts on welfare. Table 36 shows that we observe almost no treatment effect on other secondary impact measures, although some are likely underpowered. Treated respondents score slightly lower on a set of six questions about how HIV can be transmitted, although this is only marginally significant (p=0.095) and baseline scores are already high. The treatment group are no less likely to report being HIV-positive or having an STI. However, we do not place too much weight on this result, since it is unlikely that many respondents got newly tested for HIV or STIs in the time window between the start of the programme and the endline survey (four months). There is also no reduction in the likelihood of being pregnant, although we observe only a handful of pregnancies at baseline and endline, given that not being pregnant was a restriction for being accepted into the study sample.

Reassuringly, we see no negative effects of the treatment on self-reported well-being, as measured through a survey module with twelve separate indicators. This is in contrast to the findings of Ashraf et al. (2014a) who see a decrease in women's reported well-being after being assigned to receive concealable contraceptives. The difference may reflect the fact that condom use is ultimately something agreed on and observed by both partners, rather than something hidden and conducive to moral hazard. Nor do we observe any adverse effects on reports of physical or emotional violence, as measured through five separate indicators.

Table 36: Treatment effects – other outcome variables

	(1) **HIV knowledge (score 0-6)**	(2) **HIV positive**	(3) **STI last 3 months**	(4) **Well-being (score 0-12)**	(5) **Violence (score 0-5)**
Treatment	-0.136*	-0.021	0.021	0.171	0.076
	(0.081)	(0.042)	(0.033)	(0.260)	(0.195)
Facilitator f.e.'s	✓	✓	✓	✓	✓
Observations	219	196	185	212	162
Control mean endline	5.758	0.313	0.054	8.135	1.149

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are as follows, all measured at endline: column 1, a score from six questions testing knowledge about how HIV can and cannot be transmitted; column 2, a self-reported dummy for HIV-positive status; column 3, a self-reported dummy for having had an STI in the last three months; column 4, a score from twelve questions on well-being (higher scores indicate greater well-being); column 5 a score from five questions about emotional and physical violence (a higher score indicates greater violence). N=232 for the endline survey excluding attritters, except the violence module where N=162 since these questions were only enumerated to women in a stable relationship. Missing observations reflect not applicable, does not want to answer, and cases where the facilitator fixed effect perfectly predicts the outcome variable. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17), since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

This may reflect the careful design of the programme, to mitigate the possibility that women suggesting new contraceptive use might face backlash from their partner.

3.5.6 *Robustness*

3.5.6.1 *Bounding treatment effects for attrition*

As mentioned in Section 3.4.1, attrition is significantly higher in the control group. For robustness we therefore conduct a bounds analysis on the main treatment effects following Lee (2009). Table 49 in Appendix 3.7.3 details the results. Whilst the lower bounds cannot rule out a treatment effect of zero for the various measures of female condom use, these bounds do not include facilitator fixed effects, since whether attrition is higher in the treatment or control group varies by facilitator and thus the monotonicity assumption required for Lee bounds fails. We are able to rule out any sizeable negative effects, and the upper bounds are large and highly significant.

3.5.6.2 *Diary robustness checks*

As robustness checks, we also re-estimate the individual-level impacts using the diary data. First, to check that the diary respondents are representative of the full sample in terms of the impacts estimated from the survey data, we repeat estimation of Equation 28 with the survey data but only for diary respondents. Table 50 in Appendix 3.7.3 confirms that the results are not different from the main sample, although the result for female condom use in the last 30 days loses significance due to the loss of sample size.

Next, we re-estimate impact using the main outcome variables as measured from the diary data. We first use the linear probability fixed effects panel model from Equation 29.[42] Table 51 in Appendix 3.7.3 show the results. Again, we see significant impacts for the use of female condoms during the full endline period and the last 30 days, while we see no significant impacts on male condom use and on female condom use in the last fourteen days. We also estimate ANCOVA specficiations, for comparability with the main results.[43] Table 52 in Appendix 3.7.3 shows the results. Consistent with the survey data, we see significant and positive impacts on ever use and use in the last 30 days of female condoms. The impact on use of female condoms in the last 14 days is no longer significant, due to the loss of sample size. We see no significant impacts for the use of male condoms.

42 Again, similar results are obtained with a logit specification; results available on request.

43 Since the nature of the diary data is such that baseline observations are missing for some respondents, we follow McKenzie (2015) and estimate:

$$\Pr[Y_{if1} = 1 | Y_{if0}, treat_{if}, \eta_f] = \alpha + \delta Y_{if0} + \beta_1 treat_{if} + \gamma missbase_{if} + \sigma missbase_{if} * Y_{if0} + \eta_f \quad (30)$$

where Y_{if1} is the outcome variable of interest for the endline period. Y_{if0} is the value of the outcome variable for the baseline period. $missbase_{if}$ is a dummy equal to one if the respondent is missing the value of the outcome variable during the baseline period. $missbase_{if} * Y_{if0}$ then sets the baseline value to zero in the case that it is missing. Inclusion of this dummy means that δ is estimated only for respondents whose baseline data is not missing. $treat_{if}$ is again a dummy for being in the treatment group, and η_f is again a facilitator fixed effect. β_1 represents the intent-to-treat effect, this time as estimated on the diary subsample.

3.6 CONCLUSION

In terms of implications for policies to combat HIV/AIDS and other STIs, our results imply that female condoms may be the only viable dual protection method for some women with low bargaining power. Thus for policymakers considering whether to subsidise distribution of female condoms, the correct cost comparison in contexts where women have low bargaining power is not to male condoms but rather to antiretroviral therapies and other costs associated with unprotected sex. Including female condoms into sex education programmes could also be a welfare-improving policy with low marginal cost, avoiding the fixed costs of setting up a separate condom intervention like that in this study. To aid more comprehensive welfare calculations and funding decisions, it would be useful for future studies to determine individual men and women's willingness-to-pay for male and female condoms, once they have tried both technologies. Evidence on longer-term adoption of female condoms would also be welcome, given that our study and studies in the medical literature have tended to focus on the first three to six months.

More broadly, this paper has highlighted how low female bargaining power may lead to under-adoption of technologies that improve household welfare, in cases where women have a stronger preference for adoption or face higher costs of non-adoption. There are many other examples of welfare-improving technologies where women may have a stronger willingness to adopt. For instance, women may have a higher demand for insurance, given evidence that they are more risk-averse. Women may also have a higher willingness-to-pay for household sanitation technologies such as private toilets, since women face larger stigma and risks to their personal safety from using facilities outside of the household. Women may also have a stronger preference for delaying the marriage and childbearing of adolescent girls, if women's preferences for lower fertility and higher birth spacing apply to their children. In each of these cases, information and social norm campaigns targeted specifically at men may be necessary to increase welfare-improving investments and adoption (Stopnitzky, 2017). If such campaigns still cannot resolve under-adoption, then providing versions of the technology that

are more acceptable to men, or bundling technologies with goods for which men have a high demand, may offer a second-best solution. These remain important topics for future research.

3.7 APPENDICES

3.7.1 *Proofs*

Proof that the couple's optimal choice of H is increasing in α:

The first order condition for the maximisation problem given in Equation 16 is:

$$V_P\left(P\left(H\right), H; \alpha\right) P'\left(H\right) + V_H\left(P\left(H\right), H; \alpha\right) = 0. \tag{31}$$

The second order condition is:

$$\left[V_{PP} P' + V_{PH}\right] P' + V_P P'' + V_{HP} P' + V_{HH} < 0. \tag{32}$$

Given weak concavity of the production frontier, $P' < 0$ and $P'' \leqslant 0$.. Equation 32 is therefore satisfied for any strictly quasi-concave preferences such that $V_{PP} < 0$, $V_{HH} < 0$, $V_{PH} = V_{HP} = 0$. This includes the Cobb-Douglas preferences given in Equation 14.

Differentiating the first order condition with respect to α yields:

$$\left\{\left[V_{PP} P'\left(H\right) + V_{PH}\right] P' + V_P P'' + V_{HP} P' + V_{HH}\right\} H_\alpha + V_{P\alpha} P' + V_{H\alpha} = 0 \tag{33}$$

which rearranges to

$$H_\alpha = -\frac{V_{P\alpha} P' + V_{H\alpha}}{\left[V_{PP} P'\left(H\right) + V_{PH}\right] P' + V_P P'' + V_{HP} P' + V_{HH}} \tag{34}$$

Note that if $\beta^m > \beta^f$, $V_{P\alpha} < 0$. Similarly, if $\beta^m > \beta^f$ then $V_{H\alpha} > 0$. Given that that $P_H < 0$, it follows that $V_{P\alpha} P_H + V_{H\alpha}$ is positive. Hence the numerator of Equation 34 is positive. The denominator of Equation 34 is negative by the second order condition. Hence $H_\alpha > 0$. *QED*.

3.7.2 *Additional descriptive data*

Table 37: Baseline balance excluding attritters

	Mean	Control Mean	Treatment Mean	t-test	Total N	Control N	Treatment N
Demographics							
Age	30.80	30.65	30.93	-0.25	232	107	125
Years of education	6.30	6.36	6.25	0.27	231	107	124
Literate	0.86	0.86	0.86	-0.10	230	106	124
HH head	0.23	0.21	0.25	-0.59	232	107	125
Income							
Has job	0.37	0.41	0.34	1.10	229	105	124
Personal income last 30 days	813.15	916.36	724.80	1.04	232	107	125
Relationships							
In a stable relationship (incl. married)	0.85	0.85	0.86	-0.12	232	107	125
Married (officially or unofficially)	0.62	0.62	0.62	-0.06	231	107	124
Years relation	9.23	9.32	9.16	0.14	184	84	100
# Partners last 12 months	0.93	0.92	0.94	-0.58	232	107	125
Sexual behaviour							
Pregnant	0.06	0.05	0.07	-0.80	231	106	125
HIV positive (self-report)	0.30	0.34	0.27	1.01	202	95	107
STD last 3 months (self-report)	0.14	0.15	0.12	0.58	205	92	113
Names FC as contraceptive	0.43	0.46	0.40	0.91	230	107	123
Contraceptive use							
Ever used FC	0.08	0.07	0.08	-0.15	232	107	125
Ever used MC	0.74	0.77	0.71	0.94	232	107	125
Ever used other	0.72	0.71	0.74	-0.43	232	107	125
Never used any	0.07	0.08	0.06	0.58	232	107	125
Used FC last 30 days	0.03	0.02	0.04	-0.97	232	107	125
Used MC last 30 days	0.33	0.30	0.36	-0.98	232	107	125
Current use FC	0.03	0.03	0.03	-0.18	232	107	125
Current use MC	0.38	0.36	0.39	-0.43	232	107	125
Current use other	0.38	0.38	0.38	0.11	232	107	125
Current use none	0.37	0.38	0.35	0.49	232	107	125

Notes: N=232 in the balanced sample excluding attritters. Lower sample sizes reflect observations that are missing or not applicable. "Treatment Mean" contains all individuals assigned to the treatment group, whether or not they attended the sessions. Column 4 presents the test statistic for the null hypothesis that the mean in the treatment group is equal to the mean in the control group. "Ever used other" and "Current use other" refer to use of any other modern contraceptive method apart from condoms, e.g. the pill, injectables, or an IUD.

Table 38: Baseline balance on covariates – diary subsample

	Mean	Control Mean	Treatment Mean	t-test	Total N	Control N	Treatment N
Demographics							
Age	31.40	31.93	30.86	0.48	57	29	28
Education	5.91	5.45	6.39	-1.22	57	29	28
Literate	0.84	0.86	0.82	0.36	56	28	28
HH head	0.30	0.21	0.39	-1.53	57	29	28
Income							
Has job	0.37	0.38	0.36	0.17	57	29	28
Personal income last 30 days	987.72	927.59	1050.00	-0.30	57	29	28
Relationships							
In a stable relationship (incl. married)	0.82	0.83	0.82	0.06	57	29	28
Married (officially or unofficially)	0.53	0.59	0.46	0.91	57	29	28
Years relation	11.78	12.82	10.58	0.86	41	22	19
# Partners last 12 months	0.89	0.86	0.93	-0.69	57	29	28
Sexual behaviour							
Pregnant	0.00	0.00	0.00	0.00	57	29	28
HIV positive (self-report)	0.33	0.38	0.28	0.70	49	24	25
STD last 3 months (self-report)	0.12	0.12	0.12	0.05	49	24	25
Names FC as contraceptive	0.27	0.31	0.22	0.74	56	29	27

Notes: N=57 in the balanced panel. Lower sample sizes reflect observations that are missing or not applicable. "Treatment Mean" contains all individuals assigned to the treatment group, whether or not they attended the sessions. Column 4 presents the t-test statistic for the null hypothesis that the mean in the treatment group is equal to the mean in the control group.

Table 39: Baseline balance on use – diary subsample

	Mean	Control Mean	Treatment Mean	t-test	Total N	Control N	Treatment N
Ever use survey							
Ever used FC	0.05	0.07	0.04	0.56	57	29	28
Ever used MC	0.77	0.79	0.75	0.38	57	29	28
Ever used other	0.84	0.86	0.82	0.41	57	29	28
Use last 30 days survey							
Used FC last 30 days	0.04	0.00	0.07	-1.44	57	29	28
Used MC last 30 days	0.35	0.28	0.43	-1.20	57	29	28
Current use survey							
Current use FC	0.02	0.00	0.04	-1.00	57	29	28
Current use MC	0.40	0.38	0.43	-0.37	57	29	28
Current use other	0.47	0.41	0.54	-0.91	57	29	28
Current use none	0.26	0.34	0.18	1.43	57	29	28
Ever use diaries							
Used FC in baseline weeks	0.04	0.03	0.04	-0.02	57	29	28
Used MC in baseline weeks	0.44	0.41	0.46	-0.38	57	29	28
Use last 30 days							
Used FC in baseline last 30 days	0.04	0.03	0.04	-0.02	57	29	28
Used MC in baseline last 30 days	0.53	0.52	0.54	-0.14	57	29	28
Use last 14 days							
Used FC in baseline last 14 days	0.04	0.03	0.04	-0.02	57	29	28
Used MC in baseline last 14 days	0.39	0.34	0.43	-0.64	57	29	28
Impact on discussions and sex acts							
% sex acts with discussion about condom use	0.16	0.13	0.18	-0.57	57	29	28
% sex acts with female-initiated discussion about condom use	0.12	0.07	0.18	-1.50	57	29	28
# sex acts per week	0.89	0.89	0.88	0.05	56	28	28

Notes: N=57 in the balanced panel. Lower sample sizes reflect observations that are missing or not applicable. "Treatment Mean" contains all individuals assigned to the treatment group, whether or not they attended the sessions. Column 4 presents the test statistic for the null hypothesis that the mean in the treatment group is equal to the mean in the control group. "Ever used other" and "Current use other" refer to use of any other modern contraceptive method apart from condoms, e.g. the pill, injectables, or an IUD. Percentage use represents the percentage of sex acts per individual which were protected by female condoms, male condoms and other contraceptives respectively.

Table 40: Diary sample representativeness of full sample – covariates

	Mean	Survey Mean	Diary sub-sample Mean	t-test	Survey N	Diary subsample N
Demographics						
Age	30.95	30.74	31.92	-0.90	231	50
Years of education	6.21	6.30	5.76	1.17	230	50
Literate	0.85	0.86	0.82	0.73	229	49
HH head	0.25	0.23	0.34	-1.45	231	50
Income						
Has job	0.38	0.37	0.42	-0.61	228	50
Personal income last 30 days	871.71	816.67	1126.00	-1.27	231	50
Relationships						
In a stable relationship (incl. married)	0.85	0.85	0.84	0.22	231	50
Married (officially or unofficially)	0.60	0.62	0.50	1.50	230	50
Years relation	9.70	9.20	12.19	-1.97*	183	37
# Partners last 12 months	0.93	0.93	0.92	0.20	231	50
Sexual behaviour						
Pregnant	0.05	0.06	0.00	3.85***	230	50
HIV positive (self-report)	0.31	0.30	0.35	-0.56	201	43
STI last 3 months (self-report)	0.13	0.14	0.12	0.38	204	43
Names FC as contraceptive	0.41	0.43	0.31	1.64	229	49
Bargaining power (principle components)						
Decision-making	1.74	1.69	1.98	-1.58	160	34
Negative relation	2.92	2.93	2.87	0.36	160	34
Violence	3.45	3.43	3.54	-0.51	160	34
Assets 1	0.79	0.80	0.78	0.09	204	43
Assets 2	0.48	0.48	0.47	0.05	204	43
Assets 3	1.21	1.23	1.14	0.70	204	43

Notes: N=232 in the balanced panel, of which N=57 are in the subsample who respond to the diaries. Lower sample sizes reflect observations that are missing or not applicable. "Survey Mean" contains all individuals in the balanced panel, whether or not they participated in the diaries. "Diary subsample Mean" contains just those individuals who responded to the diaries. Column 4 presents the t-test statistic for the null hypothesis that the mean in the diary subsample is equal to the mean in the full sample.

Table 41: Diary sample representativeness of full sample – baseline use

	Mean	Survey Mean	Diaries Mean	t-test	Survey N	Diaries N
Ever use survey						
Ever used FC	0.08	0.09	0.05	0.94	297	55
Ever used MC	0.75	0.75	0.78	-0.56	297	55
Ever used other	0.74	0.72	0.85	-2.45**	297	55
Use last 30 days survey						
Used FC last 30 days	0.03	0.03	0.04	-0.35	297	55
Used MC last 30 days	0.32	0.32	0.36	-0.67	297	55
Current use survey						
Current use FC	0.02	0.02	0.02	0.27	297	55
Current use MC	0.39	0.39	0.40	-0.08	297	55
Current use other	0.41	0.39	0.49	-1.36	297	55

Notes: N=57 are in the subsample of survey respondents who respond to the diaries. N=298 are in the full baseline study sample. Lower sample sizes reflect observations that are missing. All values taken from the baseline survey. "Ever used other" and "Current use other" refer to use of any other modern contraceptive method apart from condoms, e.g. the pill, injectables, or an IUD. Column 4 presents the t-test statistic for the null hypothesis that the mean in the diary subsample is equal to the mean in the full sample.

Table 42: Balance – principal components of bargaining power

	Mean	Control Mean	Treatment Mean	t-test	Total N	Control N	Treatment N
Assets 1	0.76	0.81	0.71	0.82	263	128	135
Assets 2	0.43	0.41	0.45	-0.33	263	128	135
Assets 3	1.18	1.24	1.13	0.94	263	128	135
Violence	3.46	3.44	3.48	-0.32	206	99	107
Negative relation	2.83	2.84	2.81	0.24	206	99	107
Decision-making	1.79	1.87	1.71	1.17	206	99	107

Notes: All values taken from the baseline survey. "Assets 1", "Assets 2" and "Assets 3" represent the three principal components loaded by the assets module as identified in Table 23. "Decision-making", "Negative relation" and "Violence" represent the principal component loaded by each of these survey modules as identified in Table 24. All variables are coded such that a higher value proxies greater bargaining power for the respondent. All components are scaled such that the least empowered woman on that component has a score of zero. They are also normalised such that a one point increase in each component represents an increase of one standard deviation. "Treatment Mean" contains all individuals assigned to the treatment group, whether or not they attended the sessions. Column 4 presents the test statistic for the null hypothesis that the mean in the treatment group is equal to the mean in the control group. All values taken from the baseline survey (N=298). The assets module was enumerated to all women who lived with their partner at baseline, including a few who did not claim to be in a stable relationship (N=264). The decision-making module was enumerated to all respondents (N=298), except the questions "who has more influence" and "who has more power" which were asked only of women in a stable relationship at baseline (N=250). Negative relationship and emotional and physical violence questions were also only asked of to women who were in a stable relationship at baseline (N=250). Any lower sample sizes reflect values missing or unwillingness to answer.

