

The Role of Environmental Shocks in Shaping Prosocial Behavior

Jan Duchoslav

Thesis committee

Promotor

Prof. Dr E.H. Bulte
Professor of Development Economics
Wageningen University & Research

Co-promotor

Dr F. Cecchi
Post-doctoral researcher, Development Economics Group
Wageningen University & Research
University of Groningen, the Netherlands

Other members

Dr M. Bauer, CERGE-EI and Charles University, Praha, Czech Republic
Prof. Dr A.S. Bedi, Erasmus University Rotterdam, the Netherlands
Prof. Dr E.H.P. Frankema, Wageningen University & Research
Prof. Dr S.T. Holden, Norwegian University of Life Sciences, Ås, Norway

This research was conducted under the auspices of the Wageningen School of Social Science (WASS)

The Role of Environmental Shocks in Shaping Prosocial Behavior

Jan Duchoslav

Thesis

submitted in fulfilment of the requirements for the degree of doctor at
Wageningen University

by the authority of the Rector Magnificus

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Friday 23 June 2017

at 1:30 p.m. in the Aula.

Jan Duchoslav

The Role of Environmental Shocks in Shaping Prosocial Behavior

190 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2017)

With references, with summary in English

ISBN 978-94-6343-147-7

DOI: 10.18174/410541

In the memory of my grandparents

Contents

	Page
Contents	vii
Chapter 1 Introduction	1
Chapter 2 Prenatal Temperature Shocks	13
Chapter 3 Prenatal Stress	45
Chapter 4 Institutional Change	83
Chapter 5 Towards Prosocial Policy	115
Chapter 6 Discussion and Conclusions	145
References	157
Summary	185
Acknowledgements	189

Chapter 1

Introduction

1.1 Overview

Prosocial behavior – a range of individual actions that not only take into account individual benefits, but also those of the group – forms one of the building blocks of human society, and having helped a small East African ape become a globally dominant species, it underlies our evolutionary success (Nowak and Highfield, 2011, p.17). Humanity should not, however, rest on its laurels. In the words of the renowned theoretical physicist and cosmologist Stephen Hawking (2016):

...now, more than at any time in our history, our species needs to work together. We face awesome environmental challenges: climate change, food production, overpopulation, the decimation of other species, epidemic disease, acidification of the oceans.

Together, they are a reminder that we are at the most dangerous moment in the development of humanity. We now have the technology to destroy the planet on which we live, but have not yet developed the ability to escape it. Perhaps in a few hundred years, we will have established human colonies amid the stars, but right now we only have one planet, and we need to work together to protect it.

Unfortunately, some of the very environmental challenges which necessitate our cooperation undermine it at the same time. Taking a broad look at changes in both the physical and social environment, this thesis explores the ways in which environmental shocks affect human willingness to cooperate. Based mostly on field-experimental evidence from Uganda, it first identifies the negative effects that natural and man-made environmental shocks faced by unborn children in the womb have on their later-life cooperativeness. It then explores how an abrupt change in the social environment – the introduction of a formal health insurance – crowds out cooperation in informal sharing arrangements. It ends on a more positive note, outlining possible policy tools to foster prosocial behavior.

Each of the main chapters can be read as a self-standing research article. However, they have one common message: There are many – sometimes unexpected – ways in which we can lose our preference for cooperation, but there are also ways in

which prosocial behavior can be successfully nurtured.

1.2 Origins of prosocial behavior

The willingness to make personal sacrifices for the benefit of the group forms one of the cornerstones of human society (Bowles and Gintis, 2011). However, prosocial behavior is by far not unique to *homo sapiens*. Well documented examples of complex cooperation in other species (with arguably lower cognitive abilities than humans) range from hunting wolf packs (Macdonald et al., 2004), through certain bird species helping strangers build their nests (Skutch, 1935), to ants building living rafts to survive floods (Cassill et al., 2015), and bees sacrificing their lives to protect the hive (Shorter and Rueppell, 2012). Many have seen this as an indication that prosociality is a natural preference which evolved in numerous species – including humans – and with which we are born (Rushton, 1991; Wesson, 1991; Nowak, 2006).

For centuries, the notion of such innate altruism would seem perfectly natural to religiously inspired thinkers such as Thomas Aquinas, who considered virtue to be of divine origin (Aquinas, 1274, I, q.100, a.1, co.1). However, early modern philosophers started finding inconsistencies in this ethical paradigm, moving so far from the idea of innate goodness that at the brink of the industrial revolution, Adam Smith (1776, book 1, chapter 2) could famously write:

It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages.

Smith was not the first to entertain the thought that selfish motivations can lead to cooperative behavior, but his formulation of this idea has survived centuries, and inspired an entire academic discipline.

As much as self-interest can and does foster cooperation in many circumstances (as aptly illustrated by Smith), it prevents it in many others, especially in

“social dilemma” situations where individual and social welfare maxima do not overlap, such as in various versions of the famous Prisoner’s Dilemma (Flood, 1958). Viewed through the lens of Darwinian theory of evolution, this creates an apparent paradox: If non-cooperation is the dominant strategy in so many situations, the process of natural selection should prevent cooperative behavior from developing on a large scale, which is in direct contradiction to the empirical evidence mentioned at the beginning of this section. An important step towards reconciling this seeming contradiction was made by Trivers (1971), who proposed a formal model to account for natural selection of reciprocally altruistic behavior, and Axelrod and Hamilton (1981), who extended it to the context of the Prisoner’s Dilemma, showing that cooperation can establish itself as an evolutionary stable strategy in a population of self-interested individuals. Finally, building on extensive empirical evidence, Gintis et al. (2003) show that a small number of prosocial individuals can help establish cooperative norms even in a society with an initially self-interested majority.

Cooperation has thus become ubiquitous in human societies, and virtually every society has developed strong mechanisms which discourage its members from deviating from the prosocial norm, and which enforce the return to the norm if they do stray too far from it (Henrich et al., 2001, 2006). These enforcement mechanisms typically rely on the willingness of individuals to incur the cost of punishing antisocial behavior towards themselves and – crucially – towards others (Fehr and Gächter, 2002). Such “altruistic punishment” is in itself another form of prosocial behavior (Yamagishi, 1986) which has over time evolved as a dominant strategy (Ostrom et al., 1992; Fehr and Gintis, 2007; Hauert et al., 2007; Buckholtz and Marois, 2012). Through their culture of prosocial behavior, societies thus further nurture their individual members’ prosociality. In other words, prosocial preferences are endogenous with respect to an individual’s environment (Bowles, 1998).

In their attempts to explain prosocial behavior, different academic disciplines have historically tended to emphasize either the importance of individual characteristics and preferences or that of culture, social norms, and institutions. Psychologists

and economists typically pay more attention to the former, while sociologists and anthropologists gravitate to the latter, often falling on the opposite sides of the larger nature versus nurture argument. Nevertheless, there is now a growing consensus that rather than operating independently, nature and nurture influence each other, and it is their interplay that largely determines individual prosocial preferences (Rutter, 1997; Goldschmidt, 2006).

Insofar as cooperation is essential to much of economic activity, this has some important implications for economic development. On the one hand, it is affected by preferences that have developed over – and are transmitted across – many generations (Spolaore and Wacziarg, 2013). On the other, economic development can also be hampered or buoyed up by changes in the environment within which individuals operate and to which their preferences are – contrarily to accepted economic wisdom (see for example Stigler and Becker, 1977) – endogenous (Fehr and Hoff, 2011). The overarching aim of this thesis is to shed more light on how environmental changes affect prosocial behavior.

1.3 Exogenous shocks and cooperation

A recent but growing body of research shows that environmental shocks affect individual preferences even before one's birth. Various markers of conditions in utero such as weight and length at birth or the ratio between the lengths of the index and ring fingers (second-to-fourth digit ratio, or 2D:4D), have been linked to later-life outcomes, including prosocial preferences.

As a direct consequence of intrauterine growth retardation, preterm birth, or both, low birth weight is a telltale sign of adverse conditions in utero. The exact nature of the physiological processes that lead to low weight at birth (often collectively referred to as intrauterine programming) are still subject to vigorous scientific debate, but there is growing consensus that they may involve decreased fetal nutritional intake (whether due directly to low maternal nutritional intake or to suboptimal placental size, blood flow or function), low fetal oxygen supply, and hormonal imbalances. These can in turn be triggered by conditions as diverse

as maternal malnutrition, stress, disease, substance abuse, and environmental exposure (such as to high altitude or ambient temperature) (Fowden et al., 2006). As a general marker of unfavorable intrauterine conditions, low birth weight has been linked to various later-life outcomes ranging from cardiovascular disease (Barker, 1990) to lower income (Black et al., 2007) and reduced cognitive abilities (Hack et al., 2005), which in turn reduce the taste for cooperation (Moore et al., 1998; Zhang et al., 2015).

Considering the abundant evidence that ambient temperature during gestation is one of the factors affecting birth weight (Wells and Cole, 2002; Lawlor et al., 2005; Deschênes et al., 2009), its effects on later-life outcomes in general and social preferences in particular have received surprisingly little attention. By investigating the effects of prenatal temperature shocks on cooperative behavior in public goods games, I fill in part of this gap in Chapter 2, where I use historical climate data combined with data on child and adult behavior in public goods games to show that abnormally high ambient temperatures during gestation decrease individual contributions to the public good.

The 2D:4D ratio is determined in early stages of fetal development by epigenetic modifications, i.e. changes to gene expressions which take place without a change in the DNA sequence (Jirtle and Skinner, 2007). It reflects relative levels of androgen and estrogen hormones in the womb during early gestation, and remains constant throughout the rest of an individual's life (Lutchmaya et al., 2004; Zheng and Cohn, 2011). Hormonal imbalance is one of the physiological reactions to trauma (de Kloet et al., 2008; Song et al., 2008; Steudte et al., 2011), and is transmitted from mother to fetus during pregnancy (Mancuso et al., 2004; Weinstock, 2008). The 2D:4D ratio is thus a useful phenotypic marker of maternal distress during early gestation.

Variations in hormone levels have been found to influence a wide range of behaviors (Mehta and Josephs, 2011). Increased levels of testosterone – an androgen hormone – cause an increase in fair behavior in ultimatum bargaining (Eisenegger et al., 2010), while increased levels of oxytocin – a hormone whose secretion is regulated by estrogen levels – cause an increase in trust (Kosfeld et al., 2005) and in parochial

altruism (De Dreu et al., 2010). The 2D:4D ratio itself predicts prosocial behavior in several lab experiments (Buser, 2012). Building on these findings, I investigate the effects of war-induced prenatal stress on cooperation. In Chapter 3, I show that fetal hormone exposure manifested in the 2D:4D ratio explains why children whose caregivers suffer from post-traumatic stress disorder are more likely to free-ride in a public goods game than children of non-traumatized caregivers.

It is, however, not only the physical aspect of people's environment that shapes their preferences; its social aspect is equally important in this respect. By restricting individual behavior within defined limits, institutions – “the humanly devised constraints that shape human interaction” (North, 1990, p. 3) – effectively mold people's preferences, including the prosocial ones (Bowles et al., 2003). In social dilemma situations where individual interests and social welfare are at odds, institutions that reward cooperators, punish free-riders, or do both, are vital to sustaining high levels of cooperation (Kosfeld et al., 2009; Sutter et al., 2010). This applies to both formal and informal institutions, where the former are typically thought of as exogenous constraints enforced by an outside party, and the latter as endogenously determined self-enforcing rules which result from the sub-game perfect equilibrium of a repeated game (Greif, 1993; Aoki, 2001a,b).

As formal institutions expand into areas of human interaction hitherto governed by informal arrangements, the former may substitute or complement the latter, depending on the context (see e.g. Roland, 2004; Williamson, 2009; Grzymala-Busse, 2010; Alesina and Giuliano, 2015; Chandrasekhar et al., 2016). Institutional innovations in the formal sphere may therefore crowd out, or solidify, existing informal arrangements, with far reaching consequences for economic development (Pejovich, 1999). Chapter 4 illustrates the effects of such interaction between formal and informal institutions on the cooperation fostered by the latter by showing how the introduction of formal community-level health insurance reduces cooperative behavior in other (informal) settings.

1.4 Towards prosocial policy

Chapters 2, 3 and 4 (introduced in section 1.3) outline several ways in which changes in people's physical and social environment hamper prosocial behavior. While reversing these changes – or preventing them altogether – would likely have positive effects on prosociality, few policymakers are in a position to reverse climate change or prevent violent conflict, no matter how desirable such achievements would be in light of all their other potential benefits. The complexity of these problems puts the policies that are required to resolve them squarely out of the scope of this thesis. Reverting to informal arrangements after formal institutions have been introduced would possibly be a less complex task, but one of dubious overall merit considering that formal institutions are scarcely designed to undermine cooperation, and that they may be fundamental to economic development in their intended purpose (Rodrik et al., 2004).

The environmental shocks discussed so far affect prosocial preferences, which give people an intrinsic motivation to cooperate. Prosocial preferences are fundamental to prosocial behavior, but are not the only source of motivation for individuals to behave prosocially. Ariely et al. (2009) identify two other categories of motives: extrinsic motivation – material rewards associated with prosocial behavior – and image motivation – the utility an individual derives from being perceived by others as being prosocial. Similarly to formal and informal institutions, intrinsic, extrinsic, and image-related motives have the potential of complementing each other, but also of crowding each other out. Extrinsic motivation has the standard direct price effect on prosocial behavior, making it more or less attractive (depending on whether the incentives take the form of rewards or punishment), but it also affects prosocial behavior indirectly through psychological channels. Extrinsic incentives can put a value on the social good provided by the incentivized individual. If this value is lower than the individual previously assumed, he may react by decreasing the effort put into providing the good (Gneezy et al., 2011). Moreover, extrinsic incentives can create doubts about the intrinsic motives for behaving prosocially, decreasing the probability of being perceived as intrinsically prosocial, thus crowding out the image motives for prosocial behavior (Bénabou

and Tirole, 2006). The image value of prosocial behavior is especially likely to suffer when both the behavior in question and the extrinsic rewards are visible, and less so when either or both cannot be readily observed (Ariely et al., 2009). There is nonetheless abundant empirical evidence that incentivizing prosocial behavior can backfire, and does so in many – although not all – cases (Titmuss, 1970; Ariely et al., 2009; Lacetera and Macis, 2010a,b; Gneezy et al., 2011).

Yet, thanks to the relative ease of its provision, extrinsic motivation is a popular policy tool to foster prosocial behavior, and in light of the difficulties associated with addressing the root causes of low intrinsic motivation, it is not one to be readily dismissed. To illustrate the potential of extrinsic incentives as a policy tool for promoting prosocial behavior, I take a look at a specific example of its implementation. In Chapter 5, I evaluate the impact of performance-based financing in small healthcare facilities on their output. I show that financial rewards for increased staff effort can lead to a sustained increase in the provision of healthcare, demonstrating how simple policy measures which exploit extrinsic incentives can alleviate the symptoms of low intrinsic motivation.

1.5 Objectives

Economic production is largely based on cooperation, and the process of economic development involves many social dilemmas. It is therefore crucial to understand how the preferences which guide our behavior vis-à-vis these situations are shaped. The ability and willingness to work for the benefit of the group rather than just one's own has evolved over many generations, and is – to some extent – innate to any healthy human being (Rushton, 1991; Spolaore and Wacziarg, 2013). At the same time, individual prosocial preferences are – also to a certain extent – endogenous to the physical and social environment within which we operate (Bowles et al., 2003; Fehr and Hoff, 2011). The overarching objective of this thesis is to identify ways in which environmental changes affect intrinsic prosocial preferences, and to outline a possible direction for fixing any such negative effects.

1.6 Methodology

To advance knowledge, economics relies on iterations between formulating theoretical propositions about the functioning of the world, and their empirical testing. Theory provides empiricists with testable models and hypotheses, while empirical research identifies unexpected behavioral patterns which can inform more accurate theories (Becker, 1993). In this sense, economics is not much different from other scientific disciplines (though its classification as a science is sometimes disputed even by economists themselves – see e.g. Eichner, 1983). Where economics does, however, differ from natural and life sciences is its reliance on statistical analysis of observational data (typically gathered using questionnaires in the case of microeconomics) rather than on experimental methods to test hypotheses. Economists’ reliance on statistics is traditionally so heavy that they have even coined a special term for using statistical inference to analyze economic phenomena – econometrics. Nevertheless, empirical economists have for the past few decades been borrowing the experimental method, designing incentive-compatible experiments that randomly assign subjects to treatment and control conditions. Inspired by natural sciences, these have been conducted in the laboratory (Smith, 1976, 1982, 2010; Wilde, 1981; Levitt and List, 2007), but also in the field (Harrison and List, 2004; Levitt and List, 2009; Gerber and Green, 2012).

1.6.1 Field experiments

Field experiments span a wide range of types in terms of the level of their contextualization (List, 2006). Closest to the laboratory setting, lab-in-the-field experiments (or artefactual field experiments) simply transplant the laboratory to the field, facilitating the recruitment of subjects from the population of interest, but providing them with no further context (Binswanger, 1980; Bouma et al., 2008; Voors et al., 2011). Framed field experiments contextualize the experimental setting by directly comparing it to real life situations (Reichhuber et al., 2009; Alpizar et al., 2011; Fochmann and Weimann, 2013). In randomized controlled

trials (or natural field experiments), researchers design the treatment and control conditions, and randomly assign subjects to them, observing the subjects' natural behavior, doing away with the artificiality of choices in a laboratory environment (Duflo and Kremer, 2005; Björkman and Svensson, 2009; Duflo et al., 2011; Peabody et al., 2013). On the far end of the contextualization spectrum, natural experiments take advantage of naturally occurring random variation in a condition of interest to provide the treatment and control assignment (van Os and Selten, 1998; Almond, 2006; Banerjee et al., 2010; Almond and Mazumder, 2011).

In line with economic tradition, the research on which this thesis is based makes frequent use of econometric methods, ranging from basic regression analysis to the generalized method of moments. However, it also exploits the advantages of experimental methods to both elicit preferences which would be difficult to observe in non-incentivized settings, and to provide (quasi) natural variation in environmental conditions. Preference for cooperation – one of the main themes of this thesis – is elicited using lab-in-the-field experiments in Chapters 2 and 3, and framed field experiments in Chapters 2 and 4. Thus elicited preferences are then treated as an outcome of random variation in natural (Chapter 2) and man-made (Chapter 3) environmental shocks, or of quasi-random assignment to treatment (Chapter 4). The research described in Chapter 5 mostly uses standard econometric methods.

1.6.2 Impact evaluation and academic research

Funded for the most part by the The Netherlands Foundation for Scientific Research (NWO), the research underlying this thesis was conducted in various parts of Uganda as part of a broader world-wide effort to evaluate development projects (co-)financed by the Dutch Ministry of Foreign Affairs. Impact evaluations in general aim to provide the foundations for successful development initiatives by showing what interventions work (or not), as well as how and under which conditions they work. The rationale behind conducting this particular impact evaluation, which painted a mixed picture of the efficacy of development aid, was to do just that, so as to inform future funding policies to maximize their

impact. The findings presented in this thesis thus go far beyond the original intent of the underlying studies at minimal additional cost. Besides its scientific contribution, this work can serve as an example of the synergies which lie at the intersection of practical development work and academic research, and which often remain unexploited.

1.7 Outline

The rest of this book is organized as follows. Chapters 2, 3, and 4 investigate the effects of environmental shocks on people's willingness to behave prosocially as elicited in incentivized public goods games. Chapter 2 shows that later-life cooperation in public goods games is negatively affected by ambient temperature shocks faced before an individual's birth. Chapter 3 looks at a similar effect of maternal stress, transmitted to the fetus by epigenetic processes. Moving away from changes in physical environment, Chapter 4 instead focuses on a change in the social environment, showing how the introduction of formal health insurance erodes cooperation in informal sharing arrangements. By evaluating the impact of performance-based financing on staff effort in healthcare facilities, Chapter 5 outlines a possible way to remedy low levels of prosocial behavior. Chapter 6 concludes.

Chapter 2

Prenatal Temperature Shocks Reduce Cooperation

Evidence from Public Goods Games in Uganda

Abstract: Climate change has not only led to a sustained rise in mean global temperature over the past decades, but also increased the frequency of extreme weather events. This chapter explores the effect of temperature shocks in utero on later-life taste for cooperation. Using historical climate data combined with data on child and adult behavior in public goods games, we show that abnormally high ambient temperatures during gestation decrease individual contributions to the public good in a statistically and economically significant way. A 1 standard deviation rise in mean ambient temperature during gestation lowers children's cooperation rate in a dichotomous public goods game by 11% points, and the reduced taste for cooperation lasts into adulthood.

This chapter is based on:

Duchoslav, J. (2017). Prenatal Temperature Shocks Reduce Cooperation: Evidence from Public Goods Games in Uganda. Working paper.

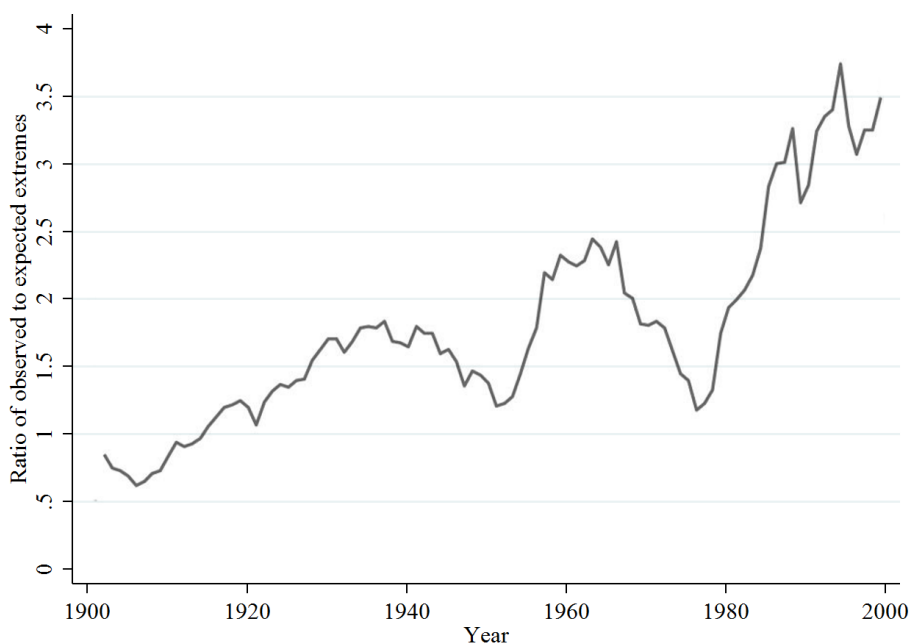
2.1 Introduction

Climate scientists have reached solid consensus that global climate change is occurring over a decade ago (Oreskes, 2005). There has been a sustained rise of mean global temperature, and extreme temperatures have become increasingly common (see Figure 2.1, adapted from Coumou and Rahmstorf (2012)). The focus of scientific discourse on the topic has therefore shifted towards estimating the economic implications of future climate change as well as finding feasible, effective countermeasures and mitigation strategies (Dell et al., 2009). The severity of the former justifies the costs of the latter. Careful assessment of the damage function is thus of utmost importance.

Recent contributions to this literature have investigated the effects of immediate temperature on outcomes ranging from economic production (Dell et al., 2009; Burke et al., 2015) through the onset of conflict (Hsiang et al., 2013), to human reproductive behavior—with consequences for physical health and educational outcomes of the offspring (Wilde et al., 2017). Inspired by another growing body of literature – that on fetal origins, i.e. the impact of events during gestation on later-life outcomes – I take a step back and consider behavioral implications of temperature shocks in utero. Using historical variation in ambient temperature as a natural experiment, and behavior in an incentivized public goods game as an outcome measure, I assess the impact of unusually high temperatures during gestation on later-life taste for cooperation – a preference essential to much economic production. I find that abnormally high ambient temperatures during gestation significantly reduce cooperativeness in children, and that this effect lasts into adulthood.

I first provide a brief overview of the fetal origins literature in Section 2.2. In Section 2.3, I describe the experimental design and my empirical strategy. The results are presented in Section 2.4, and Section 2.5 concludes.

Figure 2.1 – Extreme temperatures



2.2 Fetal origins

Stemming from an observation by the epidemiologist David Barker (1990) that low birth-weight and premature birth are associated with coronary heart disease in later life, the fetal origins literature has grown considerably beyond the medical field into other domains including economics, psychology or management science. Conditions in utero and their proxies have now been linked to later life outcomes ranging from educational achievement (Bhutta et al., 2002; Almond, 2006) through trading ability (Coates et al., 2009) to sexual identity (Csathó et al., 2003). Using the 1918 influenza pandemic as a natural experiment, Almond (2006) finds that mother's illness during pregnancy reduces the educational attainment and income of the offspring. Other natural experiments make use of the Ramadan (Almond

and Mazumder, 2011) and the Nazi invasion of the Netherlands (van Os and Selten, 1998) to show that fasting and stress (respectively) during pregnancy increase the chance of mental disability in the offspring. In a similar fashion, Banerjee et al. (2010) use the case of the advancing phylloxera infestation of French vineyards to show that negative income shocks during gestation reduce adult height—a marker of overall health (Banerjee et al., 2010).

In another strand of the fetal origins literature, various markers of conditions in utero such as preterm birth, weight and length at birth or the ratio between the lengths of the index and ring fingers (2D:4D) are linked to later-life outcomes. Preterm birth predicts poor educational attainment (Bhutta et al., 2002), while low birth weight is associated with reduced cognitive abilities (Hack et al., 2005), which in turn reduce the taste for cooperation (Moore et al., 1998; Zhang et al., 2015). Low weight at birth also predicts behavioral problems (Hille et al., 2007), and low income (Black et al., 2007). The 2D:4D ratio – a marker of prenatal stress – predicts risk and prosocial preferences (Sapienza et al., 2009; Buser, 2012, and Chapter 3 of this thesis). Considering that ambient temperature during gestation is known to affect preterm birth (Lajinian et al., 1997; Yackerson et al., 2008; Flouris et al., 2009) and birth weight (Wells and Cole, 2002; Lawlor et al., 2005; Deschênes et al., 2009), its effects on later-life outcomes have received surprisingly little attention. To be sure, much of this non-experimental strand of fetal origins literature consists of comparative cohort studies without sufficient controls for socioeconomic and behavioral confounders (Zhang et al. (2015), Black et al. (2007), Dell et al. (2009), and Deschênes et al. (2009) being noteworthy exceptions), and is therefore prone to suffer from omitted variable bias. Moreover, there is no consensus regarding the physiological mechanisms linking ambient temperature during gestation to birth outcomes.¹ Taken as a whole, this body of literature nonetheless points toward a link between ambient temperature during gestation

¹Reduced fetal nutrient intake caused directly by altered maternal nutrient intake or indirectly by impaired placental function is often mentioned as a potential mechanism through which temperature shocks during pregnancy affect birth outcomes, as are hormonal imbalances caused by heat shocks. There is some evidence that these mechanisms are at play in other mammals such as sheep (Hansen, 2009; Bloomfield et al., 2003), but studies of their role in human gestation are rare (Yajnik, 2002; Smith and Nicholson, 2007).

and later-life outcomes. The methodologically well-executed study by Deschênes et al. (2009) (as well as that by Lawlor et al. (2005), which may suffer from an omitted variable bias) further suggests that it is relative – rather than absolute – temperature shocks that matter in this respect.²

To my knowledge, the hitherto only study to look at the effect of in utero temperature shocks on later-life outcomes links temperature during gestation to depression in adulthood (Adhvaryu et al., 2015). The present chapter fills in part of the remaining gap by studying the effects of ambient temperature during gestation on the taste of cooperation.