3.7.3 *Additional analyses*

Table 43: Treatment effects on male condom use, by baseline male condom use

	(1) **Ever MC No MC at baseline**	(2) **Ever MC Current MC at baseline**	(3) **Last 30 days MC No MC at baseline**	(4) **Last 30 days MC Current MC at baseline**	(5) **Current MC No MC at baseline**	(6) **Current MC Current MC at baseline**
Treatment	-0.000	0.002	0.032	-0.158	0.101	-0.002
	(0.059)	(0.053)	(0.065)	(0.097)	(0.068)	(0.105)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	141	86	141	86	141	86
Control mean endline	0.754	0.946	0.231	0.595	0.215	0.595
$\chi^2(1) : (a) = (b)$		0.00		2.65		0.68
$Pr > \chi^2$		0.973		0.103		0.411

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of male condoms. Columns 1-2 refer to whether the respondent has ever used that method, columns 3-4 to whether she has used it in the last 30 days, and columns 5-6 to whether she is currently using it. Odd-numbered columns present results for the subsample of individuals not currently using male condoms at baseline; even-numbered columns present results for the subsample of individuals who are currently using male condoms at baseline. The bottom two rows present chi-squared statistics and their p-values for the test that the treatment effect is the same across the two subsamples. These are obtained from seemingly unrelated estimations. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table 44: Treatment effects on female condom use, by baseline contraceptive use

	(1) **Ever FC No method at baseline**	(2) **Ever FC Some method at baseline**	(3) **Last 30 days FC No method at baseline**	(4) **Last 30 days FC Some method at baseline**	(5) **Current FC No method at baseline**	(6) **Current FC Some method at baseline**
Treatment	0.216***	0.168***	0.094**	0.023	0.125**	0.044
	(0.066)	(0.051)	(0.042)	(0.025)	(0.050)	(0.037)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	84	143	84	143	84	143
Control mean endline	0.100	0.081	0.000	0.016	0.000	0.032
$\chi^2(1) : (a) = (b)$		0.32		2.10		1.64
$Pr > \chi^2$		0.570		0.147		0.200

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms. Columns 1-2 refer to whether the respondent has ever used that method, columns 3-4 to whether she has used it in the last 30 days, and columns 5-6 to whether she is currently using it. Odd-numbered columns present results for the subsample of individuals not currently using any contraceptive at baseline; even-numbered columns present results for the subsample of individuals who are currently using some contraceptive method at baseline. The bottom two rows present chi-squared statistics and their p-values for the test that the treatment effect is the same across the two subsamples. These are obtained from seemingly unrelated estimations. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table 45: Treatment effects – heterogeneity by relationship status

	(1) **Ever FC**	(2) **Ever MC**	(3) **Last 30 days FC**	(4) **Last 30 days MC**	(5) **Current FC**	(6) **Current MC**
Treatment	0.358***	-0.089	0.040	0.061	0.165*	0.179
	(0.103)	(0.112)	(0.054)	(0.153)	(0.088)	(0.150)
Stable relationship	0.030	-0.038	0.007	-0.052	0.024	-0.064
	(0.051)	(0.078)	(0.020)	(0.120)	(0.024)	(0.109)
Treat*Stable relationship	-0.202*	0.090	0.009	-0.132	-0.102	-0.141
	(0.109)	(0.121)	(0.064)	(0.166)	(0.093)	(0.162)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	227	227	220	221	227	227
Control mean endline	0.088	0.824	0.010	0.366	0.020	0.353

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. "Stable relationship" is a dummy equal to one if the respondent reports being in a stable relationship at baseline. Dependent variables are binary indicators for the current use of female and male condoms. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17) since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table 46: Impacts on current use of female condoms – heterogeneity by bargaining power

	(1) β / (s.e.)	(2) β / (s.e.)	(3) β / (s.e.)	(4) β / (s.e.)	(5) β / (s.e.)	(6) β / (s.e.)
Treatment	0.128***	0.111***	0.159***	-0.174*	0.220*	0.227**
	(0.045)	(0.039)	(0.051)	(0.095)	(0.119)	(0.101)
Assets 1	-0.002					
	(0.010)					
Treat*Assets1	-0.044**					
	(0.017)					
Assets 2		0.004				
		(0.010)				
Treat*Assets2		-0.031**				
		(0.015)				
Assets 3			-0.001			
			(0.005)			
Treat*Assets3			-0.054***			
			(0.018)			
Violence				-0.007		
				(0.008)		
Treat*Violence				0.076**		
				(0.031)		
Negative relation					-0.002	
					(0.009)	
Treat*Negative relation					-0.043	
					(0.039)	
Decision-making						-0.021
						(0.025)
Treat*Decision						-0.079*
						(0.042)
Controls	✓	✓	✓	✓	✓	✓
Observations	198	198	198	157	157	157
Control mean endline	0.020	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Excluding attritters, N=198 women are in a stable relationship at baseline. N=206 answer the assets module, including some women who live with their partner but do not report being in a stable relationship. N=160 have non-missing values for all of the decision-making, negative relation and violence questions (which is required for the principal components to be calculated) and all of the controls. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table 23. "Decision-making", "Negative relation" and "Violence" are the first principal component from each of these modules, as identified in Table 24. Dependent variables are binary indicators for current use of female condoms. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Regressions do not include facilitator fixed effects due to loss of sample size where baseline use perfectly predicts endline use conditional on a given facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table 47: Impacts on male condom use – heterogeneity by baseline HIV status

	(1) **Ever MC HIV positive at baseline**	(2) **Ever MC HIV negative at baseline**	(3) **Last 30 days MC HIV positive at baseline**	(4) **Last 30 days MC HIV negative at baseline**	(5) **Current MC HIV positive at baseline**	(6) **Current MC HIV negative at baseline**
Treatment	0.055	0.007	-0.034	-0.025	0.094	0.056
	(0.110)	(0.048)	(0.134)	(0.072)	(0.141)	(0.075)
Facilitator f.e.'s	✓	✓	✓	✓	✓	✓
Observations	59	138	59	138	59	138
Control mean endline	0.833	0.833	0.467	0.333	0.467	0.333
$\chi^2(1) : (a) = (b)$		0.16		0.00		.
$Pr > \chi^2$		0.692		0.951		.

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of male condoms. Columns 1-2 refer to whether the respondent has used ever used male condoms, columns 3-4 to whether she has used them in the last 30 days, and columns 5-6 to whether she is currently using them. Odd-numbered columns present results for the subsample of individuals who report being HIV-positive at baseline; even-numbered columns present results for the subsample who report being HIV-negative at baseline. The bottom two rows present chi-squared statistics and their p-values for the test that the treatment effect is the same across the two subsamples. These are obtained from seemingly unrelated estimations. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. All regressions include facilitator fixed effects (N=17), since randomisation was stratified on facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation.

Table 48: Treatment effects – heterogeneity by baseline bargaining power & risk beliefs

	(1) β / (s.e.)	(2) β / (s.e.)	(3) β / (s.e.)	(4) β / (s.e.)	(5) β / (s.e.)	(6) β / (s.e.)
Treatment	0.242***	0.209***	0.264***	0.298**	0.238**	0.039
	(0.084)	(0.078)	(0.087)	(0.135)	(0.115)	(0.119)
Believes high risk to self	-0.027	-0.023	-0.032	0.002	-0.013	-0.012
	(0.042)	(0.042)	(0.043)	(0.043)	(0.043)	(0.042)
Treatment*Believes high risk to self	-0.154**	-0.143*	-0.148*	-0.151*	-0.117	-0.122
	(0.078)	(0.078)	(0.078)	(0.081)	(0.087)	(0.081)
Treat*Assets1	-0.054**					
	(0.023)					
Assets 1	0.004					
	(0.011)					
Treat*Assets2		-0.038**				
		(0.017)				
Assets 2		0.008				
		(0.011)				
Treat*Assets3			-0.059**			
			(0.024)			
Assets 3			-0.001			
			(0.006)			
Treat*Decision				-0.062		
				(0.048)		
Decision-making				-0.034		
				(0.027)		
Treat*Negative relation					-0.023	
					(0.036)	
Negative relation					-0.010	
					(0.013)	
Treat*Violence						0.038
						(0.030)
Violence						0.012
						(0.013)
Controls	✓	✓	✓	✓	✓	✓
Observations	197	197	197	156	156	156
Control mean endline	0.020	0.020	0.020	0.020	0.020	0.020

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. N=198 women are in a stable relationship at baseline. N=160 have non-missing values for all of the decision-making, negative relation and violence questions (required for the principal components to be calculated) and all of the controls. N=206 answer the assets module, including some women who live with their partner but do not report being in a stable relationship. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. "Decision-making", "Negative relation" and "Violence" are the first principal component from each of these modules, as identified in Table 24. "Assets 1", "Assets 2" and "Assets 3" are the first three principal components from the assets module, as identified in Table 23. Dependent variables are binary indicators for the current use of female condoms. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Regressions do not include facilitator fixed effects due to loss of sample size where baseline use perfectly predicts endline use conditional on a given facilitator. Standard errors are robust to individual-level heteroskedasticity, since this was the level of randomisation. "Believes high risk to self" is an indicator equal to one if the respondent's answer to the question "what is your risk of contracting HIV/AIDS in a case of unprotected sex?" was above the median on a 1-5 scale. In practice this corresponds to an answer of "5, very risky" since the median response was "4, risky". Controls are: respondent's age, education, and income in the last 30 days; whether the respondent has a job, is married or in a stable relationship, and whether the respondent is the household head.

Table 49: Lee bounds – primary outcome variables

	(1) **Ever FC**	(2) **Ever MC**	(3) **Ever other**	(4) **Last 30 days FC**	(5) **Last 30 days MC**	(6) **Current FC**	(7) **Current MC**	(8) **Current other**
Upper	0.223***	0.090	0.169*	0.066**	0.005	0.112***	0.137*	0.106
	(0.058)	(0.084)	(0.086)	(0.028)	(0.073)	(0.037)	(0.077)	(0.079)
Lower	0.045	-0.087	-0.008	-0.010	-0.173**	-0.020	-0.040	-0.071
	(0.079)	(0.061)	(0.065)	(0.010)	(0.087)	(0.014)	(0.084)	(0.084)
95% C.I. Upper bound	0.318	0.228	0.311	0.112	0.124	0.173	0.264	0.237
95% C.I. Lower bound	-0.084	-0.187	-0.115	-0.026	-0.315	-0.042	-0.178	-0.209
Proportion trimmed	0.150	0.150	0.150	0.150	0.150	0.150	0.150	0.150
Observations	298	298	298	298	298	298	298	298
Control mean endline	0.088	0.824	0.735	0.010	0.363	0.020	0.353	0.412

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms, male condoms and other modern contraceptive methods. Columns 1-3 refer to whether the respondent has ever used the method, columns 4 and 5 to whether she has used it in the last 30 days (this was only asked for condoms, not for other contraceptive methods), and columns 6-8 to whether she is currently using it. Bounds do not include facilitator fixed effects, as attrition is not monotonic on treatment status conditional on facilitator fixed effects.

Table 50: Treatment effects – survey variables, diary subsample

	(1) Ever FC	(2) Ever MC	(3) Ever other	(4) Last 30 days FC	(5) Last 30 days MC	(6) Current FC	(7) Current MC	(8) Current other
Treatment	0.249***	-0.040	0.040	0.087	-0.107	0.125*	0.069	0.068
	(0.094)	(0.093)	(0.066)	(0.059)	(0.132)	(0.068)	(0.133)	(0.108)
Observations	50	50	50	50	50	50	50	50
Control mean endline	0.088	0.824	0.735	0.010	0.363	0.020	0.353	0.412

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are binary indicators for the use of contraceptives at the level of the respondent. "Treatment" is an indicator for observations in the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. All regressions are linear probability models with the respondent as the unit of observation. Facilitator fixed effects are dropped because facilitator perfectly predicts outcomes for many observations in this subsample. Standard errors are clustered at the individual level, since this was the level of randomisation.

Table 51: Impacts on condom use – fixed effects panel estimator, diary subsample

	(1) **FC full endline period** β / (s.e.)	(2) **MC full endline period** β / (s.e.)	(3) **FC last 30 days** β / (s.e.)	(4) **MC last 30 days** β / (s.e.)	(5) **FC last 14 days** β / (s.e.)	(6) **MC last 14 days** β / (s.e.)
Treat*endline	0.120**	-0.148	0.123**	-0.033	0.054	-0.099
	(0.054)	(0.093)	(0.050)	(0.103)	(0.045)	(0.113)
Individual f.e.'s	✓	✓	✓	✓	✓	✓
Facilitator*endline f.e.'s	✓	✓	✓	✓	✓	✓
Observations	383	383	252	252	175	175
Control mean	0.020	0.350	0.015	0.374	0.021	0.412

Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. Dependent variables are binary indicators for male and female condom use at the sex act level, measured each week. "Treat*endline" is an indicator for observations in the treatment group during the endline period. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treat*endline" is the intent-to-treat effect. All regressions are respondent level OLS fixed effects models with the sex act as the unit of observation. All specifications include individual fixed effects (N=57) and facilitator*endline fixed effects (N=17) since randomisation was stratified on facilitator.

Table 52: Impacts on condom use – ANCOVA specification, diary subsample

	(1) **FC full endline period** β / (s.e.)	(2) **MC full endline period** β / (s.e.)	(3) **FC last 30 days** β / (s.e.)	(4) **MC last 30 days** β / (s.e.)	(5) **FC last 14 days** β / (s.e.)	(6) **MC last 14 days** β / (s.e.)
Treatment	0.169*	-0.135	0.139*	0.005	0.069	-0.099
	(0.088)	(0.132)	(0.079)	(0.124)	(0.066)	(0.119)
Observations	57	57	57	57	57	57
Control mean	0.069	0.552	0.034	0.448	0.034	0.448

Notes: Significance levels $p < 0.10^{*}$, $p < 0.05^{**}$, $p < 0.01^{***}$. "Treatment" is a dummy for being assigned to the treatment group. Not all respondents assigned to treatment attended the sessions, thus the coefficient on "Treatment" is the intent-to-treat effect. Dependent variables are binary indicators for the use of female condoms and male condoms. Columns 1 and 2 refer to whether the respondent reports using that type of condom at least once during the full endline period in the diaries, columns 3 and 4 to whether she reports using it at least once in the last 30 days of the diaries, and columns 5 and 6 to whether she reports using it at least once in the last two weeks of the diaries. All regressions are linear probability model ANCOVA specifications, including the baseline value of the dependent variable as a regressor. Facilitator fixed effects are not included because the sample is small with 57 endline observations, implying that some facilitators perfectly predict the endline outcome variable. Standard errors are bootstrapped with 10,000 replications, clustered at the respondent level.

4

CAN SELF-FUNDED MICROFINANCE GROUPS MATCH SAVERS WITH BORROWERS? EVIDENCE FROM MALAWI

with Marcel Fafchamps

4.1 INTRODUCTION

Group-based informal financial institutions are common in developing countries, offering access to credit, insurance, and saving in contexts where formal institutions are lacking. Rotating savings and credit associations (ROSCAs) in particular are a widespread means of generating financial intermediation, creating both a source of credit (Besley et al., 1993) and a commitment savings technology (Anderson and Baland, 2002; Ambec and Treich, 2007; Basu, 2011). Joint-liability microfinance has also harnessed the group structure in order to expand formal lending on a large scale (Besley and Coate, 1995; Ghatak and Guinnane, 1999). Furthermore, some clients may also use microfinance loans as a commitment savings device (Bauer et al., 2012).

More recently, researchers and practitioners have focused attention on self-funded microfinance groups — often known as Village Savings and Loan Associations (VSLAs), or Self-Help Groups (SHGs).[1] Such groups encourage members to save and to distribute loans among their members out of the group's accumulated savings. They therefore act as a source of microfinance, but do not rely on external lines of credit. These groups have been heavily promoted

1 Other names for such groups include VSLs, SBGs, SILCs, and SfCs depending on the NGO responsible.

by NGOs, particularly in Sub-Saharan Africa and India, and have over 100 million members worldwide (Greaney et al., 2016).[2] Impact evaluations have shown positive effects of access to VSLAs on household food security (Beaman et al., 2014; Ksoll et al., 2016).[3] Researchers have also considered how the design of such groups might screen credit risk (Greaney et al., 2016), and how there may be a trade-off between encouraging poorer individuals to join and reducing groups' capacity to lend (Burlando et al., 2017). However, very little is known about how members actually sort across groups in equilibrium, and consequently the extent to which such groups may generate financial intermediation between savers and borrowers, increasing welfare.

This paper therefore asks whether VSLAs bring together individuals who seek to save with individuals who may benefit from access to credit. In particular, we first examine whether those engaged in agriculture sort into groups with those engaged in non-farm activities, i.e. negative assortative matching on occupation. This could allow farmers to smooth income from the harvest over the year, whilst enabling those engaged in small business to take short-term loans for investment. Second, we also highlight how VSLAs can be seen as a multi-faceted commitment savings technology. This may attract present-biased individuals who are sophisticated enough to have a demand for commitment. We therefore ask whether such individuals join VSLAs and sort into groups with time-consistent individuals, who may provide commitment in exchange for access to low-interest borrowing. This would again lead to negative assortative matching, this time on measures of present-bias.[4] Both of these scenarios would imply that the groups help to reallocate capital away from those with a demand for

2 Such groups typically do not form without NGO involvement for several reasons. First, the NGO can help overcome any inherent coordination problems by bringing people together and highlighting the benefits of financial services and intermediation. Secondly, the system of governance and accounting used to protect members against theft, default or improper conduct by other members is quite sophisticated, and so NGO-provided training is key. Thirdly, the NGO typically provides access to resources such account books and safe boxes, although the group is usually asked to purchase these out of its own funds. Section 4.2.1 describes the design of such groups in more detail.

3 Ksoll et al. (2016) suggest that this may be linked to increased agricultural investments. Similar to evaluations of formal microfinance, neither of these studies finds significant effects of VSLAs on business profits, health, education or female empowerment. However, both sets of authors note that this may be an artefact of short evaluation time-frames, as most of those who joined Self-Help Groups had completed at most one savings cycle by the time of the endline surveys.

4 Strictly speaking, this would imply negative assortative matching on measures of *sophisticated* present-bias. However, like many field studies, we have a proxy of present-bias but not one of sophistication. See Section 4.3.4 for details.

saving and towards those with a demand for credit, enhancing efficiency and the welfare of both savers and borrowers.

Yet despite the potential benefits, it may be that such sorting does not materialise. For example, there may be informational or enforcement frictions associated with interacting with members of another occupation. Alternatively, individuals' tastes or social norms may encourage grouping along the lines of characteristics such as age and wealth, rather than savings and borrowing needs. The extent to which sorting is actually achieved is therefore a key question in assessing the efficacy of such groups. It can also help inform their future design, for example whom NGOs should target when recruiting members, and whether and how they might chose to assign members across groups.

To test whether VSLAs appear to achieve financial intermediation, we provide an empirical analysis of sorting using data from 150 VSLAs in Malawi. In the Summer of 2013, we revisited all VSLA groups that were originally trained as part of an intervention in 2009-11.[5] We enumerated a census of all members, past and present, and elicited basic demographic and membership information for each member. Given that we collected this information two to four years after the groups were initially trained, our data is uniquely suited to studying the long-run equilibrium sorting of members across groups, and the long-term functioning of the groups more generally. In particular, individuals have had time to learn about the savings and borrowing technologies provided by VSLAs and the benefits of grouping with different members. Accordingly, they have had chance to join VSLA groups, switch across groups at the end of savings cycles, and indeed drop out of groups altogether.