2.3 Experimental design and data

Employing new data from behavioral games, anthropometric measurements and an extensive socioeconomic survey conducted in Uganda, I exploit the quasi-experimental variation in weather to gauge the impact of prenatal temperature shocks on later-life cooperation.

I use several distinct datasets in my analysis.³ Temperature data come from Willmott and Matsuura’s (2015) gridded monthly time series interpolated from weather station observations. I combine the temperature values with my main and secondary self-collected datasets. The main set contains data from a survey of primary school pupils from Northern Uganda, and also includes their choices in a one-shot dichotomous public goods game, as well as their anthropometric measurements. The secondary set contains data from a household survey from Southern Uganda, and records of the behavior of the representatives of these households in a standard public goods game.

²Wells and Cole (2002) come to a different conclusion, providing between-population evidence that absolute – rather than relative – temperature shocks during gestation drive changes in birth-outcomes. Although they control for various confounding factors such as income and nutritional intake, these are aggregated at the national and yearly levels. Due to this design, the study cannot distinguish between regular seasonal variation in temperature and relative temperature shocks.

³See Appendix Appendix 2.B: for an overview of variable definitions.

2.3.1 Temperature

I construct my measure of ambient air temperature during gestation using Willmott and Matsuura's (2015) historical time series – one of two publicly available datasets with values spatially intrapolated from terrestrial weather station measurements. Although the alternative dataset produced by Harris et al. (2014) is generally more popular, that of Willmott and Matsuura is better suited for my purposes as it uses a much denser set of weather stations in East Africa (as well as globally).⁴ The dataset contains a single temperature value for each historical month and spatial grid cell of $0.5^\circ \times 0.5^\circ$ (roughly $55\text{km} \times 55\text{km}$ in Uganda).

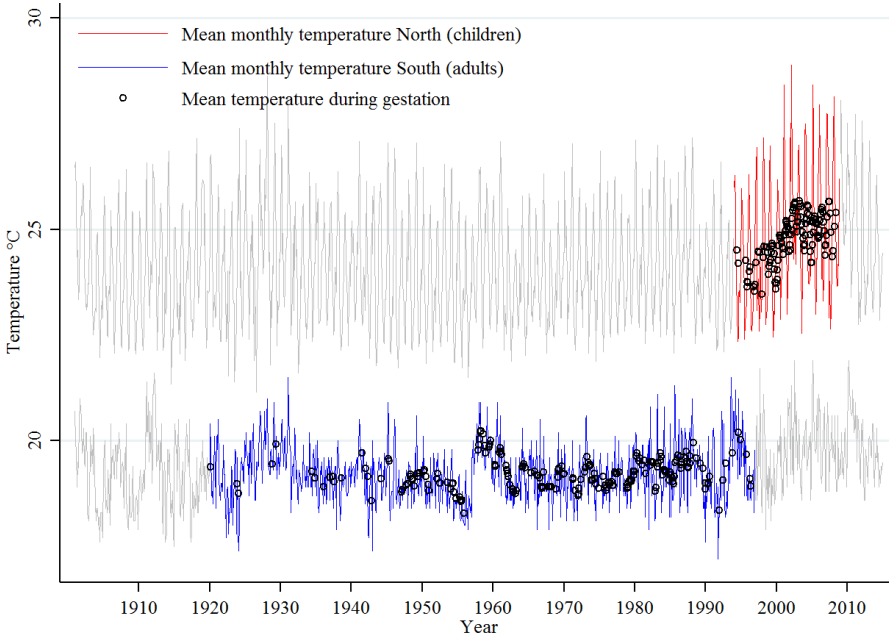
My behavioral and survey data come from two clusters of locations—one in Northern Uganda, spanning four neighboring grid cells (2.5°N – 3.5°N , 32.5°E – 33.5°E), and one from a single grid cell in Southern Uganda (0.5°S – 1.0°S , 30.0°E – 30.5°E). Since the differences between the values in the four neighboring cells in any given month are minimal, I use their mean values for all observations

⁴The dataset produced by Harris et al. (2014) at the Climate Research Unit at the University of East Anglia is clearly more popular than that assembled by Willmott and Matsuura (2015) at the University of Delaware. While Willmott and Matsuura's dataset has been cited by 149 studies since its publication in 2009, the CRU dataset has been cited 751 times since its publication five years later according to Google Scholar. In 2015 alone, the CRU dataset was cited in 226 publications on Africa, compared to only 20 citations of Willmott and Matsuura (including citations of previous versions of their dataset). I find the preference for Harris et al. odd, given that their data is based on a much sparser set of weather stations than Willmott and Matsuura's. Willmott and Matsuura use measurements from between 3 and 19 (on average 8) weather stations within a 5° (about 555 km) radius of my main research site in Northern Uganda, and between 3 and 19 (with an average of 6) stations within a 5° radius from my secondary research site in Southern Uganda, depending on the month and year of measurement. Harris et al. use only 0 to 4 (on average 3) and 0 to 2 (on average 1) stations respectively in the same regions, and until 1941, the nearest weather station used in their dataset was Harare— 21° (2300 km) from my main research site and 17° (1900 km) from my secondary research site. Willmott and Matsuura's data predict over 25% of the variation in actual temperature data from Entebbe, Uganda, while Harris et al.'s predict less than 7%. I therefore consider Willmott and Matsuura (2015) superior to Harris et al. (2014) in the East African context, despite the overwhelming popularity of the latter.

in the northern cluster, obtaining a single monthly temperature value for each of the two location clusters. To obtain the value of ambient temperature during the gestation of a respondent in the northern cluster, I average the temperature in the northern location during the month of his or her birth and in the preceding eight months. The values for respondents from the southern locations are constructed analogously. For an individual born in November, for example, I average the monthly values from March until November.

The historical monthly means are plotted in Figure 2.2, where the gray curves represent the monthly temperature means in the two clusters, the red curve represents the temperatures in the months in which the respondents in the northern cluster (main sample) were gestating, and the blue curve denotes the temperatures in the months in which the respondents in the southern cluster (secondary sample) were gestating. The mean values of ambient air temperature during the gestation (9 months) of individual respondents are denoted by black circles.

Mean temperatures of 9-month-long gestational periods have, by construction, a much smaller variance than monthly mean temperatures (as Figure 2.2 illustrates). Similarly, the variance of monthly means is smaller than that of daily means. Basing my analysis on overall mean values therefore somewhat reduces its sensitivity. Using more detailed temperature data such as a set of 9 monthly values for each individual would not, however, correspond to the level of precision with which I can determine the dates of conception – and thus the periods of gestation – of the respondents. By its nature, I can only infer an individual’s probable date of conception from their date of birth. The possibility of premature and late births introduces in such inference a level of uncertainty which is only aggravated by the fact that I only know the month (rather than the exact date) of birth of my respondents. The margin of error associated with these imprecisions can easily be more than a month. In extreme cases, there would thus be no overlap between actual and assumed values of temperatures in any given month of gestation. Using instead the mean value over the whole assumed period of gestation largely reduces the effect of such inaccuracies.

Figure 2.2 – Temperature trends

2.3.2 Main sample

My main sample consists of 531 children and their caregivers from Pader district in Northern Uganda. The children come from 42 primary schools visited in June and July 2014. In each school, 16 pupils were randomly selected from a list of those enrolled at the beginning of the year.⁵

I measure children's and caregivers' willingness to cooperate by involving them in a one-shot dichotomous public goods game similar to those in Cárdenas et al. (2009) and Barr et al. (2014). In each school, children were randomly assigned to groups of 8, but were not told which other 7 children (of the 15 participating in that school) belonged to their group. Each child then anonymously selected either

⁵Out of a total of 672, the caregivers of 141 pupils did not know their children's birth date. These children were excluded from my analysis.

a “private card” or a “group card”.⁶ By choosing the private card, the respondent allotted 4 candies to himself, but none to the other unknown members of the group. By selecting the group card, the respondent instead ensured 1 candy for each of the 8 group members, including himself (see Appendix 2.C: for a reproduction of the two cards). In this set up, total welfare is maximized when all 8 game participants opt for the group card, such that they each receive 8 candies. A sole free rider selecting the private card would receive 11 candies, but in the Nash equilibrium, everyone selects the private card and ends up with only 4 candies each.

Caregivers played a similar public goods game, but made their decisions in the isolated environment of their home, unaware of the identity of the other 8 participants with whom they were grouped. If they chose the private card, they received 4,000 UGX (roughly 1.5 USD at the time). Choosing the group card instead meant an allocation of 1,000 UGX to each anonymous member of the group, including themselves. In the Nash equilibrium, each participant thus received 4,000 UGX, total welfare was maximized at a return of 8,000 UGX for each group member, and a sole free rider would earn 11,000 UGX. 26% of the children chose the cooperative option, while the cooperation rate among their caregivers was 34%.

The descriptive statistics for the children are presented in Table 2.1. The mean ambient temperature faced by the mothers of the children in my sample during their pregnancy was 24.9°C. The children are on average 10 years old, and girls and boys are equally represented. The height of the children in the sample is practically identical to the mean for their age, but their body mass is 1.39 standard deviation below the mean for their age (de Onis et al., 2007).⁷ This suggests that some may have been nutritionally deprived in their early life, which could confound

⁶Contrarily to many public goods games in which participants can choose their preferred contribution level, I opted for a dichotomous choice, effectively reducing the game to a prisoner’s dilemma: respondents could either cooperate or not. While this reduced my ability to pick up the nuances present in the experimental sample, I believe that it facilitated the decision making process, especially for the youngest.

⁷Based on WHO recommendations for treating outliers (de Onis et al., 2007), I truncate the anthropometric data at 6 standard deviations from the mean. This results in 15 and 14 dropped observations for height-for-age and BMI-for-age respectively.

my results (I address this issue below). Children's cognitive ability was measured through standard Raven's progressive matrices (Kaplan and Saccuzzo, 2012). It is an intelligence quotient (IQ) adjusted for age and scaled relative to the sample (with a mean at 100 and a standard deviation of 15).

I further proxy for child prenatal stress by the second-to-fourth (2D:4D) digit ratio (a marker of hormonal exposure in utero). The lengths of the index and ring fingers were measured on the palmar surface of the right hand, from the midpoint of the palmar digital crease to the tip of the finger. Given the constraints due to the remoteness of the field location,⁸ the measurement is only precise to the nearest 1 mm, resulting in measurement error of $\pm 3.3\%$ at the mean of the estimates.⁹ While the precision of this measurement is still well below that obtained in laboratory settings (see Voracek et al., 2007), my measurements should be at least as accurate as those in other field studies which often only report whether the index finger is longer, shorter, or of the same length as the ring finger (Buser, 2012).

Child postnatal conflict exposure – a potentially important confounding factor considering that most children in the sample were born during a period of civil war in Northern Uganda – is a composite measure derived from the exposure of the caregiver and the child's year of birth. Given their young age at the time of the conflict, children were not asked any war-related questions. Instead, I use caregiver responses to an adapted version of the War Trauma Questionnaire (Macksoud, 1992; Papageorgiou et al., 2000).¹⁰ It consists of 23 yes-no questions about various violent events witnessed by the caregiver, from which I construct a conflict exposure index using the number of positive responses as a measure of exposure (Bellows and Miguel, 2009) and normalizing it for the sample. To proxy the child's postnatal conflict exposure I weight the caregiver's conflict exposure index by the portion of violence their child could have potentially witnessed after

⁸The state of the art in measuring finger lengths is to use an office scanner to take a perfectly flat image of the palmar surface of the hand, and computer software to measure the exact lengths. I instead used clipboards and tape measures.

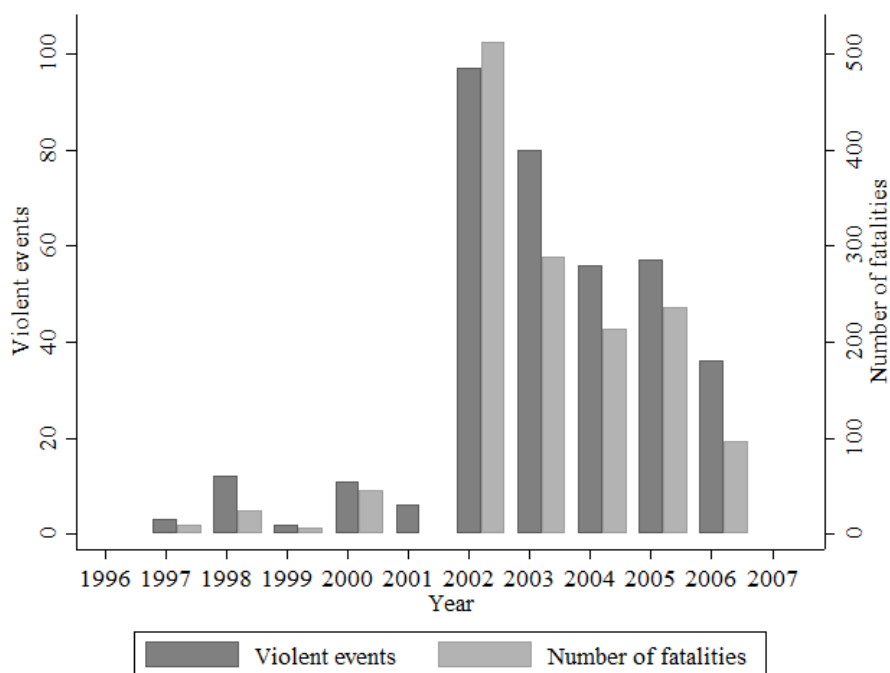
⁹A pilot in which 30 raters each separately measured the digit lengths of 35 individuals revealed comparable margins of error.

¹⁰Any questions about shelling and bombardment are irrelevant in the Ugandan setting, and were therefore omitted from the questionnaire.

birth. To obtain the weights, I divide the number of civilian fatalities that occurred in Pader district following the child's birth by the total number of civilian facilities recorded in the district throughout the length of the conflict (Figure 2.3).¹¹ For example, a child born in December 2003 – by which time 62% of reported fatalities took place – whose caregiver's conflict exposure is 87% is likely to have witnessed 38% of the violence that the caregiver was exposed to. For my purposes, the child's conflict exposure index would therefore be 33% ($(1 - 0.62) \times 0.87 = 0.33$).

Finally, precipitation and consumer prices during gestation are constructed analogously to the temperature variable.

Figure 2.3 – Civilian targeting



To account for further environmental and genetic effects on preferences, I also

¹¹Source: ACLED Version 5, 1997-2013

Table 2.1 – Descriptive statistics (children)

Variable	Mean	Std. Dev.	Min.	Max.	N
Temperature	24.915	0.505	23.464	25.681	531
Cooperation	0.256	0.437	0	1	531
Female	0.493	0.5	0	1	531
Age in years	10.478	2.731	6	20	531
Height-for-age	-0.025	1.005	-3.391	3.571	516
BMI-for-age	-1.387	0.935	-3.99	2.26	517
IQ-for-age	100.619	14.815	73.993	148.458	488
Prenatal stress	0	1	-3.273	5.028	215
Postnatal conflict exposure	0	1	-4.971	2.885	477
Precipitation	5.408	1.233	2.889	8.347	531
Consumer prices	-2.113	2.163	-5.954	4.937	531

interviewed each child’s main caregiver—the adult household member with whom the child spends most time. The descriptive statistics for the caregivers are presented in Table 2.2. About half of the caregivers in my sample chose to cooperate in the public goods game. Caregivers are on average 41 years of age,¹² 58% are female. Additionally, I collected information about their education level and risk preferences. All caregivers were exposed to at least some kind of conflict-related violence, though the level of exposure varies greatly.¹³ Almost all respondents are Christian and belong to the Acholi ethnic group. A typical household is composed of 8 people. I also collected information about their relative asset wealth (Sahn and Stifel, 2003).

In my setting, information about the current main caregivers can only serve as a proxy for environmental and genetic influences to which the children have been subjected throughout their lives. Of the 531 caregivers in my sample, only 265 are biological mothers of the children, while 206 are their biological fathers. The remaining 60 were grandparents, uncles or aunts, other relatives, and siblings (in descending order of prevalence). One caregiver was not related to the child at all. Nonetheless, the average caregiver in my sample had been taking care of the child

¹²Nine caregivers did not know their age, reducing the number of observations to 522.

¹³Three caregivers refused to complete the conflict exposure module of the survey, reducing the number of observations to 528. This reduction carries over the measure of child postnatal conflict exposure, which is derived from that of their caregiver.

Table 2.2 – Descriptive statistics (caregivers)

Variable	Mean	Std. Dev.	Min.	Max.	N
Age in years	40.588	9.738	20	94	522
Age at birth of child	30.207	9.545	10	85	522
Female	0.58	0.494	0	1	531
Acholi	0.976	0.155	0	1	531
Christian	0.998	0.043	0	1	531
Years of education	3.84	3.431	0	18	531
Married	0.772	0.42	0	1	531
Functional literacy	0.234	0.423	0	1	531
Conflict exposure	0	1	-3.035	1.688	528
Household size	7.945	2.491	2	20	531
Wealth	0	1	-1.54	3.068	531
Risk averseion	0.516	0.5	0	1	531
Cooperation	0.335	0.473	0	1	531

for 82% of the child’s life, making the information about the caregivers a strong proxy for the environment surrounding the children.

2.3.3 Secondary sample

My main dataset contains rich information about the children and their environment, but suffers from two important shortcomings. The first is its conflict setting. If temperature shocks invite conflict (O’Loughlin et al., 2012; Hsiang et al., 2013), then the physiological effects of ambient temperature during gestation would be hard to disentangle from the effects of temperature-induced conflict. Second, it does not allow us to repeat the analysis using the caregiver’s behavior and temperature during their gestation, because I only know the caregivers’ year of birth. This means that I cannot tell whether the behavioral effects that temperature shocks in utero have on children last into adulthood. To address these concerns, I turn to a second sample of 257 adults from Sheema district in Southern Uganda, which was untouched by the conflict in the north.

In July and August 2014 I visited 45 villages in the district, and surveyed a random sample of 10 households per village selected from a census. A randomly selected adult representative of each surveyed household was invited to participate in an

incentivized public goods game.¹⁴

The game was played in groups of 5 participants who could anonymously decide to contribute between 0 and 5 tokens (worth 1,000 UGX or 0.38 USD each) to a common pot, keeping the rest for themselves. Shared funds were doubled and redistributed equally (after rounding). After an initial practice round, 5 rounds of the game were played with each group, though the participants did not know beforehand how many rounds the game would last.¹⁵ One round was selected at random for payment.

In this design, total welfare is maximized when all participants contribute their entire endowment of 5 tokens to the common pot, receiving 10 each in return. Nevertheless, free riders could receive up to 13 tokens, and the Nash equilibrium is reached with all players keeping their 5 tokens. On average, participants contributed 3.44 tokens to the public good. The descriptive statistics for the game participants are presented in Table 2.3. The mean ambient temperature faced by the mothers of the adults in my sample during their gestation was 19.2°C. The participants are on average 42 years old, a third are female, and 84% are married. On average, they fell just short of completing primary education, and half are functionally literate. Nearly the whole sample is ethnically Ankole and Christian by religion.

2.4 Analysis and results

2.4.1 Main finding

I hypothesize that exposure to high ambient temperatures during an individual's gestation may impact his or her later-life preference for cooperation. Combining the findings of Wells and Cole (2002), Lawlor et al. (2005) and Deschênes et al.

¹⁴Out of a total of 450 randomly selected household representatives, 193 either did not know their month and year of birth, or did not show up to play the public goods game. These people were excluded from my analysis.

¹⁵Withholding the information about the exact length of the game helps ensure that all rounds, including the final one, are played in a strategic way.

Table 2.3 – Descriptive statistics (adults)

Variable	Mean	Std. Dev.	Min.	Max.	N
Temperature	19.222	0.347	18.278	20.233	257
Contribution	3.437	1.174	0	5	257
Age in years	42.004	15.711	16	92	257
Female	0.335	0.473	0	1	257
Years of education	6.743	3.866	0	17	257
Married	0.844	0.363	0	1	257
Munyankole	0.977	0.151	0	1	257
Christian	0.996	0.062	0	1	257
Functional literacy	0.498	0.501	0	1	257
Wealth	0	1	-1.212	5.515	257

(2009) with those of Hack et al. (2005) and Zhang et al. (2015), I expect prenatal exposure to high ambient temperatures to reduce cooperative behavior. I first analyze this relationship by fitting the following probit model:

$$Pr(Cooperation_{iym} = 1 | Temperature_{ym}) = \Phi(\alpha + \beta Temperature_{ym}) \quad (2.1)$$

where $Cooperation_{iym}$ equals 1 if child i born in month m of year y and attending school s selects the cooperative option, and $Temperature_{ym}$ is the mean ambient temperature during the likely gestation of children born in month m of year y . Standard errors are clustered at the level of running month of birth.

I find that exposure to high ambient temperature during gestation is negatively correlated with the child's probability of contribution to the public good. Parametrically, a 1 °C increase in mean ambient temperature during gestation reduces the child's probability of contribution by 7.5% points (Table 2.4, column 1). At mean prevalence of 25.6%, this is equivalent to a 29% reduction in the likelihood of cooperation.

To account for non-temperature seasonal confounds and unobserved background characteristics potentially related to season of birth similar to those described by Buckles and Hungerman (2013) in the United States, I include calendar month fixed effects, which only increases the magnitude of the detected effect of temperature (Table 2.4, column 2). The relationship could potentially also be driven by other child characteristics. Prosocial preferences develop throughout

childhood and adolescence, and become increasingly gender-dependent with approaching adulthood (Eisenberg et al., 2006). Controlling for age, gender and their interaction, however, does not change the interpretation of the result (Table 2.4, column 3), nor does controlling for caregiver characteristics and school fixed effects to account for family and peer demographics (Table 2.4, columns 4-5), both of which have been linked to children's prosocial behavior (Eisenberg et al., 2006). Gradually including these additional controls, I reach the following fully specified model:

$$\begin{aligned} Pr(Cooperation_{iyms} = 1 | Temperature_{ym}, \mathbf{x}_{iyms}, \mathbf{z}_{iyms}) = \\ \Phi(\alpha + \beta Temperature_{ym} + \boldsymbol{\gamma}' \mathbf{x}_{iyms} + \boldsymbol{\delta}' \mathbf{z}_{iyms} + \zeta_m + \eta_s) \end{aligned} \quad (2.2)$$

where $Cooperation_{iyms}$ equals 1 if child i born in month m of year y and attending school s selects the cooperative option, $Temperature_{ym}$ is the mean ambient temperature during the likely gestation of children born in month m of year y , \mathbf{x}_{iyms} is a vector of individual child characteristics (female, age, age×female), \mathbf{z}_{iyms} is a vector of caregiver characteristics (female, age, age×female, Acholi, years of education), ζ_m are month-of-birth fixed effects, and η_s are school fixed effects. Standard errors are clustered at the level of running month of birth. In all specifications, I find a strong negative effect of ambient temperature during gestation on children's cooperation.

Table 2.4 – High temperature decreases taste for cooperation

	Child cooperation				
	(1)	(2)	(3)	(4)	(5)
Temperature	-0.075** (0.036)	-0.132** (0.052)	-0.143** (0.072)	-0.141* (0.076)	-0.208*** (0.067)
Month of birth FE	N	Y	Y	Y	Y
Child characteristics	N	N	Y	Y	Y
Caregiver characteristics	N	N	N	Y	Y
School FE	N	N	N	N	Y
N	531	531	531	522	499
Area under ROC curve	0.556	0.610	0.620	0.625	0.742

Notes: Probit marginal effects. SE clustered at the level of running month of birth in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child characteristics: Female, Age (in months), Age×Female. Caregiver characteristics: Female, Age (in years), Age×Female, Acholi, Years of education.

Result 1: Exposure to abnormally high ambient temperature during gestation decreases later-life taste for cooperation. A 1 °C (1 s.d.) increase in mean ambient temperature during gestation decreases the probability of cooperation in a public goods game by up to 21% points (11% points), leading to a 17% (8%) drop in total welfare.

The result is robust to excluding high-leverage and high-influence observations (see Table 2.A.1 in the Appendix), as well as to limiting the analysis to children born in the same area (see Table 2.A.2). Including the mean air temperature during a 9-month period 1 year prior to the assumed period of gestation as a placebo treatment leaves the result unaffected (see Table 2.A.3), as does assuming other periods of gestation (see Table 2.A.4).¹⁶

2.4.2 Indirect temperature effects and other factors

Both in theory and in my data, temperature is strongly negatively correlated with precipitation, which in turn affects agricultural yields and – by extension – food prices. The combination of high temperatures and low precipitation during gestation could thus lead to malnutrition in infancy, whose negative consequences for the child’s cognitive abilities can persist for years (Beckett et al., 2006; Figlio et al., 2014). On the other hand, low precipitation levels decrease the likelihood of malaria contraction (Craig et al., 1999), and could thus also have a positive effect on later-life outcomes (Barreca, 2010). I control for the environmental covariates and the indicators of early-life deprivation separately by estimating the following

¹⁶In my analysis, I assume the period of gestation to correspond with the calendar month of the respondent’s birth and the previous 8 months. To obtain the mean temperature during the gestational period of a respondent born in November, for example, I average the mean temperatures from March until November. As discussed in Section 2.3.1, this is quite a simplification. A child born full term on the 1st of November would have gestated between February and October, while a child born 1 month prematurely on the 30th of November would have gestated between April and November. As a robustness check, I re-estimate model (2.2) using these two extremes as alternative individual regressors, as well as each others instruments in an instrumental variable probit setting.

probit models:

$$\begin{aligned} Pr(Cooperation_{iym s} = 1 | Temperature_{ym}, \mathbf{x}_{iym s}, \mathbf{z}_{iym s}, \mathbf{e}_{ym}) = \\ \Phi(\alpha + \beta Temperature_{ym} + \boldsymbol{\gamma}' \mathbf{x}_{iym s} + \boldsymbol{\delta}' \mathbf{z}_{iym s} + \zeta_m + \eta_s + \boldsymbol{\theta}' \mathbf{e}_{ym}) \end{aligned} \quad (2.3)$$

$$\begin{aligned} Pr(Cooperation_{iym s} = 1 | Temperature_{ym}, \mathbf{x}_{iym s}, \mathbf{z}_{iym s}, \mathbf{c}_{iym s}) = \\ \Phi(\alpha + \beta Temperature_{ym} + \boldsymbol{\gamma}' \mathbf{x}_{iym s} + \boldsymbol{\delta}' \mathbf{z}_{iym s} + \zeta_m + \eta_s + \boldsymbol{\iota}' \mathbf{c}_{iym s}) \end{aligned} \quad (2.4)$$

where \mathbf{e}_{ym} is a vector of environmental variables *Precipitation* (monthly mean of daily precipitation) and *Consumer prices* (linearly detrended consumer price index) in month m of year y , $\mathbf{c}_{iym s}$ is a vector of personal characteristics *Height-for-age*, *BMI-for-age* and *IQ-for-age* of child i born in month m of year y and living in village s , and all other notation has the same meaning as in (2.2). Combining (2.3) and (2.4), I get the following fully specified model:

$$\begin{aligned} Pr(Cooperation_{iym s} = 1 | Temperature_{ym}, \mathbf{x}_{iym s}, \mathbf{z}_{iym s}, \mathbf{e}_{ym}, \mathbf{c}_{iym s}) = \\ \Phi(\alpha + \beta Temperature_{ym} + \boldsymbol{\gamma}' \mathbf{x}_{iym s} + \boldsymbol{\delta}' \mathbf{z}_{iym s} + \zeta_m + \eta_s + \boldsymbol{\theta}' \mathbf{e}_{ym} + \boldsymbol{\iota}' \mathbf{c}_{iym s}) \end{aligned} \quad (2.5)$$

where all notation is the same as in (2.3) and (2.4).