We employ a dyadic regression framework to test the predictions about sorting on occupation and present-bias. Specifically, we construct all pairs of members from the census. We then estimate dyadic logit equations to examine the determinants of any two members being members of the *same* group, given that by virtue of being in our dataset, both are both members of *some* group. If the fact that two members are different on a given characteristic is positively correlated with them being members of the same group, then this suggests that

5 This is the intervention evaluated in Ksoll et al. (2016); see Section 4.3.1 for details.

there is negative assortative matching on that characteristic, and vice versa for positive assortative matching. In light of the predictions, we test for evidence of negative assortative matching on occupation. We are also able to match a subset of members to a panel dataset containing measures of time preferences. For these individuals, we also test for evidence of negative assortative matching on present-bias.

Our results show a mixed picture in terms of how much financial intermediation is being realised through VSLAs. In stark contrast to the prediction of negative assortative matching on occupation, we in fact observe positive assortative matching. Within a village, if two members have the same primary occupation, this increases the probability that they are members of the same group by over half of the average probability of being in the same group. This lack of negative assorting on occupation suggests that there are unrealised gains from financial intermediation between individuals from different occupations. This may be because individuals find that the informational and enforcement advantages of grouping with people from the same occupation outweigh the financial intermediation benefits of matching with individuals who have a demand for borrowing and saving at different times.

On the other hand, in line with the prediction for efficient matching on time preferences, there is strong evidence of negative assortative matching on present-bias. Again, the magnitude of this effect is large: if one individual is present-biased and the other is not present-biased, then this again increases the probability that they are members of the same group by over half of the average probability. This negative assortative matching suggests that the VSLA groups do at least create a degree of financial intermediation between commitment savers and borrowers. This may be welfare-enhancing — both for commitment savers, given the low access to commitment savings technologies in these communities, and for borrowers, given the low access to credit. Nevertheless, NGOs and policymakers might be concerned if present-biased individuals' willingness to pay for commitment leads them to accept a very low interest rate when lending out their savings to borrowers. We offer evidence that this does not appear to be the case, as interest rates on lending are in line with the "fair" benchmark of members' average long-run monthly discount rates.

4.1.1 *Contribution to existing literature*

The two papers most closely related to ours are Greaney et al. (2016) and Burlando et al. (2017), both of whom consider the design of saving and borrowing groups and the composition of their members. Greaney et al. (2016) run a field experiment in Kenya, Tanzania and Uganda to compare the performance of groups who pay for their own training to groups whose training is NGO-funded. They find that groups who pay for their own training do better across a number of group-level indicators, such as loan repayment rates and average member business profits. They model borrowing motives for joining Self-Help Groups, and argue that making groups pay for their own training reduces adverse selection by making high-risk individuals select out of membership. On the other hand, this may also drive out the most needy.

Burlando et al. (2017) randomise the proportion of ultra-poor members assigned across groups in Uganda. They find that groups who are assigned a high proportion of ultra-poor members exhibit significantly lower total savings and borrowing midway through the cycle, although these differences disappear by the end of the cycle. This is suggestive of a trade-off between including ultra-poor members and increasing credit constraints at the group level, because ultra-poor members contribute fewer savings to the group's fund. Moreover, the ultra-poor members are the ones who appear to suffer from the increase in credit constraints when a group is assigned a higher proportion of ultra-poor members.

Our analysis complements these papers in two ways. First, we consider the related but separate question of how, conditional on participating, members sort across groups in equilibrium. As outlined above, sorting may be crucial in determining how well these groups are able meet the financial needs of their members. Second, we also highlight the nature of these groups as offering a commitment savings technology. As we argue, the fact that many individuals likely join such groups out of a demand for commitment may have important implications in terms of interest rates, lending dynamics and ultimately welfare.

Our motivation is conceptually similar to a body of work which examines sorting in informal financial institutions more broadly: Ghatak (2000) and Ahlin (2016) develop theoretical models of how sorting might enable the efficient pricing of risk in the context of joint-liability microfinance; Banerjee et al. (1994) study a similar problem in the design of credit cooperatives; and Eeckhout and Munshi (2010) examine empirical patterns of sorting into chit funds in India.[6] Our contribution differs insofar as we examine sorting in groups when some members are primarily interested in saving, although others remain primarily interested in borrowing.[7]

Our empirical strategy builds on work by Arcand and Fafchamps (2012) who use a dyadic regression framework to study sorting and inclusiveness in community-based organisations. Other authors have used dyadic analysis to examine sorting on risk preferences for risk-sharing games (Attanasio et al., 2012; Barr et al., 2012). Our approach adds to this literature in that we study sorting on time preferences. We also do so in the context of a fully-fledged programme setting, rather than a framed field experiment.

The remainder of the paper is organised as follows. Section 4.2 explains the design of VSLAs in more detail, and makes predictions about the sorting of members across groups. Section 4.3 describes the survey and key aspects of the data. Section 4.4 details the empirical strategy and results. Section 4.5 provides robustness checks. Section 4.6 concludes.

6 "Chit fund" is the term used in India to denote a bidding ROSCA.

7 As mentioned earlier, ROSCAs and chit funds also potentially offer a commitment savings technology. However, Eeckhout and Munshi (2010) model all members as joining with the sole motivation of getting an early turn in the rotation and thus obtaining a *de facto* loan.

4.2 CONCEPTUAL FRAMEWORK

4.2.1 *Design of VSLAs*

The design of self-funded microfinance groups and the procedures used to train members are highly similar across NGOs and countries. Our data come from VSLAs in Malawi.[8] In the intervention we study, the NGO first holds an initial information meeting in each community targeted. After this, individuals are invited to self-select into groups of 15-25 members in order to participate. With the help of the NGO, each group then purchases a cash box with three separate locks and elects three different members to act as key-holders, thereby reducing the probability that any funds placed into the box will be subject to theft. The NGO or an NGO-trained field agent then assists the group in writing a constitution, and trains the group in financial literacy and account-keeping over a period of several weeks. The group then begins to hold weekly meetings, at which each group member must make savings by purchasing between one and five "shares" of a fixed, small value.[9] After a month, members can also start requesting to take loans, to be repaid at a fixed monthly rate of interest. At the end of each cycle — usually a year — the group's total remaining savings funds plus the successfully-recovered loans and the loan interest are "shared out" in proportion to individual members' savings (hence the term "shares").

Comparing their structure to that of other savings and credit institutions, VSLAs lie somewhere between credit cooperatives and ROSCAs. VSLAs bear a strong resemblance to credit cooperatives and credit unions, but are generally much smaller and much less formalised. For example, VSLAs typically have no legal status, unlike some of the larger rural credit unions. As a consequence, VSLAs are likely to be more reliant on interpersonal relationships

8 VSLAs appear to be the first version of "Self-Help Groups" to be introduced to Sub-Saharan Africa, by CARE International in Niger in 1991. See http:\www.vsla.net.

9 In principle, each member is also required to make a small weekly contribution to the group's insurance fund, to cover events such as illness or death of a group member's relatives. However, in practice we found that most groups set the level of such contributions to be very small: around 20 Malawi Kwacha (MK), equivalent to $0.06). Some groups had even dropped this component altogether, citing past disagreements about payouts.

for monitoring and enforcement of loan repayment, which may have both advantages and disadvantages. Insofar as VSLAs are informal savings and credit groups which rely on "social capital" to sustain them, they are close in spirit to ROSCAs. However, VSLAs are more sophisticated than ROSCAs in that they allow for a much greater degree of flexibility. On the savings side, each member can choose to buy between one and five shares each week, rather than committing to a fixed payment which is the same for all members and all weeks. On the lending side, members can demand the size and timing of the loans they receive, rather than having to wait for their turn in a rotation.[10] Meanwhile, compared to joint-liability microfinance, VSLAs are distinct insofar as they offer a pure savings technology as well as credit.[11]

VSLAs can be seen as a multi-faceted commitment savings technology. First, there is a deposit commitment, since all members are required to purchase at least one share per week.[12] Second, VSLAs enforce a withdrawal commitment, since savings made into the box (i.e. all shares purchased) cannot be liquidated until the end of the savings cycle.[13] Third, the group meetings likely provide other "soft" deposit commitment technologies: active peer pressure to save; a desire to save in order to appear reliable to one's peers (Breza and Chandrasekhar, 2015); and reminders to save (Kast et al., 2012). Finally, any member requesting a loan must prove that it is for a good purpose, and the whole group must agree for the loan to be granted. Thus it is unlikely that members can undo the commitment savings features of the group by borrowing for consumption, except in cases of demonstrable emergencies. VSLAs may therefore offer an attractive package to individuals who have problems with self-control or with shielding resources from their spouse or relatives, and who are sophisticated enough to recognise the potential value of commitment.[14]

10 Bidding ROSCAs do allow members some choice over when they receive the pot; however, each member can still only receive the pot once and cannot choose the pot size.

11 VSLA interventions are also typically targeted at communities which are under-served even by microfinance lenders.

12 In practice we see that some groups occasionally relax this requirement, allowing individuals to purchase zero shares in certain weeks. It is likely that the groups somehow strike a balance between commitment and liquidity to cover shocks, by relaxing the deposit commitment but keeping the withdrawal commitment strong.

13 There is a provision that individuals can withdraw a small number of their own shares in the case of a medical emergency or similar. However, we rarely see instances of this when examining the individual account books.

14 Anecdotally, members do seem sophisticated in this way: when we asked about reasons for joining, many individuals told us that being in a VSLA "overcomes the temptation of spending savings kept at home", or "addresses the problems encountered within the household".

4.2.2 *Sorting across VSLAs*

4.2.2.1 *Sorting on occupation*

From a financial intermediation perspective, the most salient characteristic on which we would expect to see sorting is occupation. The population of the survey area is overwhelmingly engaged in one of two occupations: agriculture, and small business.[15] These occupations have very different cash-flow needs. Farming households have a large outflow of funds for investment during the planting season in January, and receive one major inflow of funds just after the maize and rice harvest in April.[16] Meanwhile households engaged in business and other non-farm activities typically have frequent inflows and outflows of income, and may require access to capital throughout the year. In terms of sorting, broadly three scenarios are therefore possible:

1. EFFICIENT SORTING: Given the cash-flow patterns described above, efficiency would imply negative assortative matching of VSLA members on occupation — i.e. those in agriculture sorting into groups with those engaged in small business. If this were to occur, the farmers would expect to act as savers for most of the cycle, except for occasional out-of-season agricultural investments and emergencies. Meanwhile those in non-farm activities would borrow out of the farmers' savings, thereby generating dividends for the farmers through loan interest repayments. Such sorting might be achieved through direct bargaining, or through the loan interest rate: competition for borrowing funds amongst non-farmers within a given VSLA would drive up the interest rate on lending, eventually pushing some non-farmers to move to other VSLA groups with relatively few non-farmers.

15 Typical businesses in the area include selling vegetables and goods from nearby markets, bricklaying and carpentry, driving bicycle taxis, and sewing.

16 There is also a second, smaller harvest for cassava which takes place in November. This is mainly used to supplement household consumption during the lean season of December-February.

2. SORTING WITH SOCIAL CONSTRAINTS: On the other hand, the fact that most VSLA lending is done on a limited-liability basis may prevent this efficient sorting from occurring.[17] If members' ability to monitor and punish delinquent borrowers is much stronger when the borrower has the same occupation, then this may overwhelm the benefits of negative assorting and instead lead to positive assortative matching on occupation. A similar argument would apply if transaction costs are much lower among those with the same occupation, or if individuals derive much greater social benefit from interacting in a VSLA with others in the same occupation.

3. SORTING WITH NO CREDIT: Finally, it may be that VSLAs are unable to engage in providing much access to investment credit altogether. This may occur if even the most tightly-knit groups are unable to enforce high rates of loan repayment, or if savers are so risk-averse that they are unwilling to lend out their savings. In this case, members would join VSLAs almost exclusively in order to use them as a savings technology. We would therefore expect to see zero assortative matching on occupation, or positive assortative matching if there are social benefits as described above.

4.2.2.2 *Sorting on time preferences*

As detailed above, VSLAs also offer a commitment savings technology. This may attract sophisticated present-biased individuals to become members in order to save.[18] If so, again broadly three scenarios are possible.

1. EFFICIENT SORTING: The first scenario is if these present-biased individuals sort into groups with borrowers. This would lead us to observe negative assortative matching

17 The member's own shares can be seized as collateral in the case of non-repayment, but these may not be enough to cover the value of the loan.

18 We focus on self-control problems — i.e. present-bias — as a driver of demand for commitment savings. However, in our empirical analysis we also perform analogous tests of sorting along measures of "other-control" problems, namely household bargaining power and income hiding.

on present-bias, since borrowers will be mainly be time-consistent. The latter holds because time-consistent members are less likely to have a demand for commitment, and so are more likely to have joined for borrowing motives.[19]

This scenario is efficient from a financial intermediation perspective, since it allows commitment savers' capital to act as a source of credit to borrowers. Again, the sorting may occur directly, if members of a village have a good idea about one another's preferences. Alternatively, it may occur via the interest rate: commitment savers by definition have a weakly positive willingness to pay for commitment, and thus are willing to accept a weakly lower interest rate (compared to regular savers or prospective borrowers) when lending out their savings deposits. Thus prospective borrowers who seek out the lowest loan interest rate should end up sorting into groups with commitment savers. Furthermore, there may be credit rationing even at the equilibrium loan interest rate (Burlando et al., 2017). Thus competition for funds may further drive prospective borrowers into groups with commitment savers, rather than other borrowers.[20]

2. SORTING WITH SOCIAL CONSTRAINTS: On the other hand, it may be that social norms dictate sorting into groups on the basis of characteristics such as social status or family ties, and that these norms are so strong as to overwhelm any sorting along savings and borrowing characteristics. In this case we would observe zero assortative matching on present-bias, conditional on these types of characteristics.

3. SORTING WITH NO BORROWING: A third possible scenario is again if the credit function of VSLAs becomes negligible, due to lack of repayment enforceability. If so, both present-biased and and time-consistent individuals may still join VSLAs, but any time-consistent

19 Time-consistent individuals may also join for "other-control" motives, or indeed as a way to undertake precautionary savings. It is also theoretically possible that they might engage in speculative behaviour, if they seek to "pile in" towards the end of the savings cycle in an attempt to suck out any profits from lending up until that point. However, the scope for such behaviour is limited by the rule of a maximum share purchase of five shares per week.

20 Even if some present-biased individuals also seek to borrow, groups may learn that these present-biased individuals are less reliable for repayment.

individuals who do so must also be seeking a saving technology.[21] If VSLAs act as pure savings groups, then it may be that everybody would prefer time-consistent saving partners over present-biased saving partners, given that time-consistent individuals are likely to save more reliably and thus to provide a good influence. If so, then time-consistent savers would group together, leaving present-biased savers to group together, and we would observe positive assortative matching on present-bias.[22] Alternatively, it may be that commitment savers prefer to partner with other commitment savers — analogous to anti-addiction or weight-loss groups. This would again lead to positive assortative matching on present-bias.

4.3 SURVEY AND DATA

4.3.1 *Description of VSLA survey*

We now turn to testing these predictions empirically. In the Summer of 2013 we surveyed 150 VSLA groups in Karonga District, northern Malawi (Figure 11). These VSLAs were originally formed as part of a cluster-randomised controlled trial, which ran from 2009 to 2011. Forty-six villages were included in the initial study, half of which were invited to form groups and began to receive VSLA training in late 2009-early 2010 (treated villages), and half of which only received training in late 2011 (control villages). The intervention was implemented by the Rockwool Foundation and CCAP Synod of Livingstonia Development Department (SOLDEV), and the results of the impact evaluation are detailed in Ksoll et al. (2016).[23] Since

21 A time-consistent individual may still save in a VSLA at zero nominal interest if saving at home carries a negative interest rate, for example due to risk of theft.

22 Side payments could theoretically complicate the sorting process, as commitment savers might be willing to make larger transfers to time-consistent savers in order to obtain their good influence as fellow group members. However, the possibility of sustaining multiple bilateral side payments amongst group members seems unlikely. More likely is that this willingness to pay for time-consistent partners would be reflected in the interest rate on lending, but this takes us back to the first scenario as outlined above.

23 See http://www.rockwoolfonden.dk and http://www.ccapsolinia.org/ for details. The training of these groups was funded by the Rockwool Foundation, rather than by members themselves. Greaney et al. (2016) show that whether the NGO or members pay for training affects who participates in VSLAs. Thus our empirical results on sorting may only be representative of groups in which the NGO pay for the training. However, this is still by far the most widely-used model for Self-Help Group interventions.

we visited the area two years after the control villages were phased into treatment, our sample covers all VSLA groups that were eventually trained by SOLDEV in both treatment and control villages, and we do not exploit the initial randomisation.[24] Two very remote control villages dropped out of the programme in 2011 and never established any groups, hence our 2013 sample covers the remaining 44 villages.

Figure 11: Location of Karonga District within Malawi



The survey protocol was as follows. We contacted each group via the NGO, who invited all group members to a meeting at the group's usual meeting place. We first explained the purpose of our survey, and obtained the consent of all group members to share their information. The data collection then proceeded in three steps. First, we used the set of individual account books to construct a roster of all group members, past and present. We then elicited basic demographic information for each member as well as their membership history, by reading out each member's name to the group and then asking a series of questions about that individual. This constitutes the census data, which is the main focus of the analysis in this paper. Second, we also enumerated a short group survey covering the group's history and practices,

24 Anecdotally, we learned that a number of "replication" groups did form without SOLDEV training, either autonomously or with the help of members of SOLDEV-trained groups who had been encouraged to teach others the VSLA methodology. Data on such informal groups was difficult to obtain. Our results are therefore valid for all members of official VSLA groups in the area.

such as the typical use of savings and loan funds, the interest rate charged on loans, and the typical punishment for late loan repayments. Finally, we photographed each individual's account book, which details their weekly savings decisions and their borrowing behaviour for the current cycle. In what follows we draw on the group survey and account books as supplementary data, for example when considering the use of funds and interest rates.

4.3.2 *Matching to household survey data*

To test whether members sort on present-bias, we require a measure of individuals' time preferences.[25] Ideally this measure would have been elicited prior to members joining, to rule out reverse causation if being members of the same group leads individuals to have a greater similarity or differences in their choices over time. To obtain such a measure, we matched the members of our 2013 census by name and village back to the 2009-11 panel dataset which was collected for the initial impact evaluation.[26] The 2009-11 panel covers a stratified random sample of households from the treatment and control villages.[27] Since the 2009-11 panel contains only a random subsample of each village's population — whereas our 2013 member census covers all members — many members in our member census were not interviewed as part of the 2009-11 panel. Overall we are able to match around a fifth of the members from our 2013 census (722 out of 3,801) to the 2009-11 panel. From now on we refer to this subsample as "matched individuals".[28] These individuals are evenly spread across groups: we matched

25 Given that we enumerated the 2013 census to the whole group simultaneously, we were restricted to asking about a reasonably small set of individual characteristics. We therefore did not attempt to measure preferences, and so cannot offer a test of the predictions on present-bias from Section 4.2.2.2 on the whole census.

26 We report time preference values from the 2010 wave, since the 2009 wave did not include the far frame for female respondents. However, only a very small number of the VSLA groups had begun to form by early 2010, and thus time preferences are still plausibly exogenous to the characteristics of other group members.

27 By construction, the 2009-11 panel therefore includes some individuals who after baseline went to become members of VSLAs and whom we match to our 2013 member census, and others who did not and who therefore do not appear in our 2013 member census. The 2009 baseline was stratified insofar as households who declared an interest in joining VSLAs were over-sampled. For us this just increases the probability that we are able to match members of our 2013 census to the 2009-11 panel.