I find that high precipitation during gestation decreases children's taste for cooperation, while consumer prices do not significantly affect it (Table 2.5, columns 1 and 3). This suggests that – at least in the context of Northern Uganda – the effects of precipitation during gestation on later-life prosocial preferences via exposure to malaria dominate those via agricultural yields. Importantly, however, they do not wash away the effect of temperature itself (Table 2.5, columns 4 and 6).

High cognitive abilities proxied by the age-adjusted IQ predict higher probability of contributing to the public good in accordance with Zhang et al. (2015). From Beckett et al. (2006) and Figlio et al. (2014), I would expect height-for-age and BMI-for-age – both markers of early-life nutritional deprivation – to also be positively correlated with child cooperation. Instead, I estimate their effects to be statistically insignificant and significantly negative respectively (see Table 2.5,

Table 2.5 – Other environmental factors and early life deprivation

	(1)	(2)	(3)	(4)	(5)	(6)
Temperature				-0.212** (0.083)	-0.201*** (0.068)	-0.215*** (0.080)
Precipitation	-0.035** (0.016)		-0.047*** (0.016)	-0.038** (0.017)		-0.049*** (0.017)
Consumer prices	0.015* (0.009)		0.012 (0.009)	-0.002 (0.011)		-0.006 (0.010)
Height-for-age		0.022 (0.020)	0.027 (0.020)		0.018 (0.019)	0.024 (0.019)
BMI-for-age		-0.042** (0.020)	-0.050*** (0.020)		-0.046** (0.020)	-0.049** (0.020)
IQ-for-age		0.003*** (0.001)	0.004*** (0.001)		0.004*** (0.001)	0.004*** (0.001)
Child & caregiver characteristics	Y	Y	Y	Y	Y	Y
Month of birth & School FE	Y	Y	Y	Y	Y	Y
N	499	437	437	499	437	437
Area under ROC curve	0.740	0.763	0.771	0.750	0.773	0.780

Notes: Probit marginal effects. SE clustered at the level of running month of birth in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child characteristics: Female, Age (in months), Age \times Female. Caregiver characteristics: Female, Age (in years), Age \times Female, Acholi, Years of education.

columns 2-3). Their inclusion in the model does not however alter my main result (Table 2.5, columns 5-6).

There is increasingly conclusive evidence that high temperatures may trigger or intensify violent conflict (O’Loughlin et al., 2012; Hsiang et al., 2013). Pre- and post-natal exposure to conflict have in turn been found to influence social preferences: Conflict-induced prenatal stress reduces contributions to the public good in later life (see Chapter 3), while post-natal exposure leads to more prosocial behavior within close networks (Voors et al., 2012; Bauer et al., 2014; Gilligan et al., 2014). Many of the children in my sample were born during a period of civil war in Northern Uganda. Using a sub-sample for which information on war exposure and prenatal stress is available,¹⁷ I control for potential conflict-related confounds by estimating the following fully specified probit model:

$$\begin{aligned} Pr(Cooperation_{iyms} = 1 | Temperature_{ym}, \mathbf{x}_{iyms}, \mathbf{z}_{iyms}, \mathbf{w}_{iyms}) = \\ \Phi(\alpha + \beta Temperature_{ym} + \boldsymbol{\gamma}' \mathbf{x}_{iyms} + \boldsymbol{\delta}' \mathbf{z}_{iyms} + \zeta_m + \eta_s + \boldsymbol{\kappa}' \mathbf{w}_{iyms}) \end{aligned} \quad (2.6)$$

where \mathbf{w}_{iyms} is a vector of conflict exposure measures of child i born in month m of year y and living in village s , comprising *Prenatal stress* (a z-score of the reverse 2D:4D ratio—a marker of prenatal stress) and a measure of *Postnatal conflict exposure*. All other notation is the same as in (2.2).

I find that prenatal stress reduces the taste for cooperation, as does postnatal conflict exposure (Table 2.6, column 1). While the former is in line with the findings in Chapter 3, the latter contrasts with other studies (Voors et al., 2012; Bauer et al., 2014; Gilligan et al., 2014). Nonetheless, the inclusion of these war-related controls does not wash away the effect of ambient temperature during gestation; it rather makes it stronger (Table 2.6, column 2).

¹⁷Part of the children in this study were also interviewed in 2012, at which time I measured the lengths of their fingers to calculate the 2D:4D ratio. I made the same measurements for this study, but after explaining to the research assistants that the digit ratio is “usually around 1”, the frequency of precisely that value being reported increased dramatically. While I do not believe that this was a result of intentional misenumeration, it does constitute a heavy bias, forcing me to discard the 2014 2D:4D values. This reduced the available sample to those children interviewed in 2012.

Table 2.6 – Conflict exposure

	Child cooperation	
	(1)	(2)
Temperature		-0.355*** (0.110)
Prenatal stress	-0.081** (0.035)	-0.083** (0.035)
Postnatal conflict exposure	-0.036* (0.022)	-0.045** (0.021)
Child characteristics	Y	Y
Month of birth & School FE	Y	Y
<i>N</i>	164	164
Area under ROC curve	0.825	0.841

Notes: Probit marginal effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. SE clustered at the level of running month of birth in parentheses. Child characteristics: Female, Age, Age², Age × Female, Age² × Female.

The preferences of children may be influenced by those of their caregivers through both environmental and – when the two are blood related – genetic mechanisms (Dohmen et al., 2012). I control for caregiver preferences in the following fully specified probit model:

$$\Pr(\text{Cooperation}_{iym} = 1 | \text{Temperature}_{ym}, \mathbf{x}_{iym}, \mathbf{z}_{iym}, \mathbf{p}_{iym}) = \Phi(\alpha + \beta \text{Temperature}_{ym} + \gamma' \mathbf{x}_{iym} + \delta' \mathbf{z}_{iym} + \zeta_m + \eta_s + \mathbf{X}' \mathbf{p}_{iym}) \quad (2.7)$$

where \mathbf{p}_{iym} is a vector of the preferences of the caregiver of child i born in month m of year y and living in village s , comprising *Caregiver cooperation* (a dummy equal to 1 if the caregiver chose the cooperative option) and *Caregiver risk aversion* (a dummy equal to one if the caregiver chose the least risky of a series of lotteries). All other notation is the same as in (2.2).

I find that a child's social preferences are strongly correlated with the social preferences of their main caregiver, but not with the caregiver's risk preferences. Children are about 11% points more likely to contribute to the public good if their main caregiver contributes to it as well in a separate game (Table 2.7, column 1). The effect of ambient temperature during gestation is however not affected by these controls (Table 2.7, column 2), and the results hold when analysis is restricted to caregivers who are biological parents of their children (Table 2.7,

columns 3-4).

Table 2.7 – Caregiver preferences

	Child cooperation			
	All caregivers	Biological parents only		
	(1)	(2)	(3)	(4)
Temperature		-0.199*** (0.068)		-0.190*** (0.072)
Caregiver cooperation	0.105*** (0.039)	0.102*** (0.039)	0.113*** (0.042)	0.115*** (0.041)
Caregiver risk aversion	-0.030 (0.035)	-0.023 (0.034)	-0.007 (0.038)	0.001 (0.038)
Child characteristics	Y	Y	Y	Y
Month of birth & School FE	Y	Y	Y	Y
<i>N</i>	499	499	443	443
Area under ROC curve	0.738	0.751	0.724	0.738

Notes: Probit marginal effects. SE clustered at the level of running month of birth in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child characteristics: Female, Age (in months), Age×Female. Caregiver characteristics: Female, Age (in years), Age×Female, Acholi, Years of education.

Finally, it is conceivable that different types of parents are more likely to conceive at times with different weather and climate patterns. If the different types of parents would also have different social preferences, such self-selection could bias my results. In my setting, much of any such bias should be absorbed by the month of birth fixed effects. To further verify that no self-selection bias is present, I regress a battery of caregiver characteristics on mean temperature during the child's gestation according to the following model:

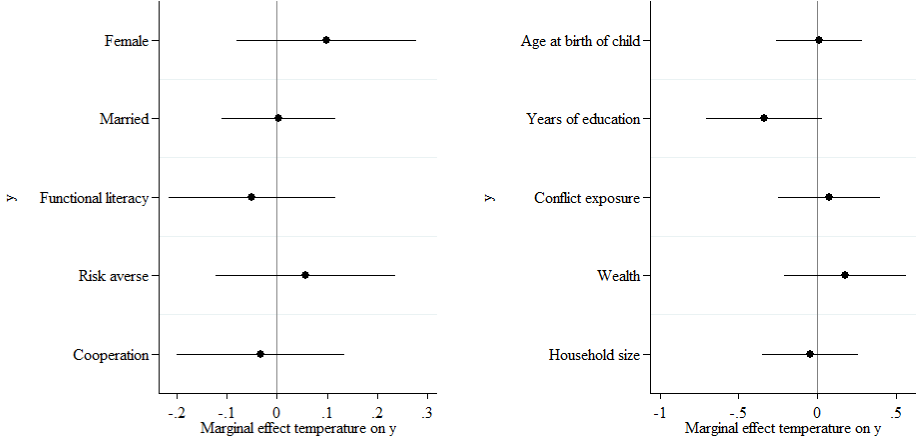
$$y_{iyms} = \alpha + \beta Temp_{ym} + \gamma' \mathbf{x}_{iyms} + \delta_m + \zeta_s + \varepsilon_{iyms} \quad (2.8)$$

where y_{iyms} refers to one of the following characteristics of the caregiver of child i born in month m of year y in village s : gender, marital status, functional literacy, risk aversion, public goods game choice, age at birth of child, years of education, conflict exposure, wealth, and household size. All other notation is the same as above.

If parents did not self-select into conceiving at the onset of a particularly hot (or cold) 9-months period based on these characteristics, the estimated β coefficients should be statistically insignificant. I summarize the estimated β coefficients and

their 95% confidence intervals in Figure 2.4. As expected, none is statistically different from zero, indicating no detectable parent self-selection bias.

Figure 2.4 – Caregiver self-selection does not drive the results



Result 2: The relationship between ambient temperature during gestation and cooperation is stable and robust to controlling for other environmental factors, early life deprivation markers, pre- and post-natal conflict exposure and caregiver preferences.

2.4.3 Long-term effects

To gauge the long-term effects of ambient temperature shocks during gestation on the taste for cooperation and to test the external validity of my main finding, I apply a similar analytical approach to a sample of adults from a different part of the country playing a different type of public goods game. I first fit the following simple OLS model:

$$Contribution_{iym} = \alpha + \beta Temperature_{ym} + \varepsilon_{iym} \quad (2.9)$$

where $Contribution_{iyms}$ represents the average amount of tokens contributed to the public good by participant i born in month m of year y and living in village s . $Temperature$ has the same meaning as above, and ε_{iyms} is a stochastic error term. Standard errors are clustered at the levels of running month of birth and game group. I find a negative effect of ambient temperature during gestation on the contribution to the common pot.

To bring the model closer to (2.2), I next extend it to include all the available confounds:

$$Contribution_{iyms} = \alpha + \beta Temperature_{ym} + \gamma' \mathbf{x}_{iyms} + \delta_m + \zeta_s + \varepsilon_{iyms} \quad (2.10)$$

where \mathbf{x}_{iyms} is a vector of personal characteristics of participant i born in month m of year y and living in village s , which is comprised of *Age* (age in months), *Female* (a dummy equal to 1 if the participant is female), and their interaction. δ_m are month of year fixed effects and ζ_s village fixed effects. The rest of the notation is the same as in (2.9).

I find a negative and statistically significant effect of ambient temperature during gestation on contribution to the public good, with every 1°C increase in temperature lowering contributions to the public good by about 0.5 token or some 13%. The result is robust to outlier exclusion (see Table 2.A.5 in the Appendix) as well as to a placebo test by older temperatures (see Table 2.A.6).

Result 3: The negative effects of exposure to unusually high ambient temperature during gestation on later-life taste for cooperation last into adulthood. A 1 °C (1 s.d.) increase in mean ambient temperature during gestation decreases contributions to the public good by about 13% (5%), thus decreasing total welfare by 6% (2%).

Table 2.8 – Long-term effects

	Contribution	
	(1)	(2)
Temperature	-0.341*	-0.468**
	(0.193)	(0.204)
Personal characteristics	N	Y
Month of birth & Village FE	N	Y
N	257	257
adj. R^2	0.006	0.140

Notes: SE clustered at the level of running month of birth and game group in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Personal characteristics: Female, Age, Age \times Female.

2.5 Discussion and conclusion

When Montesquieu (1748, book 15, chapter 1: 7) wrote that excess heat makes people “slothful and dispirited,” he pointed out that the fact is often used as a justification for slavery. It is perhaps due to the negative connotations of this argument that few social scientists studied the effects of heat on human behavior until quite recently. With global climate change driving temperatures to historically unprecedented levels, this attitude has drastically shifted.

There is now some cross-country evidence suggesting that prevailing extreme temperatures negatively affect health outcomes (Wells and Cole, 2002), and hamper economic production (Burke et al., 2015). Looking exclusively at such cross-country studies, one could be tempted to conclude that it is absolute temperatures that drive health and behavioral changes, and that geographical location largely predetermines health and economic outcomes. In such a world, children in tropical countries would be born underweight (Wells and Cole, 2002), suffer from the various negative consequences of poor birth outcomes (Black et al., 2007), and grow up in inefficient economies (Burke et al., 2015). In the context of this study, they would become less cooperative than their luckier counterparts from more temperate climates.

Within-country analyses, however, paint a more complex picture. Due to their longitudinal nature, they have to control for any trends and seasonal patterns not associated with temperature (typically by including time fixed effects in their

models), effectively netting out seasonal and long-term temperature patterns as well. Their findings suggest that unexpected deviations from normal temperatures – rather than absolute temperatures – are responsible for observed health and behavioral changes (Dell et al., 2009; Deschênes et al., 2009; Hsiang et al., 2013). In the context of this chapter, one would thus expect a person born in an unusually warm year in Northern Uganda to be less cooperative than their neighbor born in an unusually cold year. One would, however, not know whether they should be more or less cooperative than somebody born on the same day in North Holland, for example.

Relying on longitudinal data from two locations in Uganda, I follow Dell et al.'s (2009) recommendation to include time fixed effects in this chapter. I find that exposure to higher than normal ambient temperatures during gestation reduces the probability that a child contributes to the public good. The estimated effect is large and robust to controlling for potential confounders including other environmental factors, markers of early-life deprivation, prenatal stress, postnatal conflict exposure and caregiver preferences. The negative effect of prenatal exposure to abnormally high temperatures lasts into adulthood.

Thus, people's willingness to cooperate – a prerequisite for much of economic production – may decline as the likelihood of extreme temperatures increases. The welfare implications of this are substantial in my stylized behavioral games. Their estimation in practice is, however, beyond the scope of this chapter, and should instead be the focus of future research. Similarly, it will be important to study the extent to which adaptation to new climatic realities may mitigate the behavioral effects of higher temperatures. Until these questions are answered, at least the possibility of such effects should be taken into account when constructing the damage function of climate change and assessing the benefits of climate policies.

Appendix 2.A: Tables

Table 2.A.1 – The main result is robust to outlier exclusion

	Child cooperation		
	All	Low leverage	Low influence
	(1)	(2)	(3)
Temperature	-0.143** (0.067)	-0.253*** (0.077)	-0.189** (0.084)
Child & caregiver characteristics	Y	Y	Y
Month of birth & School FE	Y	Y	Y
<i>N</i>	518	434	469
Area under ROC curve	0.736	0.687	0.725

Notes: Probit marginal effects. SE clustered at the level of running month of birth in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child characteristics: Female, Age (in months), Age×Female. Caregiver characteristics: Female, Age (in years), Age×Female, Acholi, Years of education. Low leverage: $\hat{h}at < \mu_{\hat{h}at} + \sigma_{\hat{h}at}$. Low influence: $|DFBETA| < 2/\sqrt{N}$.

Table 2.A.2 – The main result is robust to spatial restrictions

	Child cooperation	
	All	Born in same area
	(1)	(2)
Temperature	-0.208*** (0.067)	-0.195*** (0.066)
Child characteristics	Y	Y
Month of birth & School FE	Y	Y
<i>N</i>	499	458
Area under ROC curve	0.742	0.734

Notes: Probit marginal effects. SE clustered at the level of running month of birth in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child characteristics: Female, Age (in months), Age×Female. Caregiver characteristics: Female, Age (in years), Age×Female, Acholi, Years of education.

Table 2.A.3 – The main result is robust to a placebo test

	Child cooperation	
	(1)	(2)
Temperature	-0.208*** (0.067)	-0.187*** (0.068)
Temperature (previous year)		-0.086 (0.075)
Child characteristics	Y	Y
Month of birth & School FE	Y	Y
<i>N</i>	499	499
Area under ROC curve	0.742	0.744

Notes: Probit marginal effects. SE clustered at the level of running month of birth in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child characteristics: Female, Age (in months), Age×Female. Caregiver characteristics: Female, Age (in years), Age×Female, Acholi, Years of education.

Table 2.A.4 – The main result is robust to alternative assumptions about gestational period

	Child cooperation			
	Probit		IV Probit	
	(1)	(2)	(3)	(4)
Temperature (normal)	-0.180** (0.070)		-0.155** (0.077)	
Temperature (premature)		-0.133** (0.065)		-0.198** (0.080)
Child characteristics	Y	Y	Y	Y
Month of birth & School FE	Y	Y	Y	Y
<i>N</i>	518	518	518	518
Area under ROC curve	0.740	0.735	0.728	0.724

Notes: Probit marginal effects. SE clustered at the level of running month of birth in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Child characteristics: age, female, age×female. Normal pregnancy extreme temperature instrumented by premature pregnancy extreme temperature in (3). Premature pregnancy extreme temperature instrumented by normal pregnancy extreme temperature in (4).

Table 2.A.5 – The long-term result is robust to outlier exclusion

	Contribution		
	All (1)	Low leverage (2)	Low influence (3)
Temperature	-0.468** (0.204)	-0.489** (0.243)	-0.730*** (0.207)
Personal characteristics	Y	Y	Y
Month of birth & School FE	Y	Y	Y
<i>N</i>	257	222	222
adj. R^2	0.340	0.346	0.447

Notes: SE clustered at the level of running month of birth and game group in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Personal characteristics: Female, Age (in months), Age×Female. Low leverage: $\hat{hat} < \mu_{\hat{hat}} + \sigma_{\hat{hat}}$. Low influence: $|DFBETA| < 2/\sqrt{N}$.

Table 2.A.6 – The long-term result is robust to a placebo test

	Contribution	
	(1)	(2)
Temperature	-0.468** (0.204)	-0.403** (0.203)
Temperature (previous year)		-0.228 (0.267)
Child characteristics	Y	Y
Month of birth & School FE	Y	Y
<i>N</i>	257	257
adj. R^2	0.140	0.140

Notes: SE clustered at the level of running month of birth and game group in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Personal characteristics: Female, Age (in months), Age×Female.

Appendix 2.B: Detailed variable definitions

Temperature. Mean ambient air temperature in the month of the respondent's birth and the preceding 8 months (in °C). All values are taken from Willmott and Matsuura (2015)—a monthly time series interpolated to a $0.5^\circ \times 0.5^\circ$ grid. To obtain the mean temperature for the children, I averaged the values from four grid cells (2.5°N - 3.5°N , 32.5°E - 33.5°E) covering the former Pader district. Values for the adults in my secondary sample come from a single grid cell (0.5°S - 1.0°S , 30.0°E - 30.5°E), which covers all the surveyed villages.

Cooperation. A dummy representing the respondent's choice in a one-shot dichotomous public goods game: "group card" or "private card". The cooperative choice "group card" takes value 1, the non-cooperative choice "private card" takes value 0.

Female. A dummy taking the value of 1 if the respondent is female, 0 otherwise.

Age. Respondent's age in months (unless specified otherwise).

Height-for-age. The child's height standardized for his age class (de Onis et al., 2007).

BMI-for-age. The child's body-mass-index standardized for his age class (de Onis et al., 2007).

IQ-for-age. The child's IQ, measured using standard Raven's matrices and standardized for his age class (in sample).

2D:4D. A child level measure of the relative length of the index finger of the right hand with respect to the ring finger (in cm).

Prenatal stress. A child level variable derived by standardizing the 2D:4D (z-score): $DigitRatio_i = -\frac{2D:4D_i - \overline{2D:4D}}{\sigma}$. The negative sign is added for ease of interpretation.

(Postnatal) conflict exposure. The child's postnatal conflict exposure is proxied by the conflict exposure of his primary caregiver multiplied by the fraction of violent conflict events that took place after the child's birth. The caregiver's conflict exposure index is derived from answers to 23 war-witnessing questions (see Macksoud, 1992).

Precipitation. Mean precipitation in the month of the respondent's birth and the preceding 8 months (in mm/day). All values are taken from Willmott and Matsuura (2015)—a monthly time series interpolated to a $0.5^\circ \times 0.5^\circ$ grid. To obtain the mean temperature for the children, I averaged the values from four grid cells (2.5°N - 3.5°N , 32.5°E - 33.5°E) covering the former Pader district. Values for the adults in my secondary sample come from a single grid cell (0.5°S - 1.0°S , 30.0°E - 30.5°E), which covers all the surveyed villages.

Consumer prices. Mean linearly de-trended consumer price index in the month of the respondent's birth and the preceding 8 months (International Monetary Fund, 2016).

Age at birth of child. Caregivers age at her last birthday before the birth of the child.

Christian. A dummy equal to 1 if the respondent is Christian by religion, 0 otherwise.

Acholi. A dummy equal to 1 if the respondent is ethnically Acholi, 0 otherwise.

Munyankole. A dummy equal to 1 if the respondent is ethnically Munyankole, 0 otherwise

Years of education. The number of years of education the respondent completed.

Married. A dummy equal to 1 if the respondent is married, 0 otherwise.

Functional literacy. A dummy equal to 0 if the respondent never reads a newspaper, 1 otherwise.

Household size. The number of household members permanently residing in the homestead.

Wealth. An index constructed following Sahn and Stifel (2003) by taking the principal factor of assets possessed by the respondent's household.

Risk aversion. A dummy equal to 0 if the respondent chose the least risky of several lottery options; 1 otherwise.

Contribution. The mean number of tokens contributed over 5 rounds of a standard public goods game.

Appendix 2.C: Public goods game cards

Figure 2.C.1 – Individual card

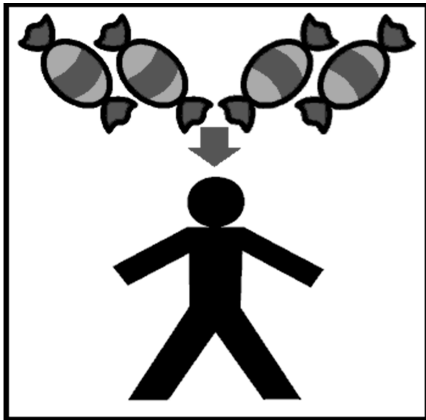
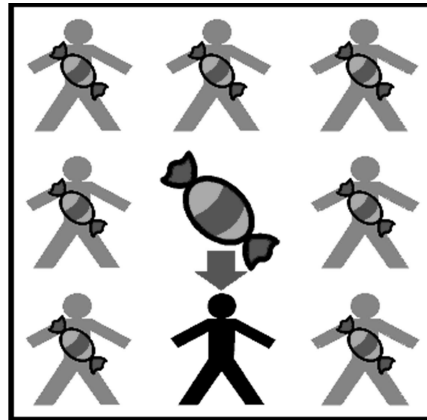


Figure 2.C.2 – Group card



Chapter 3

The Effect of Prenatal Stress on Cooperation

Evidence from Violent Conflict in Uganda

Abstract: Are preferences endogenously determined in the womb? We play a public goods game with Ugandan children born during a conflict characterized by high civilian victimization. Children whose caregivers suffer from post-traumatic stress disorder are more likely to free-ride in the game. Genetic and environmental factors alone do not explain the relationship, but children's 2D:4D digit ratio—a marker of fetal hormone exposure associated with epigenetic effects of maternal distress—does. Our findings extend the fetal origins literature to the domain of preferences. By reducing next generation's taste for cooperation, conflict may have father-reaching economic consequences than previously thought.

This chapter is based on:

Cecchi, F. and J. Duchoslav (2017). The Effect of Prenatal Stress on Cooperation: Evidence from a Violent Conflict in Uganda. Working paper.

3.1 Introduction

The nine months in utero may well be the most critical time in a person's life (Almond and Currie, 2011). Later-life characteristics are increasingly associated with "fetal origins". Suffering from severe trauma and stress during pregnancy alters the hormone exposure of the child, triggering epigenetic processes that may shape brain evolution and behavior (Dörner et al., 2001; Keverne and Curley, 2008). Among others, this was shown to affect future abilities, personality, and health trajectories. Do these in utero changes extend to the domain of social preferences?

We play a non-strategic public goods game with 442 children born in Pader district in Northern Uganda during the 1998-2006 period of intense fighting between government forces and the Lord's Resistance Army (LRA). The game requires a one-shot dichotomous choice on whether to cooperate or not—free-riding being the only dominant strategy. Simultaneously we conduct an extensive socio-economic questionnaire including violent conflict exposure (victimization), post-traumatic stress disorder (PTSD) symptoms, and a closely related public goods game with the child's main caregiver.

We find that children whose caregivers suffer from PTSD symptoms are significantly more likely to free-ride in the game. We speculate that this results from a combination of genetic and environmental intergenerational transmission mechanisms, but also epigenetic ones—through changes in hormone exposure during early fetal development. While it is impossible to measure the exact hormonal profile of fetuses later in life, we can proxy for it using one of the many biometric markers of in utero hormone exposure, which develop mostly during early gestational stages and remain relatively stable throughout life. We opt for the 2D:4D digit ratio, which measures the relative length of the index finger with respect to the ring finger, and is therefore non-intrusive and relatively easy to measure (Manning et al., 2003; Lutchmaya et al., 2004; Zheng and Cohn, 2011). Like other biometric markers of prenatal hormonal exposure, the digit ratio is established through an epigenetic process in which hormones released

by the mother's body modify the gene expressions of the fetus—rather than altering the genetic code itself. As such, it represents an interaction of nature – the child's genetic code – and nurture—the mother's hormonal reaction to environmental stimuli. In particular, the 2D:4D digit ratio is negatively related to the fetal testosterone-to-estradiol (T:E2) ratio (Manning et al., 2003; Lutchmaya et al., 2004). Higher levels of maternal distress are associated with higher fetal testosterone levels (Ward and Weisz, 1980), higher fetal T:E2 ratios (vom Saal et al., 1990), and lower offspring 2D:4D digit ratios (Lilley et al., 2010). In fact, the children in our sample exhibit significantly lower digit ratios than their biological mothers and siblings born before the war. We also find that caregiver PTSD predicts lower digit ratios in children, but caregiver digit ratios do not predict caregiver traumatization—suggesting that in our setting self-selection into trauma is not the main driver of this relationship. Under the assumption that violent conflict exposure and traumatization were conditionally unconfounded in our context (Doom and Vlassenroot, 1999; Jackson, 2002; Blattman, 2009; Blattman and Annan, 2010), we therefore argue that the digit ratio may help us discern whether prenatal stress represents a third separate causal mechanism explaining the increased free-riding.