28 The matched members should be more representative of the villages' populations than the census of all members is. For example, whilst only 25% of members in our 2013 member census are male, since women disproportionately join VSLA groups, close to 50% of the respondents in the 2009-11 survey were male due to it being representative. Thus we are disproportionately likely to match a male 2013 member back to the 2009-11 survey

at least one member in 95.3% of groups, and on average we matched 4.7 members out of an average group size of 25.3 members. Of these, 352 randomly received the full time preference module and so can be used to test the predictions of sorting on present-bias.[29]

4.3.3 *Group-level data on saving and borrowing*

From the group survey, we see that the median share price in 2013 was 100 MK, equivalent to around $0.30.[30] Members of the median group could therefore save between $0.30 and $1.50 per week, or $16-$80 per year.[31] Loan sizes vary greatly, but typical amounts for larger loans are 5,000 MK or 10,000 MK ($15 or $30).

The patterns of how savings and loan funds are used is quite distinct. The predominant use of savings is for agricultural inputs, with 58% of groups reporting that this is among the three largest uses by their members. Indeed, the groups all choose to share out in January which is during the planting season.[32] The other most prominent uses of savings are food — since the January share-out also coincides with the lean season — and durable household items, such as kitchenware. Loans on the other hand are highly concentrated on trading and business purposes: 74% of groups say this is the most important use of loans, and altogether 95% say this is among the three most important uses of loan funds. The other most commonly-reported uses of loan funds are education, emergencies, and again purchasing food. Thus at first sight there does seem to be some financial intermediation taking place within groups: some members save for the planting and lean season, whilst their savings are lent out to members who engage in trading/business activities, or who need to smooth consumption after shocks.

compared to a female 2013 member. Section 4.5 describes how all of our results are robust to re-weighting to take account of this.

29 For budgetary reasons, only a random subset of households received the full questionnaire including time and risk preference modules.

30 The exchange rate at the start of the July 2013 survey was 330 MWK = 1 USD.

31 For comparison, Malawi's GNI per capita in 2013 was $390 (http://data.worldbank.org); although the study area is a particularly poor and remote region.

32 This choice of timing itself may reflect sophistication about the effect of having cash-in-hand just before input purchase is required (Brune et al., 2011; Duflo et al., 2011).

Turning to interest rates, in 2013 the average monthly loan interest rate set by groups was 17%. Whilst this may seem high, it is close to the most natural benchmark for a "fair" interest rate available in the data: namely, individuals' average monthly discount rate in the long term. The latter is estimated to be 20% with linear utility, or lower if the curvature of the utility function is taken into account; see Section 4.3.4.2 for details.[33] At the same time, this suggests that borrowers are not taking advantage of commitment savers' willingness to pay for commitment by gaining a very low interest rate. Indeed, savers appear to earn a healthy return on their shares: the median monthly interest on saving is approximately 3%, as calculated from each group's reported annual return per share at the end of the cycle.[34]

4.3.4 *Individual member characteristics*

4.3.4.1 *Census of members*

Table 53 describes some of the key demographic characteristics of the 3,801 individuals in our member census. 73% of members report farming as their primary economic activity, whilst 21% work in business (mainly family businesses). Although the NGO imposes no rules on the gender of participants, 75% of the members are female.[35] The data also suggest a large degree of churn individuals' membership, which again points to the scope for sorting across

33 Inflation in 2010 versus 2013 should of course be taken into account in order to compare in real terms the interest rates in 2013 to the discount rates measured in 2010. Inflation in Malawi was fairly stable at around 8% y-o-y (corresponding 0.64% per month) from the beginning of 2009 until the beginning of 2012. After a devaluation of the Kwacha by 33% in May 2012, inflation spiked and ran at an average of 28% in 2013 overall (corresponding to 2.1% per month). Groups do not appear to have taken this into account in the nominal loan interest rate set at the beginning of the 2013 cycle, which in most groups remained unchanged from previous cycles. However, even the high 2013 inflation rate is still negligible on a monthly basis compared to such a high monthly loan interest rate. Thus accounting for monthly inflation does not alter our conclusion that the interest rate on loans appears to be broadly in line with monthly discount rates.

34 This is straightforward to reconcile with a monthly interest rate on borrowing of 17%, since only a fraction of the group's funds are lent out at any given time. The median annual return per share is 45%, equating to 3% monthly interest compounded on shares purchased right at the beginning of the year. However, this is a conservative approximation — given that individuals save throughout the year rather than just at the beginning, the true effective monthly interest rate will be higher.

35 This could reflect a variety of factors: for example if women have lower transaction costs or higher social benefits of membership, lower access to other financial technologies, or indeed are viewed as more financially responsible.

groups: almost a third of members (1,262 out of 3,801) join sometime after the first cycle, and 521 members have left the group at some point by 2013.

Table 53: Individual member characteristics – 2013 member census

	Full CSAE Sample				
	Mean	Std dev	Min	Max	N
Occupation					
Farmer	0.73	(0.44)	0.0	1.0	3801
Businessperson	0.21	(0.41)	0.0	1.0	3801
Other	0.06	(0.23)	0.0	1.0	3801
Demographic Variables					
Male	0.25	(0.43)	0.0	1.0	3799
Female-headed household	0.21	(0.41)	0.0	1.0	3796
Age	36.15	(12.05)	12.0	83.0	3785
Education					
Some primary educ. (only)	0.76	(0.43)	0.0	1.0	3801
Some post-primary educ.	0.18	(0.39)	0.0	1.0	3801
Literate (read & understand newspaper)	0.83	(0.38)	0.0	1.0	3795
Wealth					
Father well-off in village (scale 1-5)	3.34	(1.31)	1.0	5.0	3573
Spouse's father well-off in village (scale 1-5)	3.42	(1.32)	1.0	5.0	3566
Income Poverty Indicators					
Household well-off in group (scale 1-9)	7.52	(1.28)	1.0	9.0	3770
HH owns a bicycle	0.51	(0.50)	0.0	1.0	3796
# Goats	1.29	(2.59)	0.0	40.0	3792

Notes: All variables as measured during the 2013 member census, N=3,801 members. 519 members were no longer active, but are included to avoid selection bias. Analysis is conducted with and without these individuals, see Section 4.5. Missing observations reflect answers of "do not know" or "not applicable". Occupation denotes an individual's primary economic activity, if engaged in multiple activities.

4.3.4.2 *Time preference data*

As described above, we matched 722 individuals by name to the baseline data of the 2009-11 impact evaluation. Matching to this additional dataset yields a much richer set of data on individual characteristics for the matched individuals, as summarized in Table 55. Crucially, activities to measure time preferences were administered to both the head and the spouse in a random subset of the panel data households in 2010, and can be matched to 352 individuals.

Table 54: Individual economic activity & schooling – 2013 member census

Category	N	% of responses
Occupation	**3780**	**100%**
Farmer	2,786	73.34%
Business	793	20.87%
Self-employed	*108*	*2.84%*
Family business worker	*685*	*18.03%*
Fishing	61	1.61%
Fishing, employed	*17*	*0.45%*
Fishing, self-employed	*44*	*1.16%*
Employee	89	2.34%
Casual labour (ganyu)	24	0.63%
Student	4	0.11%
Unemployed, not seeking work	9	0.24%
Other	14	0.37%

Data from the VSLA member census conducted in July-August 2013. The total number of members is 3,801. The number of observations for each variable reflects missing values or non-applicability.

Notes: All variables as measured during the 2013 member census, N=3,801 members. 519 members were no longer active, but are included to avoid selection bias. Analysis is conducted with and without these individuals, see Section 4.5. Missing observations reflect answers of "do not know" or "not applicable". Occupation denotes an individual's primary economic activity, if engaged in multiple activities.

The time preference activities took the form of multiple price lists. Participants were first asked whether they would prefer to receive 2000 Kwacha (approximately $13) now or increasing amounts in one month.[36] This constitutes the near frame. Participants were then asked whether they would prefer to receive 2000 MK in one year or the same increasing amounts in one year and one month. This constitutes the far frame.[37] The average respondent prefers 2000 MK now to 2332 MK in one month, and 2000 MK in one year to 2402 MK in one year and one month. If participants answered such questions without considering their background consumption, and if utility was linear, this would imply an average near-frame monthly discount rate of 17% and an average far-frame monthly discount rate of 20%. However, taking into account any curvature of the utility function implies a lower discount rate (Andersen et al., 2008).

A more impatient choice in the near frame than in the far frame is taken to indicate "present-bias". The choices of members classified as "present-biased" imply an average near-frame discount rate of 16% and an average far-frame monthly discount rate of close to zero, consistent with the idea of present-biased individuals exhibiting excessive short-run discounting but more modest long-run discounting.[38] Overall 11% of individuals are classified as "present-biased". This is very similar to the rate of 10% found by Brune et al. (2011) also in rural Malawi. Other estimates from developing countries find that a larger proportion of individuals appear to be "present-biased" (Ashraf et al., 2006; Giné et al., 2016; Janssens et al., 2017). However, as discussed in chapter one, estimates of "present-bias" over money may be exaggerated at times of tight liquidity constraints (Carvalho et al., 2016). Tight liquidity constraints are much less of a concern here, since the 2009-11 panel dataset was always enumerated shortly after the harvest. If anything, we may under-estimate the number of present-biased individuals, if some present-biased individuals have enough liquidity to arbitrage experimen-

36 The exchange rate at the start of 2010 was approximately 150 MWK = 1 USD. Due to practical constraints, responses were unincentivized. The limited evidence comparing incentivized and unincentivized responses to time preference questions suggests that lack of incentivization does not bias responses, although they may be more noisy (John, 2017). This is in contrast to evidence that unincentivized elicitation may cause biases in the cases of risk preferences and beliefs.

37 The frame order was not randomised, as this was not practical in a large paper-based household survey.

38 Again, the estimate of 16% is an approximation if utility is linear, but is an upper bound if utility is concave. Those classified as "time-consistent" actually exhibit far-frame switch-points consistent with a higher far-frame discount rate than those classified as "present-biased". However, these "time-consistent" individuals appear equally impatient in the near and the far frames.

tal payments (Augenblick et al., 2015). This would reduce our chances of observing the any assortative matching, positive or negative, on this measure. Similarly, our measure does not proxy whether individuals categorised as "present-biased" are sophisticated or not. If only a subset of those individuals classified as "present-biased" are sophisticated and have a demand for commitment, then again this would reduce our chances of observing any assortative matching on this measure.

The fact that the far frame refers to one year after the near frame eliminates concerns that seasonality in consumption and liquidity constraints may act as a confound in the measure of "present-bias" (Epper, 2015). Nonetheless, individuals may still spuriously appear "present-biased" if they are expecting a decrease in the marginal rate of intertemporal substitution next year compared to this year, for example if this year's harvest was particularly bad for their household. We therefore employ a number of tests to check whether measured "present-bias" appears to be capturing a decreasing marginal rate of intertemporal substitution, rather than truly present-biased preferences.

First, if individuals who appear "present-biased" are actually those facing a higher marginal rate of intertemporal substitution than they expect to face in one year's time, we might expect the measure of "present-bias" to be correlated with measures of low recent consumption. Vice versa, we might expect individuals classified as "future-biased" (26% of the sample) to exhibit high recent consumption. In fact, the measures of "present-bias" and "future-bias" in early 2010 are completely uncorrelated with consumption from 2009. Moreover, whilst "future-bias" is correlated with better food security in 2009 (the household is 8.9 percentage points less likely to have had fewer than two meals per day on average in the last week, p-value 0.025), "present-bias" is also marginally correlated with better food security (14.0 percentage points, p-value 0.112). This suggests that individuals who appear "time-consistent", not those who appear "present-biased", are those who have experienced recent hardship.

Data on saving and borrowing also points more strongly against the idea that individuals who appear "present-biased" are actually those who are expecting to be better off next year. If this was the case, we might expect these individuals to exhibit higher recent borrowing and

lower saving. Instead, we observe a strong negative correlation between appearing "present-biased" in 2010 and data on borrowing from 2009 and 2010: in 2009, individuals categorised as "present-biased" are 13.7 percentage points less likely to have asked for a loan in the past year, 7.3 percentage points less likely to have any current loans, and have 746 MK fewer in outstanding loans; whilst in 2010 they are 16.8 percentage points less likely to have asked for a loan in the past year, and have 4957 MK fewer in outstanding loans, although the latter is marginally insignificant (p-value 0.134). All of this is more consistent with the idea that the idea that present-biased individuals may not be deemed credit-worthy by others, or may avoid seeking credit if they are aware of their own tendency to over-consume. "Present-biased" individuals also have higher total savings from 2009 (4468 MK, p-value 0.075), which further goes against the idea that individuals appear "present-biased" because they are liquidity-constrained now but anticipate higher income in the future.[39] Meanwhile, the measure of "future-bias" is uncorrelated with measures of saving and borrowing.

On the other hand, the measure of "present-bias" is marginally correlated with subjects' subjective report that they have had a bad harvest in 2010 compared to the past decade (correlation of 0.292 on a 1-5 Likert scale, p-value 0.101). We thus cannot entirely rule out that the measure of "present-bias" may in part be capturing individuals' expectations that they will be better off in a year's time. Therefore, in Section 4.4 we run a further set of tests to check that true present-bias rather than a decreasing marginal rate of intertemporal substitution is driving our results on sorting. The 2009-11 panel dataset also provides measures of the matched individuals' risk aversion — elicited using standard Binswanger lotteries — intra-household bargaining power, and more detailed measures of consumption and food security. These are also summarised in Table 55, and are used as additional controls in robustness checks.

39 The fact that "present-biased" individuals may be able to save outside of VSLAs is still consistent with them having a demand for VSLAs as a commitment savings device, since VSLAs may offer a better return than other forms of saving such as cash-under-the-mattress. Moreover, exercising self-control by oneself may be costly (Gul and Pesendorfer, 2001; Toussaert, 2015).

Table 55: Individual member characteristics – matched subsample

	Matched individuals				
	Mean	Std dev	Min	Max	N
Time Preferences					
Present-biased	0.11	(0.31)	0.0	1.0	350
Future-biased	0.26	(0.44)	0.0	1.0	350
Minimum switch-point, near frame	2332.36	(369.62)	1900	2800	352
Minimum switch-point, far frame	2401.39	(378.38)	1900	2800	352
Risk Preferences					
Risk-neutral	0.11	(0.32)	0.0	1.0	330
Intra-Household Bargaining					
Ever hides money from spouse	0.44	(0.50)	0.0	1.0	307
Female HH decision-making power (index 0-8)	2.87	(1.81)	0.0	8.0	377
Social Variables					
HH important in village decisions (scale 1-6)	3.27	(1.14)	1.0	6.0	721
HH ever speaks at village meetings	0.57	(0.50)	0.0	1.0	718
Income					
Monthly consumption per capita, MK	2176.95	(973.09)	648.7	7811.1	721
Food security poor (dummy)	0.28	(0.45)	0.0	1.0	722
Credit					
HH asked for credit in last year	0.15	(0.36)	0.0	1.0	383
HH has any loans outstanding	0.09	(0.29)	0.0	1.0	383
Total value of loans outstanding, MK	779.90	(4238.64)	0.0	45000	383

Notes: N=722 individuals are matched from the 2013 census to the 2009-11 panel data. N=383 of these individuals are matched to the longer panel survey including preference modules. Missing values reflect "do not know", "not applicable", or inconsistent answers in the case of risk preferences. All variables presented here were measured in the 2009 wave of the panel survey, except time preferences, which are taken from the 2010 survey wave since the 2009 wave did not include the far frame for females. Present-biased (future-biased) is a dummy equal to one if the response to the near frame is more impatient (patient) than the response to the far frame. Minimum switch-point is the lower bound of the interval in which the respondent switched to preferring the one-month-later payment compared to a 2000 MK payment on the earlier date. 150 MK $\approx$ 1 USD at the time of the 2009 and 2010 surveys. Risk-neutral is a dummy equal to one if the respondent prefers a 50-50 lottery to its expected value for certain, and thus could indicate risk-neutral or risk-seeking behaviour. Female HH decision-making power is constructed from questions over four types of economic decisions, scoring one if the female has some say in the decision and two if she has complete control. Malawi's GNI per capita in 2009 was $26.6/month, but these are particularly poor households in a very remote region. Food security poor is equal to one if the household reports consuming fewer than three meals yesterday.

4.3.5 *Group composition by member characteristics*

Table 56 describes the distribution of groups across villages. Thirty-five villages have at least two groups, and some villages have up to fourteen groups.[40] The presence of more than one group per village is itself suggestive of inefficiency: there is no secondary market for capital in these villages, and so VSLAs with excess capital cannot lend to other VSLAs; hence having one large VSLA per village would maximise the scope for lending out savings deposits and alleviating credit constraints. However, transaction costs and ability to monitor and sanction borrowers likely become too large past a certain group size, explaining why we observe multiple groups per village.[41] Hence sorting across groups becomes a key determinant of efficiency in this second-best scenario.

Table 56: VSLA groups per village

# groups	# villages	% of villages
1	9	20.5%
2	17	38.6%
3	4	9.1%
4	4	9.1%
5	3	6.8%
6	2	4.6%
7	2	4.6%
11	1	2.3%
13	1	2.3%
14	1	2.3%
Total	**44**	**100%**

Notes: From the 2013 member census, N=150 groups, N=44 villages. Eight of these groups were no longer active, but were included in the survey to avoid selection bias. Analysis is conducted with and without these groups, see Section 4.5.

Table 57 describes how groups are composed in terms of member characteristics. The average group size is 25 members, although groups range in size from 10 to 45 members. Groups also range in gender composition from all-male to all-female, although most groups (i.e. groups

40 Nine villages have just one group, and thus sorting is not identified in these villages.

41 Indeed, anecdotally we were told that one VSLA of around 45 members had to split into two groups, because the members found it too difficult to reach consensus decisions in such a large group.

within one standard deviation of the mean) are mixed but with a majority of female members. In some groups, as many as 62% of members come from female-headed households. There is also clear heterogeneity across groups in terms of occupational composition: some groups consist purely of farmers, whereas others contain almost no farmers. However, dyadic regression analysis is needed to determine whether such heterogeneity is evidence of individuals sorting across groups within villages, or whether it represents differences in population characteristics across villages.

Table 57: Group composition by member characteristics

Variable	**Average**	**Std. Dev**	**Min**	**Max**
# members	25.34	5.69	10	45
% members farmers	74.2	26.3	3.3	100.0
% members businesspeople	20.2	22.2	0.0	93.3
% members fisherman/woman	1.6	5.5	0.0	30.4
% members female	74.9	19.0	0.0	100.0
% members female-headed HH's	20.9	13.0	0.0	61.9
Mean age of members	36	4.6	23	49
% members literate	82.5	12.2	40.9	100.0
% members some primary only	81.6	11.4	38.7	100.0
% members own bicycle	52.1	18.6	0.0	95.0
Mean # goats owned by members	1.32	0.98	0.04	7.48

Notes: From the 2013 member census, N=150 groups, N=44 villages. Eight of these groups were no longer active, but were included in the survey to avoid selection bias. Analysis is conducted with and without these groups. Similarly, 519 of these members were no longer active, but were included in the survey for completeness. Analysis is conducted with and without these members, see Section 4.5.

4.3.6 *Dyad characteristics*

For the analysis, we construct all possible pairs – dyads – of members in the same village from the 2013 member census.[42] Of these dyads, 17% comprise two individuals who are both members of the same group, whereas the other 83% comprise two individuals who are members of different groups in the same village. Table 58 describes the dyads in more detail. When we restrict attention to the dyads in which both individuals are matched to the full version of the 2009-11 panel dataset, including the time preference modules, this gives us a sample size of 1,641 dyads.[43] Table 59 highlights key additional data for the matched dyads.

4.4 EMPIRICAL STRATEGY AND RESULTS

4.4.1 *Dyadic regression framework*

To test the predictions on sorting from Sections 4.2.2.1 and 4.2.2.2, we employ a dyadic regression framework (Fafchamps and Gubert, 2007). The intuition behind this approach is as follows: if there are multiple groups in a village, and if there is positive sorting on a given characteristic, then in equilibrium two members who are less similar on that characteristic are, *ceteris paribus*, less likely to be observed as members of the same group. Vice versa, if there is negative sorting on a given characteristic, then two members who are less similar on that characteristic are more likely to be members of the same group.

42 In practice we found it to be extremely rare that an individual would join a group outside of his or her village of residence. Thus *de facto* only the other members from an individual's village of residence are candidates to be members of the same group as that individual.

43 In terms of basic characteristics, Table 64 shows that the dyads which can be matched to the full version of the panel survey have small but significant differences from the whole universe of dyads from the 2013 member census. Therefore, we later re-run all of our time preference specifications weighting each dyad by the inverse probability of that dyad being matched to the full 2009-11 survey. This does not change our results, see Section 4.5.

Table 58: Dyad characteristics – 2013 member census

	Full CSAE Sample				
	Mean	Std dev	Min	Max	N
Membership					
Same VSLA group	0.17	(0.37)	0.0	1.0	289914
Occupation					
Same economic activity	0.56	(0.50)	0.0	1.0	289914
Absolute differences - Demographic Variables					
Male	0.38	(0.48)	0.0	1.0	289467
Female-headed household	0.32	(0.47)	0.0	1.0	288740
Age	12.64	(10.25)	0.0	65	286763
Absolute differences - Education					
Some post-primary educ.	0.33	(0.47)	0.0	1.0	289914
Literate (read & understand newspaper)	0.26	(0.44)	0.0	1.0	288342
Absolute differences - Wealth					
Father well-off in village (scale 1-5)	1.41	(1.10)	0.0	4.0	253041
Spouse's father well-off in village (scale 1-5)	1.43	(1.13)	0.0	4.0	252485
Absolute differences - Income and Poverty					
Household well-off in group (scale 1-9)	1.40	(1.15)	0.0	8.0	284485
HH owns a bicycle	0.46	(0.50)	0.0	1.0	288740
# Goats	1.84	(2.73)	0.0	40	288240
Sum - Occupation					
Farmer	1.37	(0.70)	0.0	2.0	289914
Businessperson	0.48	(0.63)	0.0	2.0	289914
Sum - Demographic Variables					
Male	0.53	(0.63)	0.0	2.0	289467
Female-headed household	0.40	(0.57)	0.0	2.0	288740
Age	72.12	(16.69)	24.0	163	286763
Sum - Education					
Some post-primary educ.	0.43	(0.58)	0.0	2.0	289914
Literate (read & understand newspaper)	1.68	(0.52)	0.0	2.0	288342
Sum - Wealth					
Father well-off in village (scale 1-5)	6.76	(1.85)	2.0	10.0	253041
Spouse's father well-off in village (scale 1-5)	6.85	(1.88)	2.0	10.0	252485
Sum - Income and Poverty					
Household well-off in group (scale 1-9)	15.04	(1.83)	3.0	18.0	284485
HH owns a bicycle	0.99	(0.73)	0.0	2.0	288740
# Goats	2.44	(3.38)	0.0	70	288240

Notes: All variables as measured during the 2013 member census, N=3,801 members. 519 of these members were no longer active, but are included to avoid selection bias. Analysis is conducted with and without these individuals, see Section 4.5. All possible dyads in which both individuals live in the same village are constructed, N=289,914. Missing observations reflect answers of "do not know" or "not applicable". Occupation denotes an individual's primary economic activity, if engaged in multiple activities.

Table 59: Dyad characteristics – matched subsample

	Matched dyads				
	Mean	Std dev	Min	Max	N
Absolute differences					
Present-biased	0.16	(0.37)	0.0	1.0	1641
Future-biased	0.37	(0.48)	0.0	1.0	1641
Minimum switch-point, near frame	366.05	(321.43)	0.0	900	1655
Minimum switch-point, far frame	351.92	(337.30)	0.0	900	1651
Risk-neutral	0.20	(0.40)	0.0	1.0	1513
Ever hides money from spouse	0.45	(0.50)	0.0	1.0	1269
Female HH decision-making power (index 0-8)	1.95	(1.57)	0.0	8.0	1914
HH important in village decisions (scale 1-6)	1.08	(1.11)	0.0	5.0	7314
HH speaks at village meetings	0.47	(0.50)	0.0	1.0	7266
Sums					
Present-biased	0.19	(0.42)	0.0	2.0	1641
Future-biased	0.53	(0.64)	0.0	2.0	1641
Minimum switch-point, near frame	4761.51	(558.00)	3800	5600	1655
Minimum switch-point, far frame	4908.17	(563.28)	3800	5600	1651
Risk-neutral	0.22	(0.44)	0.0	2.0	1513
Ever hides money from spouse	0.79	(0.71)	0.0	2.0	1269
Female HH decision-making power (index 0-8)	5.90	(2.71)	0.0	16.0	1914
HH important in village decisions (scale 1-6)	6.54	(1.69)	2.0	12.0	7314
HH ever speaks at village meetings	1.16	(0.71)	0.0	2.0	7266

Notes: N=722 individuals are matched from the 2013 census to the 2009-11 panel data. N=383 of these individuals are matched to the longer panel survey including preference modules. All possible dyads in which both individuals live in the same village are constructed, N=7,314 for the general survey and N=1,641 for the full survey including preference modules. Missing values reflect "do not know", "not applicable", or inconsistent answers in the case of risk preferences. All variables presented here were measured in the 2009 wave of the panel survey, except time preferences, which are taken from the 2010 survey wave since the 2009 wave did not include the far frame for females. Present-biased (future-biased) is a dummy equal to one if the response to the near frame is more impatient (patient) than the response to the far frame. Minimum switch-point is the lower bound of the interval in which the respondent switched to preferring the one-month-later payment compared to a 2000 MK payment on the earlier date. 150 MK $\approx$ 1 USD at the time of the 2009 and 2010 surveys. Risk-neutral is a dummy equal to one if the respondent prefers a 50-50 lottery to its expected value for certain, and thus could indicate risk-neutral or risk-seeking behaviour. Female HH decision-making power is constructed from questions over four types of economic decisions, scoring one if the female has some say in the decision and two if she has complete control. Food security poor is equal to one if the household reports consuming fewer than three meals yesterday.

Our main estimating equations are undirected dyadic logit models, with observations at the dyad level. These take the following form:

$$\begin{aligned}\Pr(D_{ij} = 1 | D_i = 1 \;\&\; D_j = 1; \mathbf{Z_i}, \mathbf{Z_j}, \mathbf{W_{ij}}, \nu) \\ = \Pr(\alpha + \boldsymbol{\beta}|\mathbf{Z_i} - \mathbf{Z_j}| + \boldsymbol{\gamma}(\mathbf{Z_i} + \mathbf{Z_j}) + \boldsymbol{\delta}\mathbf{W_{ij}} + \mu_\nu + \varepsilon_{ij} > 0) \quad (35)\end{aligned}$$

where D_i and D_j denote dummies equal to one if i and j are members of some VSLA group,[44] and D_{ij} is a dummy equal to one if i and j are members of the same group. $\mathbf{Z_i}$ and $\mathbf{Z_j}$ are vectors of i's and j's individual characteristics, in which we include measures of present-bias. We also seek to minimize omitted variable bias by controlling for a rich set of characteristics, which might also drive sorting and might be correlated with occupation and present-bias. Sections 4.4 and 4.5 detail the full set of controls used in the main specifications and robustness checks. $\mathbf{W_{ij}}$ is a vector of characteristics of the dyad, including whether i and j share the same category of occupation. μ_ν is a village fixed effect.[45] This controls for the average probability of matching in the village, which depends on the number of groups and also on the relative size of each group. The inclusion of village fixed effects also absorbs a range of factors which might affect the probability of being in the same group but remain constant at the village level, for example whether the village is served by other NGO programmes. ε_{ij} is a dyad-specific error term, which we assume takes a logistic distribution. We cluster standard errors at the village level in all estimations.[46]

It follows from the logic outlined above that an estimate of $\hat{\beta} < 0$ indicates positive assortative matching on the characteristic in question, whilst an estimate of $\hat{\beta} > 0$ indicates negative assortative matching on that characteristic. Since we estimate Equation 35 on a sample which only includes individuals who are members of at least one group, an estimate of $\hat{\gamma} > 0$ indicates that conditional on being member of at least one group, individuals with a high value of that particular variable are more likely to be members of more than one group.[47]

44 This is for notational completeness: by construction both dummies will always be equal to one in our analysis, since our data contains only members.

45 For notational convenience, we omit ν subscripts from all regressors except the fixed effects.

46 This is more conservative than the more common method of clustering by dyad (Fafchamps and Gubert, 2007).

47 We observe 146 individuals who are members of more than one group in the 2013 member census.

This increases the probability that such individuals are in the same group as a randomly-chosen other member, simply because such individuals are members of more groups.

4.4.2 *Sorting on occupation*

We begin by estimating Equation 35 on the full 2013 member census.[48] Table 60 describes the results. Most strikingly for the predictions in Section 4.2.2.1, there is evidence of strong positive assortative matching on occupation: if two individuals share the same occupation then they are 8.6 percentage points more likely to be members of the same group (p-value < 0.01). This is a large effect, equivalent to 53% of the baseline probability of being in the same group (16.1%). It therefore appear that the possibility of financial intermediation across farmers and non-farmers as outlined in scenario one, via negative assortative matching, is not being realised. Instead, evidence of positive assorting suggests that scenario two holds: informational or enforcement frictions are lower, or social benefits of participating in VSLAs are higher, within occupational groups than across them. This appears to dominate as a driver of sorting.

The other large effect in terms of size is that we observe positive assortative matching on gender: *ceteris paribus*, a male and a female are 5.3 percentage points less likely to be members of the same group than an all-male pair or an all-female pair are (p-value < 0.01). Female-headed households are also more likely to group together, although the effect size is just 1.2 percentage points (p-value < 0.01). A number of other characteristics are also highly significant, although the estimated marginal effects are small. Specifically, we observe positive assortative matching on: age; whether a member's spouse's father is relatively well off and whether a member's own father is relatively well-off compared to the rest of the village (proxies of exogenous wealth, or social class more generally); the number of goats the member's household possesses and whether her household owns a bicycle (standard poverty indicators

48 In our main analyses as we include all individuals who have ever been a member of the VSLA groups surveyed in 2013. However, our results are all robust to including only those members who are still current members in 2013, see Section 4.5.

Table 60: Dyadic regressions – 2013 member census

	(1) **Full member census** Mfx / (s.e.)
Occupation	
Same economic activity	0.086***
	(0.008)
Absolute differences	
Male	-0.053***
	(0.015)
Female-headed household	-0.012***
	(0.004)
Age	-0.001***
	(0.000)
Some post-primary educ.	-0.008
	(0.005)
Literate (read & understand newspaper)	-0.013*
	(0.007)
Father well-off in village (scale 1-5)	-0.013***
	(0.002)
Spouse's father well-off in village (scale 1-5)	-0.013***
	(0.003)
Household well-off in group (scale 1-9)	-0.010***
	(0.002)
HH owns a bicycle	-0.003
	(0.003)
# Goats	-0.006***
	(0.001)
Sums	✓
Village f.e.'s	✓
Observations	219747
Pseudo R^2	0.129
Baseline predicted probability	0.161

Notes: *, ** and *** represent p<0.10, p<0.05 and p<0.01 respectively. All variables from the 2013 member census, N=3,801 members. 519 of these members were no longer active, but are included here to avoid selection bias. Results are robust to excluding these individuals, see Section 4.5. All possible dyads in which both individuals live in the same village are constructed, N=289,914. Missing observations reflect answers of "do not know" or "not applicable". Occupation denotes an individual's primary economic activity, if engaged in multiple activities. Reported effects are marginal effects estimated at the mean.

for this region of Malawi); and whether a member's household is reported to be well-off at least compared to the rest of the group.[49] Such positive assorting may take place purely due to homophilous preferences, i.e. if members prefer to interact with other group members who are similar. Alternatively, it may be that members with similar social characteristics have similar financial needs, and so can more easily agree on the value of weekly shares, the size of the interest rate, and so on.

4.4.3 *Sorting on time preferences*

To test for evidence of sorting on present-bias, we re-estimate Equation 35 for the subsample of matched individuals whose time preferences were measured in the 2009-11 panel. Table 61 shows the effect of adding the measure of "present-bias" for these individuals. The key result is that we see strong evidence of negative assortative matching on "present-bias": the absolute difference between two members' "present-bias" carries a large, positive coefficient of 16.6 percentage points (p-value 0.013). This is consistent with the idea outlined in Section 4.2.2.2, that present-biased individuals who seek a commitment savings device sort into groups with time-consistent individuals who seek access to credit. The result is particularly striking given that we observe positive assortative matching on almost every other characteristic.

We still observe a large, positive effect of two individuals having the same primary occupation: 18.1 percentage points (p-value 0.019), which is again equivalent to over half the baseline probability of two members being in the same group in this subsample. The pattern of coefficients for the other controls is also very similar to that obtained in the full sample. Table 65 in Appendix 4.7 formally tests for equality of coefficients across the full sample and the matched subsample (excluding the measure of "present-bias", since this is not available for the full sample) and finds few significant differences. Thus it appears that the sorting observed in the matched subsample is broadly representative of the sorting observed in the full census.

49 Given that "household well-off in group" is a within-group ranking, we would expect its coefficient to be biased towards a positive value. The negative coefficient therefore suggests that individuals understood this question to be more about absolute consumption.

This suggests that the negative sorting on "present-bias" observed in the matched subsample would also hold in the full 2013 census, if measures of "present-bias" were available for all individuals. The results are also robust to re-weighting the estimations in order to make the matched subsample exactly representative of the full sample, see Section 4.5 for details.

Columns (1)-(5) of Table 62 confirm that the negative assortative matching on "present-bias" is not driven by matching on short-run or long-discount rates (or marginal rates of intertemporal substitution).[50] Column (1) repeats the preferred specification shown in Table 61 for comparison. Columns (2) and (3) show that we observe no sorting on the respondent's switch-point in the near or the far frame respectively. Similarly, column (4) shows that there is no evidence of sorting on whether the respondent is below or above the median patience in the near frame or the far frame. Thus sorting really does seem to be driven by the fact that "present-biased" individuals' long-run choices are more patient than their short-run choices — whilst "time-consistent" individuals' choices are equally impatient in the near and the far frame — as opposed to the fact that "present-biased" individuals have impatient short-run choices or patient long-run choices.

However, as discussed in Section 4.3.4.2, it is still possible that our measure of "present-bias" is instead capturing individuals who were expecting to have a lower marginal rate of intertemporal substitution in the future. If so, such individuals should have had a demand for borrowing, at least when the VSLAs were first formed. Conversely, individuals classified as "future-biased" may in fact have been anticipating a higher marginal rate of intertemporal substitution in the future, and thus may have had a demand for saving. If so, we would expect to see two things. First, in terms of who sorts into groups with "time-consistent" individuals, we would now expect those individuals spuriously classified as "future-biased" to do so. This is because individuals classified as "future-biased" would now be the ones providing savings, which "time-consistent" individuals could borrow if investment opportunities arose. We would therefore observe negative assortative matching on "future-bias". Nonetheless, we may still observe negative assortative matching on "present-bias" as well, if some "time-consistent" individuals are willing to lend to the individuals spuriously classified as

50 Table 67 in Appendix 4.7 confirms that the results in Table 62 also hold when the subsample is re-weighted.

Table 61: Dyadic regressions – matched subsample

	(1) **Matched subsample** Mfx / (s.e.)
Occupation	
Same economic activity	0.181**
	(0.077)
Absolute differences	
Present-biased	0.166**
	(0.067)
Male	-0.004
	(0.035)
Female-headed household	-0.176**
	(0.070)
Age	-0.006**
	(0.003)
Some post-primary educ.	-0.038
	(0.054)
Literate (read & understand newspaper)	-0.033
	(0.042)
Father well-off in village (scale 1-5)	-0.020
	(0.012)
Spouse's father well-off in village (scale 1-5)	-0.043***
	(0.014)
Household well-off in group (scale 1-9)	-0.044**
	(0.019)
HH owns a bicycle	0.008
	(0.030)
# Goats	-0.013*
	(0.007)
Sums	✓
Village f.e.'s	✓
Observations	1292
Pseudo R^2	0.222
Baseline predicted probability	0.296

Notes: *, ** and *** represent p<0.10, p<0.05 and p<0.01 respectively. N=722 individuals are matched from the 2013 census to the 2009-11 panel data. N=383 of these individuals are matched to the longer panel survey including preference modules. All possible dyads in which both individuals live in the same village are constructed, N=7,314 for the general survey and N=1,641 for the full survey including preference modules. Missing values reflect "do not know", "not applicable", or inconsistent answers in the case of risk preferences. Time preferences are taken from the 2010 survey, wave since the 2009 wave did not include the far frame for females. Present-biased (future-biased) is a dummy equal to one if the response to the near frame is more impatient (patient) than the response to the far frame. Reported effects are marginal effects estimated at the mean.

"present-biased" as long as a sufficiently high interest rate can be agreed. Second, we would now expect to see the strongest matching between "present-biased" individuals, who in fact are individuals with a demand for credit, and "future-biased" individuals, who in fact are individuals with a demand for saving. That is, we would observe negative assortative matching on "present-bias" particularly when "time-consistent" individuals are dropped from the sample.

Columns (5)-(7) of Table 62 show that neither of these predictions are supported by the data. Column (5) shows that, unlike "present-bias", there is no observed sorting on "future-bias". Column (6) reports the results when "time-consistent" individuals are dropped from the sample, and thus only "present-biased" and "future-biased individuals" remain. There is no significant evidence of negative assortative matching within this subsample, i.e. no evidence that "present-biased" individuals sort into groups with "future-biased" individuals.[51] In contrast, column (7) reports the results when "future-biased" individuals are dropped, leaving a sample of only "present-biased" and "time-consistent" individuals. The estimated coefficient on "present-biased" is large and highly significant. Thus there is strong evidence that "present-biased" individuals are matching with "time-consistent" individuals, and that this is driving the matching observed in the full sample. This pattern of results is thus much more in line with the predictions in Section 4.4.3, and thus the idea that our measure of "present-bias" is truly capturing individuals with present-biased preferences.

4.5 ROBUSTNESS

ADDITIONAL CONTROLS: The coefficients on economic activity and "present-bias" are robust to the inclusion of a host of additional controls from the 2009-11 panel survey, as shown in Table 63. Column (1) repeats the preferred specification shown in Table 61 for reference. Column (2) shows that the result of negative assorting on present-bias holds unconditional on

51 The sample size does drop considerably in column (6). However, the point estimate is much smaller, whilst standard errors do not increase substantially.