We show that prenatal stress does indeed affect preferences for cooperation: one standard deviation from the mean of our marker reduces the child's probability of contribution to the public good by around 7% points (i.e. -17% of the mean prevalence). To test whether this is a separate mechanism we introduce nature and nurture influences on the child, captured by the caregivers' choice in a closely related public goods game and in a simple risk game. As expected, caregiver cooperation significantly influences the child's independent decision to cooperate, while their risk preferences do not. Importantly, the effect of prenatal stress on cooperation remains robust and stable. Next we control for alternative mechanisms that may potentially yield a spurious correlation, such as early life deprivation and conflict exposure after birth (both likely correlated with prenatal stress levels), and discuss the sensitivity of our findings to exogeneity assumptions, investigating the likelihood that they are driven by unobserved characteristics.

We shed light on an alternative, epigenetic mechanism of preference transmission—beyond the standard nature-nurture debate. Preferences for cooperation may be endogenously determined right from the womb. If prenatal stress affects next generation’s taste for cooperation, violent conflict may have farther reaching socio-economic consequences than previously thought. High incidence of severe prenatal stress may affect long-run development trajectories and post-conflict recovery across generations, even if the episodes of violence are limited in time.

The rest of the chapter is organized as follows. Section 3.2 reviews the literature on the subject. Section 3.3 describes the context and background. Section 3.4 outlines the experimental design. Section 3.5 illustrates the empirical strategy and results, and Section 3.6 concludes.

3.2 Conflict, preferences, and prenatal stress

This chapter explores the fetal origins of social preferences. It builds upon three strands of literature—that on the role of violent conflict in shaping preferences, that on the consequences of trauma in utero, and that on the relationship between prenatal hormone exposure and economic behavior.

The relationship between violent conflict and the functioning of societies has been at the forefront of economic debate for years. War violence persistently impacts health, education, and poverty (Ghobarah et al., 2003; Chamarbagwala and Morán, 2011; Gates et al., 2012), but also affects individual preferences and behavior. It has been found to increase community participation and political engagement (Bellows and Miguel, 2009; Blattman, 2009), out-group aggressiveness and competitiveness (Miguel et al., 2011; Cecchi et al., 2016), as well as risk propensity, and discount rates (Voors et al., 2011; Callen et al., 2014). Individuals exposed to inter-community violence display more altruistic behavior, higher public good contributions, and trust within their networks (Voors et al., 2012; Gilligan et al., 2014). Intra-community violence instead decreases social cohesion and trust, and increases sentiments of group identity (Cassar et al., 2013; Rohner

et al., 2013). Given the existing evidence of the transmission of preferences across generations (Dohmen et al., 2012), these changes may be expected to be passed on to the next generation both genetically (i.e. in case of survival bias) and through changes in nurture. Yet, especially for children conceived during or soon after severe violence, a third mechanism may be at play: that of epigenetic alterations caused by maternal stress while still in utero. To the best of our knowledge this is the first study attempting to gauge the effect conflict-related stress on the social preferences of individuals that had yet to be born at the time that it first affected them.

Several studies have investigated the physical and psychological consequences of stress and traumatic events in utero. In fact, maternal prenatal anxiety may suppress the development of a functioning immune system, increasing the incidence of several health complications in infants (Stott, 1973). Moreover, exposure to violence during pregnancy has been found to deteriorate birth outcomes, typically in terms of birth weight, fetal growth, and preterm delivery (Mancuso et al., 2004; Lauderdale, 2006; Camacho, 2008; Black et al., 2016; Koppensteiner and Manacorda, 2016; Quintana-Domeque and Ródenas-Serrano, 2017).¹ Independent positive shocks, such as a rise in cocoa price at birth – expected to reduce financial distress in cocoa-producing areas of Ghana with respect to other regions – decrease the likelihood of mental distress during adulthood (Adhvaryu et al., 2015). Prenatal negative shocks, such as extreme weather or military invasions, are instead associated with higher prevalence of schizophrenia and autism (van Os and Selten, 1998; Walder et al., 2014).

In adults, the prolonged emotional disturbance and distress induced by conflict increases the likelihood of trauma, mental health problems, and PTSD (de Jong, 2002; Lopes Cardozo et al., 2004; Miller et al., 2008). The higher the post-traumatic hormonal release, the greater the chance that subjects develop PTSD (Delahanty et al., 2000). PTSD, in turn, increases the likelihood of

¹While the mechanisms through which maternal stress influences birth outcomes are still subject to intensive investigation, the leading hypotheses identify three channels: endocrine, inflammatory/immune, and behavioral. See Hobel et al. (2008) and Dunkel Schetter (2011) for an overview of the biopsychosocial processes.

persistent hormonal imbalance, particularly with respect to cortisol—the human stress hormone (de Kloet et al., 2008; Song et al., 2008; Steudte et al., 2011). During pregnancy, maternal stress is transmitted to the fetus through hormonal releases (Mancuso et al., 2004; Weinstock, 2008). Due to obvious ethical considerations, experimental studies of this mechanism are generally limited to laboratory rodents. Prenatally stressed rodents exhibit higher concentrations of serum testosterone and higher testosterone to estradiol ratios (Ward and Weisz, 1980; vom Saal et al., 1990). In turn, the fetal testosterone to estradiol ratio is negatively related to the 2D:4D digit ratio (Manning et al., 2003; Lutchmaya et al., 2004). Connecting the dots, Lilley et al. (2010) show that high levels of maternal corticosterone, the rodent equivalent to cortisol, are associated with lower offspring 2D:4D digit ratio. They suggest that the latter may be a useful phenotypic indicator of maternal distress during early fetal development.

The 2D:4D digit ratio measures the relative length of the index finger with respect to the ring finger. It is established through changes in gene expression which take place without a change in the DNA sequence – known as epigenetic modifications (Jirtle and Skinner, 2007) – and is widely accepted as a noninvasive marker and “lifelong signature of prenatal hormonal exposure” (Zheng and Cohn, 2011). During early fetal development, increased androgen hormones (e.g. testosterone) or the inactivation of the estrogen receptor (ER- α) stimulate the ring finger growth, which leads to a lower 2D:4D ratio. On the other hand, the addition of estrogen (e.g. estradiol) or the inactivation of the androgen receptor (AR) decrease the ring finger growth, resulting in a higher 2D:4D ratio (Lutchmaya et al., 2004; Zheng and Cohn, 2011).

But why should prenatal hormonal changes affect later-life social preferences? On the one hand, the literature on violent conflict and preferences shows that economists increasingly agree that preferences are not hard-wired traits, but rather may change in response to shocks and life experiences. This literature however falls short of investigating if preferences can be shaped right from the womb. On the other, the “fetal origins” literature does investigate the relationship between prenatal hormones and later life cognitive and behavioral outcomes. Several

studies have shown a correlation between the digit ratio and preferences. In the lab, Garbarino et al. (2010) show that a low 2D:4D is associated with greater risk-taking. This is confirmed by evidence that low digit ratio MBA students self-select more into risky finance careers (Sapienza et al., 2009), and that the financial ability among male high-frequency traders is negatively related to their 2D:4D ratio (Coates et al., 2009). Its relationship to social preferences is relatively less studied, especially on non-experimental populations. Among undergraduate students, Brañas-Garza et al. (2013) find a non-monotonic impact of the digit ratio on altruism. Also, self-assessed low digit ratios ($2D < 4D$) predict lower giving in ultimatum, trust, and public goods games (Buser, 2012). We combine the literature about the epigenetic effects of maternal distress with that on the correlation between the digit ratio and preferences, bringing them to a natural experimental setting characterized by indiscriminate violence against unarmed civilians and high prevalence of maternal traumatization. We use the 2D:4D digit ratio as a marker of maternal distress to explore how prenatal stress reflects on the preferences for cooperation of the next generation, trying to isolate its causal mechanism from other potential influences of direct and indirect violent conflict exposure.

3.3 Context and background

In the last 25 years Uganda achieved high and steady GDP growth rates, averaging about 6.7% per year (World Bank, 2016). However, since independence in 1962, Uganda has witnessed many long periods of violence and constitutional suspension and only few short periods of peace and relative prosperity. In fact, even while the country's overall growth rate was faster than that of many of its neighbors by the end of the millennium, the North was enduring the last of a long series of conflicts: Joseph Kony's LRA insurgency (1987-2006).

Violence has been escalating recurrently in Uganda since 1971, when Idi Amin took power from the discredited President Milton Obote.² Amin ruled the country until

²In 1966 Obote was implicated in a corruption scandal together with the then commander of

the 1979 Uganda-Tanzania War led to his ousting. Obote's comeback instead triggered the Ugandan Bush War against the southern rebels of the National Resistance Army (NRA) headed by the current President of Uganda, Yoweri Museveni. Obote lost power for the second time in 1985, shortly before the NRA faction assaulted Kampala—gaining the power it still holds today (Finnström, 2008).

As the balance of power shifted southwards, rebel movements in the North gathered under the flag of the LRA (Doom and Vlassenroot, 1999). Limited in numbers and resources, the LRA resorted to pillaging villages and abducting local youth: an estimated 60,000 to 80,000 people were abducted over two decades (Annan et al., 2006; Pham et al., 2007). Throughout the years, weak government responses and the setup of an Acholi self-defense militia invigorated the LRA, which scaled up operations to discipline the local population (Finnström, 2008; Dolan, 2009).³ The widespread killing and mutilation of Acholi civilians escalated dramatically after 1996, especially in the Acholi districts of Gulu, Kitgum, and Pader.⁴ In these districts, more than a quarter of the males aged 10 to 25 at the end of the war had been abducted for at least two weeks, and only 80% of them returned from captivity (Blattman and Annan, 2010).

Pader district was particularly hit from 1998 onwards, when LRA operations gained momentum and moved southwards (Figure 3.1).⁵ Civilian fatalities peaked in 2002, after the start of “Operation Iron Fist” against rebel bases in South Sudan set off a bloody reaction by LRA forces. A truce between the LRA and the government was signed in 2006, and fighting in Uganda has been sporadic ever since.⁶

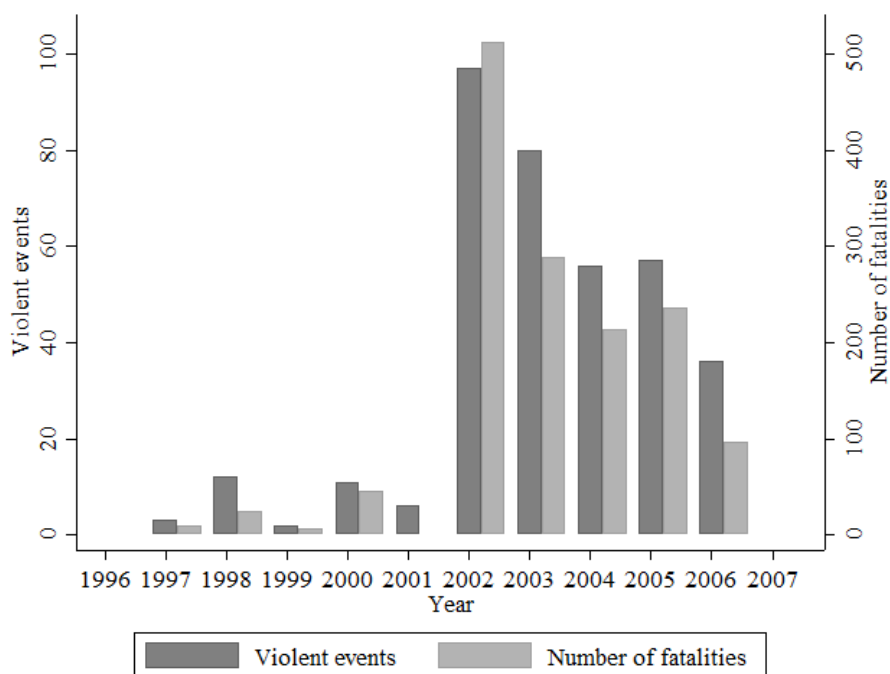
armed forces, Idi Amin. He responded by suspending the constitution.

³Civilian victimization in these years was not only the result of LRA violence, and abuse from government troops was not uncommon.

⁴Between 1994 and 2002, in response to Uganda's support for the rebels in South Sudan, the Sudanese government provided the LRA with logistic support and military equipment.