Table 62: Dyadic regressions – time preference measures, matched subsample

	(1) Mfx / (s.e.)	(2) Mfx / (s.e.)	(3) Mfx / (s.e.)	(4) Mfx / (s.e.)	(5) Mfx / (s.e.)	(6) Mfx / (s.e.)	(7) Mfx / (s.e.)
Occupation							
Same economic activity	0.181**	0.175**	0.180**	0.187**	0.178**	0.108	0.145
	(0.077)	(0.081)	(0.079)	(0.079)	(0.077)	(0.073)	(0.106)
Absolute differences							
Present-biased	0.166**		0.179**	0.175**	0.168**	0.055	0.195**
	(0.067)		(0.073)	(0.072)	(0.067)	(0.060)	(0.083)
Future-biased					-0.017		
					(0.036)		
Minimum switch-point (MK), far frame		0.000	0.001				
		(0.005)	(0.005)				
Minimum switch-point (MK), far frame		-0.003	-0.005				
		(0.006)	(0.007)				
Patience above median, near frame				0.011			
				(0.036)			
Patience above median, far frame				-0.032			
				(0.031)			
Wealth & income controls (abs. diffs)	✓	✓	✓	✓	✓	✓	✓
Demographic controls (abs. diffs)	✓	✓	✓	✓	✓	✓	✓
Sums	✓	✓	✓	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓	✓	✓	✓
Observations	1292	1292	1292	1292	1292	184	665
Pseudo R^2	0.222	0.220	0.225	0.226	0.224	0.331	0.233
Baseline predicted probability	0.296	0.296	0.296	0.296	0.296	0.273	0.293

Notes: *, ** and *** represent p<0.10, p<0.05 and p<0.01 respectively. N=722 individuals are matched from the 2013 census to the 2009-11 panel data. N=383 of these individuals are matched to the longer panel survey including preference modules. All possible dyads in which both individuals live in the same village are constructed, N=7,314 for the general survey and N=1,641 for the full survey including preference modules. Missing values reflect "do not know", "not applicable", or inconsistent answers in the case of risk preferences. All variables presented here were measured in the 2009 wave of the panel survey, except time preferences, which are taken from the 2010 survey wave since the 2009 wave did not include the far frame for females. Present-biased (future-biased) is a dummy equal to one if the response to the near frame is more impatient (patient) than the response to the far frame. Minimum switch-point is the lower bound of the interval in which the respondent switched to preferring the payment dated one month later compared to a 2000 MK payment on the earlier date. 150 MK $\approx$ 1 USD at the time of the 2009 and 2010 surveys. Patience above median is a dummy for having a switch-point below the median in that frame. Column (6) restricts the sample to "present-biased" and "future-biased" individuals, dropping "time-consistent" individuals. Column (7) restricts the sample to "present-biased" and "time-consistent" individuals, dropping "future-biased" individuals. Reported effects are marginal effects estimated at the mean.

occupation. Columns (3) and (4) show there is no evidence of sorting on being risk-neutral as compared to risk-averse, or indeed on the degree of risk-aversion. Columns (5) and (6) show that there is no evidence of sorting on measures of intra-household bargaining, either unconditionally or conditional on the measure of present-bias. This suggests that participants do not sort on "other-control" motives, in contrast to the strong sorting on "self-control" motives as proxied by present-bias.[52] Finally, column (7) shows that there is significant evidence of negative assorting on whether the individual comes from a household that speaks at village meetings — a proxy of how active or powerful the household is in local civil society. This is intuitive, as it is likely that certain "leaders" encourage other "followers" to join within each group. However, this does not change the estimated effect of present-bias, suggesting the two effects are uncorrelated.

WEIGHTING: As explained in Section 4.3.2, the 2009-11 panel survey was enumerated on a random sample of the population in each of the villages in the survey area. Yet this does not mean that the subset of 2013 members that can be matched back to the 2009-11 panel are a random subset of all 2013 members. Table 64 in Appendix 4.7shows that the matched dyads are slightly unrepresentative of the full population of 2013 dyads. However, Table 65 shows that only the estimated coefficients on gender and female-headed household are significantly different across the matched subsample and the full sample, and they maintain the same sign and significance. Nevertheless, as a robustness check, we re-weight the estimations in order to estimate the effect sizes we hypothetically would have obtained if the matched subsample were a fully random subsample of all 2013 members. To construct the weights, we first estimate a probit equation for the probability of each dyad in the 2013 data also appearing in 2009-11 data, as function of the full set of dyad characteristics listed in Table 58 and

52 The coefficients on the dyadic sums of "ever hides money from spouse" and "female HH decision-making power" are also insignificant (not shown). Thus there is no evidence that women with lower or higher household bargaining power join more groups conditional on being a member of at least one group — our best proxy of the strength of demand for VSLA participation. Other literature has examined effects of the square of household decision-making (Anderson and Baland, 2002). However, this does not make sense in a dyadic setting, as the sum or difference of two individuals' squared decision-making has no natural interpretation.

Table 63: Dyadic regressions – effects of controls, matched subsample

	(1) Mfx / (s.e.)	(2) Mfx / (s.e.)	(3) Mfx / (s.e.)	(4) Mfx / (s.e.)	(5) Mfx / (s.e.)	(6) Mfx / (s.e.)	(7) Mfx / (s.e.)
Occupation							
Same economic activity	0.181**		0.204**	0.207**	0.150**	0.154**	0.181**
	(0.077)		(0.097)	(0.096)	(0.075)	(0.072)	(0.075)
Absolute differences							
Present-biased	0.166**	0.164***	0.107*	0.115**		0.188**	0.167***
	(0.067)	(0.063)	(0.055)	(0.050)		(0.077)	(0.064)
Risk-neutral			0.016				
			(0.058)				
Risk aversion above median				-0.019			
				(0.039)			
Ever hides money from spouse					-0.006	-0.017	
					(0.029)	(0.027)	
Female HH decision-making power (index 0-8)					0.006	0.006	
					(0.011)	(0.012)	
HH important in village decisions (scale 1-6)							-0.038**
							(0.019)
HH speaks at village meetings							0.009
							(0.031)
Wealth & income controls (abs. diffs)	✓	✓	✓	✓	✓	✓	✓
Demographic controls (abs. diffs)	✓	✓	✓	✓	✓	✓	✓
Sums	✓	✓	✓	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓	✓	✓	✓
Observations	1292	1292	1031	1031	980	958	1277
Pseudo R^2	0.222	0.212	0.214	0.216	0.194	0.206	0.228
Baseline predicted probability	0.296	0.296	0.285	0.285	0.294	0.295	0.292

Notes: *, ** and *** represent p<0.10, p<0.05 and p<0.01 respectively. N=722 individuals are matched from the 2013 census to the 2009-11 panel data. N=383 of these individuals are matched to the longer panel survey including preference modules. All possible dyads in which both individuals live in the same village are constructed, N=7,314 for the general survey and N=1,641 for the full survey including preference modules. Missing values reflect "do not know", "not applicable", or inconsistent answers in the case of risk preferences. All variables presented here were measured in the 2009 wave of the panel survey, except time preferences, which are taken from the 2010 survey wave since the 2009 wave did not include the far frame for females. Present-biased (future-biased) is a dummy equal to one if the response to the near frame is more impatient (patient) than the response to the far frame. Risk-neutral is a dummy equal to one if the respondent prefers a 50-50 lottery to its expected value for certain, and thus could indicate risk-neutral or risk-seeking behaviour. Female HH decision-making power is constructed from questions over four types of economic decisions, scoring one if the female has some say in the decision and two if she has complete control. Reported effects are marginal effects estimated at the mean.

village fixed effects (results available on request).[53] We then use the estimated coefficients to generate the predicted probability that each 2013 dyad is matched to the 2009-11 data, and take the inverse of this predicted probability as the dyad-specific weight. We then re-run the various estimations shown in Table 63 re-weighted by these weights.[54] Tables 66, 67 and 68 in Appendix 4.7show that re-weighting slightly reduces the size of the estimated coefficients on occupation and present-bias, but never significantly so, and that they always maintain the same level of significance.

OTHER ROBUSTNESS CHECKS: We run a set of further robustness checks on both the full sample and the matched subsample, results of which are available on request. Results do not significantly change if we restrict the samples to current members only (i.e. dropping any individuals who are listed in the 2013 member census as having been members but having left), and if we drop the eight groups which had disbanded in 2013.[55] Nor do the results significantly change if we restrict attention to dyads in which both members are female; thus the pattern of sorting holds irrespective of gender. If we split "same occupation" and separately estimate the effect of both being a farmer, or of both being engaged in business or trade, both coefficients are still highly positive and significant, and not significantly different across the two occupations. Thus the effect of "same occupation" is not being driven by farmers or non-farmers in particular.

For the subsample of matched individuals, we run additional estimations including further controls for income and food security from the 2009-11 panel (as described in Table 55). We also re-estimate the main specifications using measures of individuals' occupations as recorded in the 2009 baseline rather than in the member census in 2013. The coefficients on same occupation and absolute difference in "present-bias" remain positive and highly significant. Finally, we re-run the estimations dropping anyone who never switched to the later amount

53 This is not equivalent to the the product of the separate probabilities that individual i and individual j are matched to the 2009-11 data, because of differences in the number of groups per village and members per group, and hence in the number of dyads across villages and groups.

54 To our knowledge this is the first use of inverse probability weighting in a dyadic specification. Such a strategy is valid since inverse probability weighting is valid for all M-estimators, of which logit models are a type.

55 We still obtained the membership data for these disbanded groups, as we wanted to avoid a selected sample of groups when examining sorting.

in one or both time frames in the time preference elicitation activity — a potential sign of misunderstanding or lack of trust. The coefficient on difference in "present-bias" remains positive and highly significant.

4.6 CONCLUSION

This paper has highlighted the potential role of VSLAs and similar groups for financial intermediation, especially in communities with low access to formal financial services. In particular, we have emphasised that VSLAs offer not only a source of credit but also a commitment savings technology. In light of this, we have asked whether individuals who have a demand for saving, and in particular commitment saving, successfully sort into groups with individuals who have a demand for borrowing. This issue has largely been overlooked in the literature on VSLA-type groups, but is likely to be crucial in determining the efficiency and equity of the financial outcomes that they generate for their members.

The fact that we find positive assortative matching on occupation suggests that members may have high costs of screening, monitoring or enforcing loan repayment when other group members are from a different occupation. This limits the scope for VSLAs to provide intermediation across occupational groups. Furthermore, positive assortative matching on occupation implies that VSLA groups may be heavily exposed to common shocks. This could increase competition for emergency consumption loans after negative covariate shocks, and more generally could increase the probability that several borrowers default at once. A possible solution may be to integrate VSLAs and similar groups into larger credit unions, or indeed to channel government funding or additional funding from MFIs through such groups. Something similar has already been achieved with Self-Help Groups in India, and more research is needed to understand whether and how this has been successful. Moreover, it is important to ask whether and how it has changed the nature of who participates, and how members sort across groups.

On the other hand, the fact that we observe negative assortative matching on a measure of "present-bias" suggests that VSLAs are already bringing together present-biased commitment savers with time-consistent individuals including prospective borrowers. This may enhance efficiency by reallocating capital from savers to borrowers. Nevertheless, there may be a concern about equity if present-biased individuals' willingness to pay for commitment leads them to accept very low interest rates when lending out their savings to borrowers. There is nothing exploitative *per se* in allowing sophisticated present-biased individuals to pay for commitment, in this case by accepting a lower interest rate.[56] However, if savers were receiving a return of close to zero or even negative values, or were experiencing high rates of default from borrowers without apparently being compensated for that risk through a high loan interest rate, then intuitively this could be grounds for concern.[57]

Reassuringly, we see that monthly interest rates on lending are broadly in line with discount rates, and that savers earn substantial positive nominal interest on their savings. Thus it does not appear that commitment savers are paying a high price for commitment. However, the risk of loan defaults should also be accounted for in any welfare calculations. Moreover, a fuller assessment of the welfare implications of VSLA-type institutions — and indeed their cost-benefit implications as an intervention — would need to take into account the returns on the projects that these groups fund. These remain interesting topics for future research.

56 Heidhues and Koszegi (2010) show that lenders can exploit partial naifs. However, in the context of VSLAs such exploitation of partial naiveté is less of a concern, since there are no penalties for failing to meet the minimum share-purchase requirement each week.

57 On the other hand, it is not clear what commitment savers' outside option would be in the absence of VSLAs, and thus what the correct counterfactual is in assessing whether commitment savers are receiving "too low" an interest rate.

4.7 APPENDICES

Table 64: Test of representativeness – matched dyads vs. 2013 census dyads

	Mean All Dyads	Mean Matched Dyads	Diff.	Std. Error	T stat	Full N	Matched N
Group membership							
Same VSLA group	0.17	0.30	-0.13***	(0.01)	-24.27	289914	7309
Occupation							
Same economic activity	0.56	0.71	-0.15***	(0.01)	-27.11	289914	7309
Absolute differences							
Male	0.38	0.39	-0.02**	(0.01)	-2.89	289467	7309
Female-headed household	0.32	0.34	-0.03***	(0.01)	-4.64	288740	7309
Age	12.64	12.17	0.46***	(0.12)	3.94	286763	7309
Some post-primary educ.	0.33	0.26	0.07***	(0.01)	13.29	289914	7309
Literate (read & understand newspaper)	0.26	0.29	-0.03***	(0.01)	-5.12	288342	7309
Father well-off in village (scale 1-5)	1.41	1.42	-0.00	(0.01)	-0.12	253041	6415
Spouse's father well-off in village (scale 1-5)	1.43	1.40	0.03*	(0.01)	2.44	252485	6954
Household well-off in group (scale 1-9)	1.40	1.29	0.11***	(0.01)	8.31	284485	7309
HH owns a bicycle	0.46	0.46	0.01	(0.01)	1.21	288740	7309
# Goats	1.84	2.30	-0.45***	(0.04)	-11.14	288240	7309
Sums							
Farmer	1.37	1.61	-0.24***	(0.01)	-33.75	289914	7309
Businessperson	0.48	0.32	0.16***	(0.01)	24.67	289914	7309
Male	0.53	0.55	-0.02**	(0.01)	-2.68	289467	7309
Female-headed household	0.40	0.45	-0.05***	(0.01)	-6.69	288740	7309
Age	72.12	77.04	-4.92***	(0.19)	-25.52	286763	7309
Some post-primary educ.	0.43	0.31	0.11***	(0.01)	18.76	289914	7309
Literate (read & understand newspaper)	1.68	1.63	0.04***	(0.01)	6.87	288342	7309
Father well-off in village (scale 1-5)	6.76	6.64	0.12***	(0.02)	5.11	253041	6415
Spouse's father well-off in village (scale 1-5)	6.85	6.87	-0.02	(0.02)	-0.84	252485	6954
Household well-off in group (scale 1-9)	15.04	15.41	-0.37***	(0.02)	-18.06	284485	7309
HH owns a bicycle	0.99	1.09	-0.10***	(0.01)	-12.09	288740	7309
# Goats	2.44	3.32	-0.89***	(0.05)	-16.90	288240	7309

Notes: All variables from the 2013 member census, N=3,801 members. N=722 of these members could be matched by name to the 2009-11 panel dataset. All possible dyads in which both individuals live in the same village are constructed for the full sample (N=289,914) and for the matched subsample (N=7,301). Missing observations reflect answers of "do not know" or "not applicable". Occupation denotes an individual's primary economic activity, if engaged in multiple activities.

Table 65: Dyadic regressions – 2013 member census and matched subsample

	(1) **Full CSAE sample** Mfx / (s.e.)	(2) **Matched subsample** Mfx / (s.e.)	P-value $\beta_1 = \beta_2$
Occupation			
Same economic activity	0.086***	0.206***	0.132
	(0.008)	(0.045)	
Absolute differences			
Male	-0.053***	-0.034	0.054*
	(0.015)	(0.028)	
Female-headed household	-0.012***	-0.065***	0.034**
	(0.004)	(0.023)	
Age	-0.001***	-0.001	0.629
	(0.000)	(0.001)	
Some post-primary educ.	-0.008	-0.005	0.724
	(0.005)	(0.021)	
Literate (read & understand newspaper)	-0.013*	-0.009	0.417
	(0.0067)	(0.018)	
Father well-off in village (scale 1-5)	-0.013***	-0.025***	0.577
	(0.002)	(0.010)	
Spouse's father well-off in village (scale 1-5)	-0.013***	-0.030**	0.379
	(0.003)	(0.012)	
Household well-off in group (scale 1-9)	-0.010***	-0.020**	0.659
	(0.003)	(0.010)	
HH owns a bicycle	-0.003	0.009	0.159
	(0.003)	(0.009)	
# Goats	-0.006***	-0.012***	0.407
	(0.002)	(0.004)	
Sums	✓	✓	
Village f.e.'s	✓	✓	
Observations	219747	5878	
Pseudo R^2	0.129	0.142	
Baseline predicted probability	0.161	0.168	

Notes: *, ** and *** represent p<0.10, p<0.05 and p<0.01 respectively. All variables from the 2013 member census, N=3,801 members. N=722 of these members could be matched by name to the 2009-11 panel dataset. All possible dyads in which both individuals live in the same village are constructed for the full sample (N=289,914) and for the matched subsample (N=7,301). Missing observations reflect answers of "do not know" or "not applicable". Occupation denotes an individual's primary economic activity, if engaged in multiple activities. Reported effects are marginal effects estimated at the mean.

Table 66: Dyadic regressions – 2013 member census, weighted

	(1) **Matched subsample** Mfx / (s.e.)
Occupation	
Same economic activity	0.098***
	(0.035)
Absolute differences	
Present-biased	0.096***
	(0.032)
Male	0.000
	(0.014)
Female-headed household	-0.094***
	(0.032)
Age	-0.004***
	(0.001)
Some post-primary educ.	-0.013
	(0.025)
Literate (read & understand newspaper)	-0.035
	(0.028)
Father well-off in village (scale 1-5)	-0.009
	(0.006)
Spouse's father well-off in village (scale 1-5)	-0.016**
	(0.007)
Household well-off in group (scale 1-9)	-0.027***
	(0.008)
HH owns a bicycle	0.008
	(0.016)
# Goats	-0.005
	(0.004)
Sums	✓
Village f.e.'s	✓
Observations	1280
Pseudo R^2	0.280
Baseline predicted probability	0.197

Notes: *, ** and *** represent p<0.10, p<0.05 and p<0.01 respectively. N=722 individuals are matched from the 2013 census to the 2009-11 panel data. N=383 of these individuals are matched to the longer panel survey including preference modules. All possible dyads in which both individuals live in the same village are constructed, N=7,314 for the general survey and N=1,641 for the full survey including preference modules. Missing values reflect "do not know", "not applicable", or inconsistent answers in the case of risk preferences. Time preferences are taken from the 2010 survey, wave since the 2009 wave did not include the far frame for females. "Present-biased" ("future-biased") is a dummy equal to one if the response to the near frame is more impatient (patient) than the response to the far frame. Reported effects are marginal effects estimated at the mean.

Table 67: Dyadic regressions – time preference measures, matched subsample, weighted

	(1) Mfx / (s.e.)	(2) Mfx / (s.e.)	(3) Mfx / (s.e.)	(4) Mfx / (s.e.)	(5) Mfx / (s.e.)
Occupation					
Same economic activity	0.098***	0.093**	0.098***	0.102***	0.097***
	(0.035)	(0.038)	(0.034)	(0.034)	(0.035)
Absolute differences					
Present-biased	0.096***		0.099***	0.097***	0.096***
	(0.032)		(0.031)	(0.030)	(0.032)
Future-biased					0.002
					(0.015)
Minimum switch-point (MK), far frame		0.000	0.001		
		(0.002)	(0.002)		
Minimum switch-point (MK), far frame		0.000	-0.001		
		(0.003)	(0.003)		
Patience above median, near frame				0.015	
				(0.017)	
Patience above median, far frame				-0.022	
				(0.016)	
Wealth & income controls (abs. diffs)	✓	✓	✓	✓	✓
Demographic controls (abs. diffs)	✓	✓	✓	✓	✓
Sums	✓	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓	✓
Observations	1280	1280	1280	1280	1280
Pseudo R^2	0.280	0.272	0.282	0.286	0.281
Baseline predicted probability	0.197	0.196	0.197	0.197	0.197

Notes: *, ** and *** represent p<0.10, p<0.05 and p<0.01 respectively. N=722 individuals are matched from the 2013 census to the 2009-11 panel data. N=383 of these individuals are matched to the longer panel survey including preference modules. All possible dyads in which both individuals live in the same village are constructed, N=7,314 for the general survey and N=1,641 for the full survey including preference modules. Missing values reflect "do not know", "not applicable", or inconsistent answers in the case of risk preferences. All variables presented here were measured in the 2009 wave of the panel survey, except time preferences, which are taken from the 2010 survey wave since the 2009 wave did not include the far frame for females. "Present-biased" ("future-biased") is a dummy equal to one if the response to the near frame is more impatient (patient) than the response to the far frame. Minimum switch-point is the lower bound of the interval in which the respondent switched to preferring the payment dated one month later compared to a 2000 MK payment on the earlier date. 150 MK $\approx$ 1 USD at the time of the 2009 and 2010 surveys. Patience above median is a dummy for having a switch-point below the median in that frame. Reported effects are marginal effects estimated at the mean.

Table 68: Dyadic regressions – effects of controls, matched subsample, weighted

	(1) Mfx / (s.e.)	(2) Mfx / (s.e.)	(3) Mfx / (s.e.)	(4) Mfx / (s.e.)	(5) Mfx / (s.e.)	(6) Mfx / (s.e.)	(7) Mfx / (s.e.)
Occupation							
Same economic activity	0.098***		0.117**	0.119**	0.088**	0.094***	0.098***
	(0.035)		(0.049)	(0.050)	(0.038)	(0.035)	(0.035)
Absolute differences							
Present-biased	0.096***	0.093***	0.052**	0.057**		0.117***	0.096***
	(0.032)	(0.028)	(0.022)	(0.023)		(0.035)	(0.031)
Risk-neutral			0.026				
			(0.036)				
Risk aversion above median				-0.020			
				(0.015)			
Ever hides money from spouse					-0.002	-0.005	
					(0.019)	(0.015)	
Female HH decision-making power (index 0-8)					0.005	0.003	
					(0.007)	(0.007)	
HH important in village decisions (scale 1-6)							-0.013
							(0.009)
HH speaks at village meetings							-0.001
							(0.012)
Wealth & income controls (abs. diffs)	✓	✓	✓	✓	✓	✓	✓
Demographic controls (abs. diffs)	✓	✓	✓	✓	✓	✓	✓
Sums	✓	✓	✓	✓	✓	✓	✓
Village f.e.'s	✓	✓	✓	✓	✓	✓	✓
Observations	1280	1280	1021	1021	969	947	1265
Pseudo R^2	0.280	0.264	0.270	0.271	0.242	0.268	0.285
Baseline predicted probability	0.197	0.191	0.189	0.189	0.193	0.197	0.194

Notes: *, ** and *** represent p<0.10, p<0.05 and p<0.01 respectively. N=722 individuals are matched from the 2013 census to the 2009-11 panel data. N=383 of these individuals are matched to the longer panel survey including preference modules. All possible dyads in which both individuals live in the same village are constructed, N=7,314 for the general survey and N=1,641 for the full survey including preference modules. Missing values reflect "do not know", "not applicable", or inconsistent answers in the case of risk preferences. All variables presented here were measured in the 2009 wave of the panel survey, except time preferences, which are taken from the 2010 survey wave since the 2009 wave did not include the far frame for females. "Present-biased" ("future-biased") is a dummy equal to one if the response to the near frame is more impatient (patient) than the response to the far frame. Risk-neutral is a dummy equal to one if the respondent prefers a 50-50 lottery to its expected value for certain, and thus could indicate risk-neutral or risk-seeking behaviour. Female HH decision-making power is constructed from questions over four types of economic decisions, scoring one if the female has some say in the decision and two if she has complete control. Reported effects are marginal effects estimated at the mean.

Part III

CONCLUSION

5

CONCLUSION

Each of the three chapters in this thesis has drawn its own specific conclusions, both judging the research question in light of the evidence presented, and making suggestions for policy and further related work. Those conclusions do not need repeating here. Instead, I take this opportunity to offer some broader thoughts on directions for future research.

This thesis has separately considered the roles of preferences, intra-household bargaining and informal institutions in relation to different cases of technology under-adoption. A promising approach going forward would be to study these different angles simultaneously for each case. An example is the recent work of Grant Miller and Mushfiq Mobarak on cookstoves in Bangladesh — a set of separate but closely related experiments disentangling the roles of information constraints, price, and gender differences in preferences and control over resources (Mobarak et al., 2012; Miller and Mobarak, 2013; Miller and Mobarak, 2014). This approach allows direct comparison of the relative importance of different constraints, and pinpoints which constraints must be alleviated either alone or simultaneously in order to stimulate adoption.

Relatedly, more research is needed on exactly whose preferences within the household must be targeted, and how, in order to promote technology adoption and other welfare-improving behaviours. The existing literature documenting that women have a stronger preference than men for household public goods tends to take these preferences as given, with the policy prescription being to place more income into the hands of women. Yet this is not always feasible, and in many situations a less costly option may be to target information or attitudinal campaigns specifically at men. For example, Stopnitzky (2017) estimates positive impacts on

latrine adoption from India's "no toilet, no bride" campaign — a series of adverts aimed at men which highlight the marriage market benefits of having a private latrine. Further, recent advertising campaigns in Sub-Saharan Africa have used male music and sports stars to promote adult male circumcision, following evidence that men's demand for circumcision is not easily shifted by information or price (Chinkhumba et al., 2014).[1] In a forthcoming field experiment in Pakistan, my co-authors and I will compare the effects of social norm interventions about domestic violence and child marriage, depending on whether they are targeted at women or at men.

Finally, whilst this thesis has considered the mediation of individuals' preferences through intra-household bargaining, very little is known about the possible reverse impact of intra-household bargaining on individuals' preferences. For example, repeated bargaining might bring partners' discount rates closer together, or allow a sophisticated partner to teach a naïve partner. A more extreme but sadly common example is the possible effect of intimate partner violence (IPV) on women's economic decision-making. IPV affects 30% of women who have been in a relationship globally, with figures much higher in some developing countries (al., 2013). It is plausible that chronic exposure to such violence may affect women's risk preferences, time preferences, and other factors affecting their economic choices. Yet this has received little attention in economic research, despite recent emphasis on the impacts of poverty, financial stress and conflict on preferences and decision-making (Mani et al., 2013; Mani and Lichand, 2016; Bauer et al., 2016). My co-authors and I are soon to run a lab-in-field experiment in Ethiopia, to examine the effects of priming women's experiences of IPV on their choices in incentivized tasks. More research on this topic would also be welcome.

1 See for example http://www.herald.co.zw/male-circumcision-advertising-cmpaign-wins-award/.

BIBLIOGRAPHY

Abdellaoui, Mohammed; Bleichrodt, Han; l'Haridon, Olivier, and Paraschiv, Corina (2013). "Is there one unifying concept of utility? An experimental comparison of utility under risk and utility over time". *Management Science* 59.9, pp. 2153–2169.

Adams, Abi; Cherchye, Laurens; De Rock, Bram, and Verriest, Ewout (2014). "Consume now or later? Time inconsistency, collective choice and revealed preference". *American Economic Review* 104.12, pp. 4147–83.

Afzal, Uzma; d'Adda, Giovanna; Fafchamps, Marcel; Quinn, Simon, and Said, Farah (2017). "Two sides of the same rupee? Comparing demand for microcredit and microsaving in a framed field experiment in rural Pakistan". *The Economic Journal*.

Agnew, Julie R; Anderson, Lisa R; Gerlach, Jeffrey R, and Szykman, Lisa R (2008). "Who chooses annuities? An experimental investigation of the role of gender, framing, and defaults". *The American Economic Review* 98.2, pp. 418–422.

Ahlin, Christian (2016). "Group lending, matching patterns, and the mystery of microcredit: Evidence from Thailand". *Working paper*.

Ambec, Stefan and Treich, Nicolas (2007). "ROSCAs as Financial Agreements to Cope with Self-control Problems". *Journal of Development Economics* 82.1, pp. 120–137.

Ambrus, Attila; Asgeirsdottir, T; Noor, Jawwad, and Sándor, László (2015). "Compensated Discount Functions – An Experiment on Integrating Rewards with Expected Income". *Working paper*.

Andersen, Steffen; Harrison, Glenn W; Lau, Morten Igel, and Rutström, E Elisabet (2006). "Elicitation using multiple price list formats". *Experimental Economics* 9.4, pp. 383–405.

Andersen, Steffen; Harrison, Glenn W; Lau, Morten I, and Rutström, E Elisabet (2008). "Eliciting risk and time preferences". *Econometrica* 76.3, pp. 583–618.

Andersen, Steffen; Harrison, Glenn; Lau, Morten, and Rutström, E Elisabet (2011). "Multiattribute Utility Theory, Intertemporal Utility and Correlation Aversion". *Working paper*.

Andersen, Steffen; Harrison, Glenn W; Lau, Morten I, and Rutström, E Elisabet (2013). "Discounting behaviour and the magnitude effect: evidence from a field experiment in Denmark". *Economica* 80.320, pp. 670–697.

Andersen, Steffen; Harrison, Glenn W; Lau, Morten I, and Rutström, E Elisabet (2014). "Discounting behavior: A reconsideration". *European Economic Review* 71, pp. 15–33.

Andersen, Steffen; Cox, James; Harrison, Glenn; Lau, Morten; Rutstroem, Elisabet, and Sadiraj, Vjollca (2015). "Asset integration and attitudes to risk: theory and evidence." *Working paper*.

Anderson, Michael L (2012). "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". *Journal of the American Statistical Association*.

Anderson, Siwan and Baland, Jean-Marie (2002). "The Economics of ROSCAs and Intrahousehold Resource Allocation". *The Quarterly Journal of Economics* 117.3, pp. 963–995.

Andreoni, James and Sprenger, Charles (2012a). "Estimating time preferences from convex budgets". *The American Economic Review* 102.7, pp. 3333–3356.

Andreoni, James and Sprenger, Charles (2012b). "Risk preferences are not time preferences". *The American Economic Review* 102.7, pp. 3357–3376.

Andreoni, James; Gravert, Christina; Kuhn, Mike; Saccardo, Silvia, and Yang, Yang (2017). "It's Just Money: Do Experimental Subjects Arbitrage Intertemporal Monetary Rewards?" *Working paper*.

Arcand, Jean-Louis and Fafchamps, Marcel (2012). "Matching in Community-Based Organizations". *Journal of Development Economics* 98.2, pp. 203–219.

Ariely, Dan and Loewenstein, George (2006). "The heat of the moment: The effect of sexual arousal on sexual decision making". *Journal of Behavioral Decision Making* 19.2, pp. 87–98.

Ashraf, Nava; Karlan, Dean, and Yin, Wesley (2006). "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines". *The Quarterly Journal of Economics* 121.2, pp. 635–672.

Ashraf, Nava; Field, Erica, and Lee, Jean (2014a). "Household bargaining and excess fertility: an experimental study in Zambia". *The American Economic Review* 104.7, pp. 2210–2237.

Ashraf, Nava; Bandiera, Oriana, and Jack, B Kelsey (2014b). "No margin, no mission? A field experiment on incentives for public service delivery". *Journal of Public Economics* 120, pp. 1–17.

Attanasio, Orazio P and Lechene, Valérie (2014). "Efficient responses to targeted cash transfers". *Journal of Political Economy* 122.1, pp. 178–222.

Attanasio, Orazio and Lechene, Valerie (2002). "Tests of income pooling in household decisions". *Review of Economic Dynamics* 5.4, pp. 720–748.

Attanasio, Orazio; Barr, Abigail; Cardenas, Juan Camilo; Genicot, Garance, and Meghir, Costas (2012). "Risk Pooling, Risk Preferences, and Social Networks". *American Economic Journal: Applied Economics* 4.2, pp. 134–67.

Augenblick, Ned and Rabin, Matthew (2015). "An Experiment on Time Preference and Misprediction in Unpleasant Tasks". *Working paper*.

Augenblick, Ned; Niederle, Muriel, and Sprenger, Charles (2015). "Working over time: dynamic inconsistency in real effort tasks". *The Quarterly Journal of Economics* 1067, p. 1115.

Bailey, Martha J (2006). "More power to the pill: the impact of contraceptive freedom on women's life cycle labor supply". *The Quarterly Journal of Economics* 121.1, pp. 289–320.

Baird, Sarah; Ferreira, Francisco HG; Özler, Berk, and Woolcock, Michael (2013). "Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: a systematic review". *Campbell Systematic Reviews* 9.8.

Baird, Sarah; Gong, Erick; McIntosh, Craig, and Özler, Berk (2014). "The heterogeneous effects of HIV testing". *Journal of Health Economics* 37, pp. 98–112.

Balakrishnan, Uttara; Haushofer, Johannes, and Jakiela, Pamela (2015). "How Soon Is Now? Evidence of Present Bias from Convex Time Budget Experiments". *IZA Discussion Paper*.

Bandiera, Oriana; Buehren, Niklas; Burgess, Robin; Goldstein, Markus; Gulesci, Selim; Rasul, Imran; Sulaiman, Munshi, et al. (2015). "Women's economic empowerment in action: Evidence from a randomized control trial in Africa". *Working paper*.

Banerjee, Abhijit V and Mullainathan, Sendhil (2008). "Limited attention and income distribution". *The American Economic Review*, pp. 489–493.

Banerjee, Abhijit V; Besley, Timothy, and Guinnane, Timothy W (1994). "Thy Neighbor's Keeper: The Design of a Credit Cooperative with Theory and a Test". *The Quarterly Journal of Economics* 109.2, pp. 491–515.

Banerjee, Abhijit and Mullainathan, Sendhil (2010). "The shape of temptation: Implications for the economic lives of the poor". *NBER Working Papers*.

Banerjee, Abhijit; Karlan, Dean; Zinman, Jonathan; Duflo, Esther; Glennerster, Rachel; Kinnan, Cynthia; Tarozzi, Alessandro; Desai, Jaikishan; Johnson, Kristin; Attanasio, Orazio, et al. (2015). "American Economic Journal: Applied Economics: Vol. 7 No. 1 (January 2015)".

Barr, Abigail; Dekker, Marleen, and Fafchamps, Marcel (2012). "Who Shares Risk with Whom under Different Enforcement Mechanisms?" *Economic Development and Cultural Change* 60.4, pp. 677–706.

Basu, Karna (2011). "Hyperbolic Discounting and the Sustainability of Rotational Savings Arrangements". *American Economic Journal: Microeconomics* 3.4, pp. 143–71.

Bauer, Michal; Chytilova, Julie, and Morduch, Jonathan (2012). "Behavioral Foundations of Microcredit: Experimental and Survey Evidence from Rural India". *American Economic Review* 102.2, pp. 1118–39.

Bauer, Michal; Blattman, Christopher; Chytilová, Julie; Henrich, Joseph; Miguel, Edward, and Mitts, Tamar (2016). "Can war foster cooperation?" *The Journal of Economic Perspectives* 30.3, pp. 249–274.

Beaman, Lori A; Karlan, Dean, and Thuysbaert, Bram (2014). "Saving for a (not so) Rainy Day: A Randomized Evaluation of Savings Groups in Mali". *Yale University Economic Growth Center Discussion Paper* 1043.

Beksinska, Mags; Smit, Jenni; Joanis, Carol, and Hart, Catherine (2012). "Practice makes perfect: reduction in female condom failures and user problems with short-term experience in a randomized trial". *Contraception* 86.2, pp. 127–131.

Besley, Timothy and Coate, Stephen (1995). "Group Lending, Repayment Incentives and Social Collateral". *Journal of Development Economics* 46.1, pp. 1–18.

Besley, Timothy; Coate, Stephen, and Loury, Glenn (1993). "The Economics of Rotating Savings and Credit Associations". *American Economic Review* 83.4, pp. 792–810.

Bjorkman Nyqvist, Martina; Corno, Lucia; Walque, Damien de, and Svensson, Jakob (2015). "Using lotteries to incentivize safer sexual behavior: evidence from a randomized controlled trial on HIV prevention". *Working Paper*.

Breza, Emily and Chandrasekhar, Arun G (2015). "Social networks, reputation and commitment: Evidence from a savings monitors experiment". *NBER Working Papers* 21169.

Browning, Martin and Chiappori, Pierre-Andre (1998). "Efficient intra-household allocations: A general characterization and empirical tests". *Econometrica*, pp. 1241–1278.

Browning, Martin and Tobacman, Jeremy (2007). "Discounting and optimism equivalences". *Working paper*.

Bruhn, Miriam and McKenzie, David (2009). "In pursuit of balance: Randomization in practice in development field experiments". *American Economic Journal: Applied Economics* 1.4, pp. 200–232.

Brune, Lasse; Gine, Xavier; Goldberg, Jessica, and Yang, Dean (2011). "Commitments to Save: a Field Experiment in Rural Malawi". *World Bank Policy Research Working Papers* 5748.

Burlando, Alfredo; Canidio, Andrea, and Selby, Rebekah (2017). "The economics of savings groups". *Journal of Development Economics*.

Carvalho, Leandro S; Meier, Stephan, and Wang, Stephanie W (2016). "Poverty and economic decision-making: Evidence from changes in financial resources at payday". *The American Economic Review* 106.2, pp. 260–284.

Chiappori, Pierre-Andre (1992). "Collective labor supply and welfare". *Journal of Political Economy* 100.3, pp. 437–467.

Chiappori, Pierre-André and Oreffice, Sonia (2008). "Birth control and female empowerment: An equilibrium analysis". *Journal of Political Economy* 116.1, pp. 113–140.

Chinkhumba, Jobiba; Godlonton, Susan, and Thornton, Rebecca (2014). "The demand for medical male circumcision". *American Economic Journal: Applied Economics*, pp. 152–177.

Choi, Kyung-Hee; Wojcicki, Janet, and Valencia-Garcia, Dellanira (2004). "Introducing and negotiating the use of female condoms in sexual relationships: Qualitative interviews with women attending a family planning clinic". *AIDS and Behavior* 8.3, pp. 251–261.

Coller, Maribeth and Williams, Melonie B (1999). "Eliciting individual discount rates". *Experimental Economics* 2.2, pp. 107–127.

Coman, Emil N; Weeks, Margaret R; Yanovitzky, Itzhak; Iordache, Eugen; Barbour, Russell; Coman, Maria A, and Huedo-Medina, Tania B (2013). "The impact of information about the female condom on female condom use among males and females from a US urban community". *AIDS and Behavior* 17.6, pp. 2194–2201.

Crosby, RA; Graham, CA; Yarber, WL, and Sanders, SA (2004). "If the condom fits, wear it: a qualitative study of young African-American men". *Sexually Transmitted Infections* 80.4, pp. 306–309.

Cubitt, Robin P and Read, Daniel (2007). "Can intertemporal choice experiments elicit time preferences for consumption?" *Experimental Economics* 10.4, pp. 369–389.

De Walque, Damien; Dow, William H; Nathan, Rose; Abdul, Ramadhani; Abilahi, Faraji; Gong, Erick; Isdahl, Zachary; Jamison, Julian; Jullu, Boniphace; Krishnan, Suneeta, et al. (2012). "Incentivising safe sex: a randomised trial of conditional cash transfers for HIV and sexually transmitted infection prevention in rural Tanzania". *BMJ Open* 2.1.

Dean, Mark and Sautmann, Anja (2014). "Credit constraints and the measurement of time preferences". *Working paper*.

Deaton, Angus (1991). "Saving and liquidity constraints". *Econometrica* 59.5, pp. 1221–1248.

Decker, Michele R; Seage III, George R; Hemenway, David; Raj, Anita; Saggurti, Niranjan; Balaiah, Donta, and Silverman, Jay G (2009). "Intimate partner violence functions as both a risk marker and risk factor for women's HIV infection: findings from Indian husband-wife dyads". *Journal of Acquired Immune Deficiency Syndromes (1999)* 51.5, p. 593.

Demirgüç-Kunt, Asli; Klapper, Leora F; Singer, Dorothe, and Van Oudheusden, Peter (2015). "The global findex database 2014: Measuring financial inclusion around the world". *World Bank Policy Research Working Paper*.

Dohmen, Thomas; Falk, Armin; Huffman, David, and Sunde, Uwe (2010). "Are risk aversion and impatience related to cognitive ability?" *The American Economic Review* 100.3, pp. 1238–1260.

Dowdy, David W; Sweat, Michael D, and Holtgrave, David R (2006). "Country-wide distribution of the nitrile female condom (FC2) in Brazil and South Africa: a cost-effectiveness analysis". *AIDS* 20.16, pp. 2091–2098.

Duflo, Esther (2003). "Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa". *The World Bank Economic Review* 17.1, pp. 1–25.

Duflo, Esther and Udry, Christopher (2004). "Intrahousehold resource allocation in Cote d'Ivoire: Social norms, separate accounts and consumption choices". *NBER Working Papers*.

Duflo, Esther; Kremer, Michael, and Robinson, Jonathan (2011). "Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya". *The American Economic Review* 101.6, pp. 2350–2390.

Duflo, Esther; Dupas, Pascaline, and Kremer, Michael (2015). "Education, HIV, and early fertility: Experimental evidence from Kenya". *The American Economic Review* 105.9, pp. 2757–2797.

Dunkle, Kristin L; Jewkes, Rachel K; Brown, Heather C; Gray, Glenda E; McIntryre, James A, and Harlow, Siobán D (2004). "Gender-based violence, relationship power, and risk of HIV infection in women attending antenatal clinics in South Africa". *The Lancet* 363.9419, pp. 1415–1421.

Dupas, Pascaline (2011). "Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya". *American Economic Journal: Applied Economics* 3.1, pp. 1–34.

Dupas, Pascaline and Robinson, Jonathan (2013a). "Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya". *American Economic Journal: Applied Economics* 5.1, pp. 163–92.

Dupas, Pascaline and Robinson, Jonathan (2013b). "Why Don't the Poor Save More? Evidence from Health Savings Experiments". *American Economic Review* 103.4, pp. 1138–71.

Eckel, Catherine C and Grossman, Philip J (2008). "Men, women and risk aversion: Experimental evidence". *Handbook of Experimental Economics Results* 1, pp. 1061–1073.

Eeckhout, Jan and Munshi, Kaivan (2010). "Matching in Informal Financial Institutions". *Journal of the European Economic Association* 8.5, pp. 947–988.

Elster, Jon (1979). *Ulysses and the Sirens: Studies in Rationaltiy and Irrationality*. Cambridge University Press.

Epper, Thomas (2015). "Income Expectations, Limited Liquidity, and Anomalies in Intertemporal Choice". *Working paper*.

Eswaran, Mukesh (2002). "The empowerment of women, fertility, and child mortality: towards a theoretical analysis". *Journal of Population Economics* 15.3, pp. 433–454.

Fafchamps, Marcel and Gubert, Flore (2007). "The Formation of Risk Sharing Networks". *Journal of Development Economics* 83.2, pp. 326–350.

Farr, Gaston; Gabelnick, Henry; Sturgen, Kim, and Dorflinger, Laneta (1994). "Contraceptive efficacy and acceptability of the female condom". *American Journal of Public Health* 84.12, pp. 1960–1964.

Feigenberg, Benjamin; Field, Erica, and Pande, Rohini (2013). "The Economic Returns to Social Interaction: Experimental Evidence from Microfinance". *Review of Economic Studies* 80.4, pp. 1459–1483.

Flood, Michael (2003). "Lust, trust and latex: Why young heterosexual men do not use condoms". *Culture, Health & Sexuality* 5.4, pp. 353–369.

Fontanet, Arnaud L; Saba, Joseph; Chandelying, Verapol; Sakondhavat, Chuanchom; Bhiraleus, Praphas; Rugpao, Sungwal; Chongsomchai, Chompilas; Kiriwat, Orawan; Tovanabutra, Sodsai; Dally, Leonard, et al. (1998). "Protection against sexually transmitted diseases by granting sex workers in Thailand the choice of using the male or female condom: results from a randomized controlled trial". *AIDS* 12.14, pp. 1851–1859.

Fortin, Bernard and Lacroix, Guy (1997). "A test of the unitary and collective models of household labour supply". *The Economic Journal* 107.443, pp. 933–955.

Galárraga, Omar; Sosa-Rubí, Sandra G; Infante, César; Gertler, Paul J, and Bertozzi, Stefano M (2014). "Willingness-to-accept reductions in HIV risks: conditional economic incentives in Mexico". *The European Journal of Health Economics* 15.1, pp. 41–55.

Gerber, Anke and Rohde, Kirsten IM (2015). "Eliciting discount functions when baseline consumption changes over time". *Journal of Economic Behavior & Organization* 116, pp. 56–64.

Gertler, Paul; Shah, Manisha, and Bertozzi, Stefano M (2005). "Risky business: the market for unprotected commercial sex". *Journal of Political Economy* 113.3, pp. 518–550.

Ghatak, Maitreesh (2000). "Screening by the Company You Keep: Joint Liability Lending and the Peer Selection Effect". *The Economic Journal* 110.465, pp. 601–31.

Ghatak, Maitreesh and Guinnane, Timothy W (1999). "The Economics of Lending with Joint Liability: Theory and Practice". *Journal of Development Economics* 60.1, pp. 195–228.

Giné, Xavier; Goldberg, Jessica; Silverman, Dan, and Yang, Dean (2016). "Revising commitments: Field evidence on the adjustment of prior choices". *The Economic Journal*.

Go, Vivian F; Sethulakshmi, C Johnson; Bentley, Margaret E; Sivaram, Sudha; Srikrishnan, Aylur K; Solomon, Suniti, and Celentano, David D (2003). "When HIV-prevention messages and gender norms clash: the impact of domestic violence on women's HIV risk in slums of Chennai, India". *AIDS and Behavior* 7.3, pp. 263–272.

Godlonton, Susan; Munthali, Alister, and Thornton, Rebecca (2016). "Responding to risk: Circumcision, information, and HIV prevention". *Review of Economics and Statistics* 98.2, pp. 333–349.

Goldin, Claudia and Katz, Lawrence F (2002). "The power of the pill: Oral contraceptives and women's career and marriage decisions". *Journal of Political Economy* 110.4, pp. 730–770.

Greaney, Brian P; Kaboski, Joseph P, and Van Leemput, Eva (2016). "Can Self-Help Groups Really Be "Self-Help"?" *The Review of Economic Studies* 83.4, pp. 1614–1644.

Gul, Faruk and Pesendorfer, Wolfgang (2001). "Temptation and self-control". *Econometrica* 69.6, pp. 1403–1435.

Halevy, Yoram (2008). "Strotz meets Allais: Diminishing impatience and the certainty effect". *The American Economic Review*, pp. 1145–1162.

Halevy, Yoram (2015). "Time consistency: Stationarity and time invariance". *Econometrica* 83.1, pp. 335–352.

Harris, Christopher and Laibson, David (2001). "Dynamic choices of hyperbolic consumers". *Econometrica* 69.4, pp. 935–957.

Harrison, Glenn W; Lau, Morten I, and Williams, Melonie B (2002). "Estimating individual discount rates in Denmark: A field experiment". *American Economic Review*, pp. 1606–1617.

Harrison, Glenn W; Lau, Morten I, and Rutström, E Elisabet (2013). "Identifying time preferences with experiments: Comment". *Center for the Economic Analysis of Risk, Working Paper* 9.

Haushofer, Johannes (2014). "The Cost of Keeping Track". *Working paper*.

Haushofer, Johannes and Fehr, Ernst (2014). "On the psychology of poverty". *Science* 344.6186, pp. 862–867.

Heidhues, Paul and Koszegi, Boton (2010). "Exploiting naivete about self-control in the credit market". *The American Economic Review* 100.5, pp. 2279–2303.

Hirshleifer, Sarojini; McKenzie, David; Almeida, Rita, and Ridao-Cano, Cristobal (2015). "The impact of vocational training for the unemployed: experimental evidence from Turkey". *The Economic Journal*.

Holland, Janet; Ramazanoglu, Caroline; Scott, Sue; Sharpe, Sue, and Thomson, Rachel (1991). "Between embarrassment and trust: young women and the diversity of condom use". *AIDS: Responses, Interventions and Care*, pp. 127–148.

Hoppe, Marilyn J; Morrison, Diane M; Gillmore, Mary Rogers; Beadnell, Blair; Higa, Darrel H, and Leigh, Barbara C (2008). "Agreement of daily diary and retrospective measures of condom use". *AIDS and Behavior* 12.1, pp. 113–117.

Janssens, Wendy; Kramer, Berber; Swart, Lisette, et al. (2017). "Be patient when measuring Hyperbolic Discounting: Stationarity, Time Consistency and Time Invariance in a Field Experiment". *Journal of Development Economics*.

John, Anett (2017). "When Commitment Fails – Evidence from a Field Experiment". *Working paper*.

Kalichman, Seth C; Williams, Ernestine A; Cherry, Charsey; Belcher, Lisa, and Nachimson, Dena (1998). "Sexual coercion, domestic violence, and negotiating condom use among low-income African American women". *Journal of Women's Health* 7.3, pp. 371–378.

Kast, Felipe; Meier, Stephan, and Pomeranz, Dina (2012). "Under-Savers Anonymous: Evidence on Self-Help Groups and Peer Pressure as a Savings Commitment Device". *NBER Working Papers* 18417.

Kaur, Supreet; Kremer, Michael, and Mullainathan, Sendhil (2015). "Self-control at work". *Journal of Political Economy* 123.6, pp. 1227–1277.

Koster, Winny; Bruinderink, Marije Groot, and Janssens, Wendy (2015). "Empowering Women or Pleasing Men? Analyzing Male Views on Female Condom Use In Zimbabwe, Nigeria and Cameroon". *International Perspectives on Sexual and Reproductive Health* 41.3, pp. 126–135.

Ksoll, Christopher; Lilleør, Helene Bie; Lønborg, Jonas Helth, and Rasmussen, Ole Dahl (2016). "Impact of Village Savings and Loan Associations: Evidence from a cluster randomized trial". *Journal of Development Economics* 120, pp. 70–85.

Laibson, David (1997). "Golden Eggs and Hyperbolic Discounting". *The Quarterly Journal of Economics* 112.2, pp. 443–478.

Laibson, David (2015). "Why don't present-biased agents make commitments?" *The American Economic Review* 105.5, pp. 267–272.

Laury, Susan (2005). "Pay one or pay all: Random selection of one choice for payment". *Andrew Young School of Policy Studies Research Paper Series* 06-13.

Lee, David S (2009). "Training, wages, and sample selection: Estimating sharp bounds on treatment effects". *The Review of Economic Studies* 76.3, pp. 1071–1102.

MacPhail, Catherine and Campbell, Catherine (2001). "'I think condoms are good but, aai, I hate those things': condom use among adolescents and young people in a Southern African township". *Social Science & Medicine* 52.11, pp. 1613–1627.

Macaluso, Maurizio; Blackwell, Richard; Jamieson, Denise J; Kulczycki, Andrzej; Chen, Michael P; Akers, Rachel; Kim, Dhong-jin, and Duerr, Ann (2007). "Efficacy of the male latex condom and of the female polyurethane condom as barriers to semen during intercourse: a randomized clinical trial". *American Journal of Epidemiology* 166.1, pp. 88–96.

Macia, Manuel; Maharaj, Pranitha, and Gresh, Ashley (2011). "Masculinity and male sexual behaviour in Mozambique". *Culture, Health & Sexuality* 13.10, pp. 1181–1192.

Mani, Anandi and Lichand, Guilherme (2016). "Cognitive Droughts". *Working paper*.

Mani, Anandi; Mullainathan, Sendhil; Shafir, Eldar, and Zhao, Jiaying (2013). "Poverty impedes cognitive function". *Science* 341.6149, pp. 976–980.

Mantell, Joanne E; Smit, Jennifer A; Exner, Theresa M; Mabude, Zonke; Hoffman, Susie; Beksinska, Mags; Kelvin, Elizabeth A; Ngoloyi, Claudia; Leu, Cheng-Shiun, and Stein, Zena A (2015). "Promoting female condom use among female university students in KwaZulu-Natal, South Africa: results of a randomized behavioral trial". *AIDS and Behavior* 19.7, pp. 1129–1140.

McKenzie, David J (2017). "Identifying and spurring high-growth entrepreneurship: experimental evidence from a business plan competition". *American Economic Review (forthcoming)*.

McKenzie, David (2015). "Another reason to prefer Ancova: dealing with changes in measurement between baseline and follow-up". *Blog post*.

Miller, Grant and Mobarak, A Mushfiq (2013). "Gender differences in preferences, intrahousehold externalities, and low demand for improved cookstoves". *NBER Working Papers*.

Miller, Grant and Mobarak, A Mushfiq (2014). "Learning about new technologies through social networks: experimental evidence on nontraditional stoves in Bangladesh". *Marketing Science* 34.4, pp. 480–499.

Mobarak, Ahmed Mushfiq; Dwivedi, Puneet; Bailis, Robert; Hildemann, Lynn, and Miller, Grant (2012). "Low demand for nontraditional cookstove technologies". *Proceedings of the National Academy of Sciences* 109.27, pp. 10815–10820.

Noor, Jawwad (2009). "Hyperbolic discounting and the standard model: Eliciting discount functions". *Journal of Economic Theory* 144.5, pp. 2077–2083.

O'Donoghue, Ted and Rabin, Matthew (2015). "Present bias: Lessons learned and to be learned". *The American Economic Review* 105.5, pp. 273–279.

Paula, Áureo De di; Shapira, Gil, and Todd, Petra E (2014). "How beliefs about HIV status affect risky behaviors: Evidence from Malawi". *Journal of Applied Econometrics* 29.6, pp. 944–964.

Peters, Anny; Jansen, Willy, and Driel, Francien van (2010). "The female condom: the international denial of a strong potential". *Reproductive Health Matters* 18.35, pp. 119–128.

Phelps, Edmund S and Pollak, Robert A (1968). "On second-best national saving and game-equilibrium growth". *The Review of Economic Studies* 35.2, pp. 185–199.

Pilz, Kevin M. (2014). "Assessment of the female condom supply chain in Mozambique". *i+solutions Report*.

Quidt, Jonathan de; Fetzer, Thiemo, and Ghatak, Maitreesh (2016). "Market Structure and Borrower Welfare in Microfinance". *Working paper*.

Rasul, Imran (2008). "Household bargaining over fertility: Theory and evidence from Malaysia". *Journal of Development Economics* 86.2, pp. 215–241.

Read, Daniel and Van Leeuwen, Barbara (1998). "Predicting hunger: The effects of appetite and delay on choice". *Organizational behavior and human decision processes* 76.2, pp. 189–205.

Ruger, Jennifer Prah (2004). "Combating HIV/AIDS in developing countries: Requires empowering people to act on their own terms". *BMJ: British Medical Journal* 329.7458, p. 121.

Saúde, Instituto Nacional de Estatística Ministério da (2015). "Inquérito de Indicadores de Imunização Malária e HIV/SIDA em Moçambique: Relatório de Indicadores Básicos de HIV". *DHS Publications*. URL: https://dhsprogram.com/pubs/pdf/PR85/PR85.pdf.

Schaner, Simone G (2011). "The Cost of Convenience? Transaction Costs, Bargaining Power, and Savings Account Use in Kenya". *Working paper*.

Schaner, Simone (2015). "Do opposites detract? Intrahousehold preference heterogeneity and inefficient strategic savings". *American Economic Journal: Applied Economics* 7.2, pp. 135–174.

Schubert, Renate; Brown, Martin; Gysler, Matthias, and Brachinger, Hans Wolfgang (1999). "Financial decision-making: are women really more risk-averse?" *The American Economic Review* 89.2, pp. 381–385.

Silverman, Jay G; Decker, Michele R; Saggurti, Niranjan; Balaiah, Donta, and Raj, Anita (2008). "Intimate partner violence and HIV infection among married Indian women". *Jama* 300.6, pp. 703–710.

Sprenger, Charles (2015). "Judging experimental evidence on dynamic inconsistency". *The American Economic Review* 105.5, pp. 280–285.

Stalgaitis, Carolyn and Glick, Sara Nelson (2014). "The use of web-based diaries in sexual risk behaviour research: a systematic review". *Sexually Transmitted Infections* 90.5, pp. 374–381.

Stopnitzky, Yaniv (2017). "No toilet no bride? Intrahousehold bargaining in male-skewed marriage markets in India". *Journal of Development Economics* 127, pp. 269–282.

Tanaka, Tomomi; Camerer, Colin F, and Nguyen, Quang (2010). "Risk and time preferences: linking experimental and household survey data from Vietnam". *The American Economic Review* 100.1, pp. 557–571.

Tarozzi, Alessandro and Mahajan, Aprajit (2011). "Time inconsistency, expectations and technology adoption: The case of insecticide treated nets". *Working paper*.

Tavory, Iddo and Swidler, Ann (2009). "Condom semiotics: meaning and condom use in rural Malawi". *American Sociological Review* 74.2, pp. 171–189.

Thaler, Richard H (1991). "Some Empirical Evidence on Dynamic Inconsistency". *Quasi-Rational Economics* 1, pp. 127–136.

Thirumurthy, Harsha; Masters, Samuel H; Rao, Samwel; Bronson, Megan A; Lanham, Michele; Omanga, Eunice; Evens, Emily, and Agot, Kawango (2014). "Effect of providing conditional economic compensation on uptake of voluntary medical male circumcision in Kenya: a randomized clinical trial". *Jama* 312.7, pp. 703–711.

Thornton, Rebecca L (2008). "The demand for, and impact of, learning HIV status". *The American Economic Review* 98.5, pp. 1829–1863.

Toussaert, Séverine (2015). "Eliciting temptation and self-control through menu choices: a lab experiment". *Working Paper*.

Trussell, James (2011). "Contraceptive failure in the United States". *Contraception* 83, pp. 397–404.

UNAIDS (2016a). "HIV prevention among adolescent girls and young women". URL: http://www.unaids.org/sites/default/files/media_asset/UNAIDS_HIV_prevention_among_adolescent_girls_and_young_women.pdf.

UNAIDS (2016b). "UNAIDS Prevention Gap Report". URL: http://www.unaids.org/sites/default/files/media_asset/2016-prevention-gap-report_en.pdf.

Ubfal, Diego (2016). "How general are time preferences? Eliciting good-specific discount rates". *Journal of Development Economics* 118, pp. 150–170.

Udry, Christopher (1996). "Gender, agricultural production, and the theory of the household". *Journal of Political Economy* 104.5, pp. 1010–1046.

Van Tassel, Eric (2004). "Household bargaining and microfinance". *Journal of Development Economics* 74.2, pp. 449–468.

Vijayakumar, Gowri; Mabude, Zonke; Smit, Jenni; Beksinska, Mags, and Lurie, Mark (2006). "A review of female-condom effectiveness: patterns of use and impact on protected sex acts and STI incidence". *International Journal of STD & AIDS* 17.10, pp. 652–659.

Wanyenze, Rhoda K; Atuyambe, Lynn; Kibirige, Vista; Mbabazi, Sarah; Tumwesigye, Nazarius M; Djurhuus, Kara, and Namale, Alice (2011). "The new female condom (FC2) in Uganda: perceptions and experiences of users and their sexual partners". *African Journal of AIDS Research* 10.3, pp. 219–224.

al., World Health Organization et (2013). "Global and regional estimates of violence against women". *WHO Publications*. URL: http://www.who.int/reproductivehealth/publications/violence/9789241564625/en/.