⁵Source: ACLED Version 5, 1997-2013

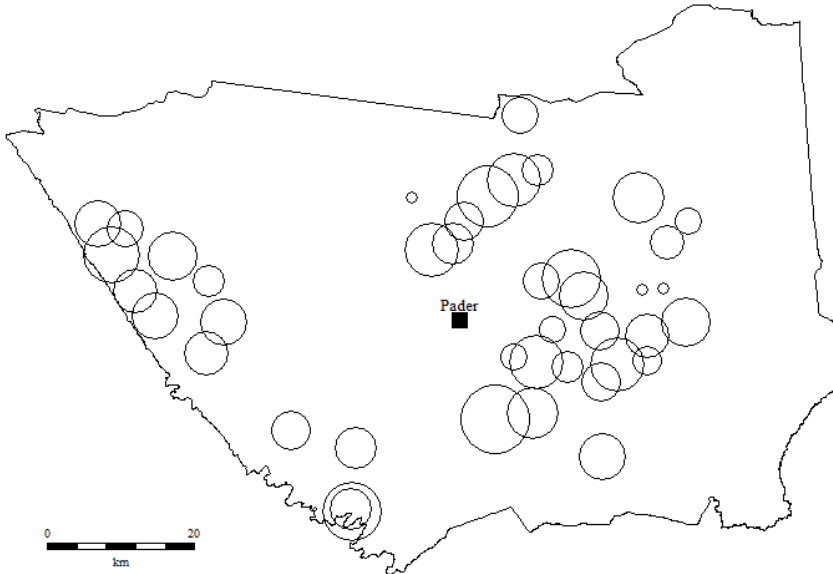
⁶The LRA has not been disarmed nor demilitarised following the 2006 truce with the Ugandan government, and has been active in the Democratic Republic of the Congo, the Central African Republic and South Sudan.

Figure 3.1 – Civilian targeting in Pader district

At its peak, LRA violence was largely indiscriminate and apparently random in nature. Small groups of roughly 15 fighters ventured into Uganda from their Sudanese bases for weeks at a time, ambushing government forces and raiding homesteads for supplies and recruits along the way. Homesteads, which tend to be relatively isolated in Northern Uganda, were targeted in an unplanned, arbitrary manner, regardless of their wealth or make-up (Doom and Vlassenroot, 1999; Jackson, 2002; Blattman, 2009). The randomness of the attacks was oftentimes deliberate, serving as a force multiplier to increase the perception of undiscerning threat in the civilian population (Vinci, 2005). Figure 3.2 shows the relative level of traumatization (PTSD), corresponding with the diameters of the circles representing each school in our sample. Coherent with literature, we find that traumatization shows no detectable spatial autocorrelation on the school

level (Moran's $I = -0.030$), nor are they correlated within schools (intra-class correlation coefficient is $\rho = 0.020$).⁷ We come back to the plausibility of selection bias and the sensitivity of our results to the unconfoundedness assumption in Section 3.5.5.

Figure 3.2 – Average traumatization per school



3.4 Experimental design

Employing new data from behavioral games, biometric measurements and an extensive socioeconomic survey, we exploit the quasi-experimental variation in war exposure generated by rebel raids in Northern Uganda villages to gauge the impact of prenatal stress on later-life cooperation.

⁷See Figure 3.B.1 for a graphic representation of the distribution of war exposure instead. See Tables 3.A.1 and 3.A.2 for full results of spatial autocorrelation and intra-class correlation analysis respectively for both war exposure and traumatization.

3.4.1 Sample and setting

Our sample includes 442 children and their caregivers from Pader district in Northern Uganda. In November 2012 we visited 42 primary schools in the district, and randomly selected 12 students from a list of pupils enlisted at the beginning of the year.⁸ The descriptive statistics for the children are presented in Table 3.1. The average digit ratio – our marker of prenatal stress – is 0.94 in our sample, and 41% of the children chose to cooperate in the public goods game. Sections 3.4.2 and 3.4.3 below discuss the details of the digit ratio measurement and of the public goods game set-up respectively. On average the children are 11 years old, half of them are female, and their body mass and height are respectively 0.4 and 1 standard deviation below the mean for their age (de Onis et al., 2007).⁹ To control for additional potential confounds, we also collect information about the children’s cognitive ability (IQ) through standard Raven’s progressive matrices (Kaplan and Saccuzzo, 2012), as well as their time and risk preferences (Voors et al., 2012). Child postnatal war exposure is a composite measure derived from the exposure of the caregiver and the child’s year of birth (see Section 3.4.4).

To account for environmental and genetic effects on preferences, we also interviewed each child’s main caregiver—the adult household member with whom the child spends most time. The descriptive statistics for the caregivers are presented in Table 3.2. About half of the caregivers in our sample chose to cooperate in the public goods game (see Section 3.4.3 for details about the public goods game). Caregivers are on average 42 years of age,¹⁰ 53% are female. Additionally, we collected information about their education level and risk preferences. All caregivers were exposed to at least some kind of conflict-related violence (though the level of exposure varies greatly in our sample),¹¹ and

⁸The randomization was stratified according to grade: 4 students were selected from grade 2, 4 from grade 4, and 4 from grade 6. Out of a total of 504, 62 students born prior to the intensification of violence in the area around 1998 were excluded from the analysis.

⁹Based on WHO recommendations for treating outliers (de Onis et al., 2007), we truncate the biometric data at 6 standard deviations from the mean. This results in 1 and 2 dropped observations for height-for-age and BMI-for-age respectively.

¹⁰Two caregivers did not know their age, reducing the number of observations to 440.

¹¹Two caregivers refused to complete the war exposure module of the survey, reducing the

Table 3.1 – Descriptive statistics (children)

Variable	Mean	Std. Dev.	Min.	Max.	Observations
2D:4D	0.943	0.042	0.733	1.111	442
Public good contribution	0.414	0.493	0	1	442
Age	11.045	2.160	6	14	442
Female	0.500	0.501	0	1	442
School grade	3.575	1.494	2	6	442
Height-for-age	-0.428	1.381	-4.82	5.80	441
BMI-for-age	-1.034	1.088	-5.98	2.01	440
IQ-for-age	95.59	15.00	79.92	162.73	442
Time preferences	0.253	0.435	0	1	442
Risk preferences	0.403	0.491	0	1	442
War exposure	0.537	0.249	0	0.958	440

Notes: See Appendix Appendix 3.C: for variable definitions.

40% suffer from PTSD symptoms. The details of war exposure and trauma measurements are discussed in Section 3.4.4 below. Almost the entire sample is ethnically Acholi and Christian by religion. Households are typically composed of 8 people. We also collected information about their relative asset wealth (Sahn and Stifel, 2003).

Table 3.2 – Descriptive statistics (caregivers)

Variable	Mean	Std. Dev.	Min.	Max.	Observations
2D:4D	0.957	0.038	0.844	1.066	154
Public good contribution	0.507	0.501	0	1	442
Age	42.03	11.87	19	92	440
Female	0.534	0.499	0	1	442
Education level	0.839	0.802	0	4	442
Risk preferences	0.312	0.464	0	1	442
War exposure	0.752	0.166	0.174	1	440
PTSD (dummy)	0.391	0.489	0	1	442
PTSD (factor)	0.000	1.000	-1.926	2.339	442
Christian	0.991	0.095	0	1	442
Acholi	0.973	0.163	0	1	442
Household size	8.093	2.988	2	25	442
Asset index	0.000	1.000	-0.571	5.507	442

Notes: See Appendix Appendix 3.C: for variable definitions.

number of observations to 440. This reduction carries over to the measure of child postnatal war exposure, which is derived from that of their caregiver.

In our setting, information about the current main caregivers can only serve as a proxy for environmental and genetic influences to which the children have been subjected throughout their lives. Of the 442 caregivers in our sample, only 190 are biological mothers. Another 163 are biological fathers, while the remaining 87 were grandparents, uncles/aunts, siblings or other relatives (in descending order of prevalence). 2 caregivers were not blood-related to the child at all. Nonetheless, the average caregiver in our sample had been taking care of the child for 93% of the child's life, making the information about the caregivers a strong proxy for the environment surrounding the children.

3.4.2 The digit ratio

The index and ring finger lengths were measured on the ventral surface of the right hand from the midpoint of the basal crease to the tip of the digit. Given the contextual constraints and instruments available, measurement precision does not exceed 1 mm, resulting in an error of $\pm 3.3\%$ at the mean of our estimations.¹² While this is still far from the precision obtained in the lab (see Voracek et al., 2007), independent raters measured the digit lengths unaware of their scientific significance; errors should therefore result in unbiased random noise.¹³

3.4.3 Public goods game

We measure children's and caregivers' willingness to cooperate by involving them in a one-shot dichotomous public goods game similar to Cárdenas et al. (2009) and Barr et al. (2014). In each school, children played in randomly assigned groups of 6, and anonymously decided whether to select a "private card" or a "group

¹²In a pilot, 30 raters separately measured 35 right hands, revealing comparable margins.

¹³Approximately two years after the main data collection we re-measured the digit lengths for a non-random sub-sample of 258 respondents. While the absolute length of the fingers had undoubtedly changed in the meanwhile, their ratio should remain relatively stable throughout lifetime. In line with expectations, the average error was $\pm 3.7\%$. Results are not driven by systematic measurement error, and excluding measures with potentially greater error does not significantly alter the results (see Table 3.A.3).

card”.¹⁴ Children could in no way infer which other 5 participants belonged to their group (out of the 11 other children selected in that school). The private card allotted 3 candies to themselves and none to other unknown group members. The group card gave instead 1 candy to each group member including themselves (a graphical representation of the two cards can be found in Figure 3.B.2). The joint surplus is therefore maximized when all participants choose the group card, such that each group member receives 6 candies. Nevertheless, free riders selecting the private card may obtain up to 8 candies. The Nash equilibrium is reached if everyone selects the private card, receiving 3 candies only. Caregivers played a very similar game, but played in groups of 12 instead of 6. The game was played in an isolated environment – typically their home – and caregivers were unaware of the identity of other participants. The private card was worth 4,000 UGX, equivalent to approximately 1.5 USD; the group card was instead worth 500 UGX. The non-cooperative equilibrium thus yielded 4,000 UGX each, joint maximization returned 6,000 UGX each, and free riders could earn as much as 9,500 UGX.¹⁵ On average, 41% of the children and 51% of the caregivers opted for the cooperative option offered by the group card.

3.4.4 War exposure and trauma

Given their young age at the time of the conflict, we do not ask children war-related questions. Instead, we use information on the individual war exposure of their caregivers, and weigh it against the war violence that happened after the year of birth of the child. We use an adapted version of the War Trauma Questionnaire (WTQ), excluding the questions about shelling and bombardment which are

¹⁴Contrarily to many public goods games in which participants can choose their preferred contribution level, we opted for a dichotomous choice: respondents could either cooperate or not. While this reduced our ability to pick up the nuances present in the experimental sample, we believe that it facilitated the decision making process, especially for the youngest.

¹⁵The variation in pay-outs between the child and caregiver versions of the game was determined during a pilot. We adjusted the relative values of the “private” and “group” cards to maximize the variance of the responses in both samples and thus facilitate parametric analysis. Specifically, the number of candies assigned by the “private” card in the child version was dropped from 4 to 3 to increase the likelihood that children would select the “group” card.

not relevant to our setting (Macksoud, 1992; Papageorgiou et al., 2000). This questionnaire provides information on 23 war related traumatic events that a person may have witnessed, rated through “yes” or “no” statements. We create a war exposure (victimization) index using the average of positive responses to these violence related questions (Bellows and Miguel, 2009). On average, caregivers responded positively to 75% of the questions, with a minimum observed exposure of 17%. To proxy the child’s postnatal war exposure we weight the caregiver’s war exposure index by the portion of violence potentially witnessed by the child after birth. To this end, we take the fraction of total civilian fatalities that occurred in Pader district following the child’s birth (see Figure 3.1).¹⁶

We measure caregiver PTSD symptoms using the civilian version of the PCL self-report checklist (Weathers et al., 1993).¹⁷ We convert individual scores into a PTSD dummy, following the recommendations of the US Department of Veteran Affairs, and the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM4).¹⁸ As an additional robustness check, we perform factor analysis to extrapolate a factor for trauma.

3.5 Analysis and results

We start with the observation that children of caregivers with PTSD symptoms are less likely to contribute to the public good. By showing that caregiver PTSD

¹⁶Source: ACLED Version 5, 1997-2013

¹⁷The civilian version of the PCL self-report checklist – a 17-item questionnaire – has been found to have strong psychometric properties, high internal consistency, and high test-retest reliability (Blanchard et al., 1996; Ruggiero et al., 2006; Conybeare et al., 2012). Moreover, it is strongly correlated with alternative measures of PTSD such as the Mississippi, MMPI-2 Keane, IES, and CAPS scales (Weathers et al., 1993; Dobie et al., 2002; Freedy et al., 2010).

¹⁸In our setting, we expect high rates of PTSD (Roberts et al., 2008; Pfeiffer and Elbert, 2011). We therefore take a conservatively high threshold for PTSD, at >66% of the maximum item score, to minimise the likelihood of false positives (Keen et al., 2008). We therefore find a lower PTSD prevalence (40%) than previous studies in the region that do not apply this correction. The DSM4 cut-off point requires at least 1 moderately positive answer in questions 1-5, 3 in 6-12, and 2 in 13-17. At the selected threshold, only 4 out of 442 caregivers do not meet this requirement. Our analysis is robust to their inclusion or exclusion from the PTSD count.

predicts lower 2D:4D digit ratio in their children, and that children born during conflict have lower digit ratios than their biological mothers and same-mother siblings, we confirm that the 2D:4D digit ratio is a reliable marker of prenatal stress. Together with caregiver preference for cooperation, it explains the entire effect of caregiver PTSD on child behavior in the public goods game. We continue by showing that the relationship between prenatal stress and contribution to the public good is robust to controlling for other mechanisms through which conflict could influence cooperative preferences, including early life deprivation and direct conflict exposure. Finally, we account for self-selection into conflict exposure and trauma along observable characteristics through propensity score matching, and show that any unobserved factor would have to have an unrealistically large effect on cooperative behavior in order to change our result.

3.5.1 Caregiver PTSD and child cooperation

We first test whether caregiver PTSD significantly changes the child's likelihood to free-ride in our public goods game. Table 3.3 shows that children whose caregivers suffer from PTSD symptoms are up to 9% points less likely to contribute to the public good (i.e. -22% of the mean prevalence). The result is robust to controlling for sub-county fixed effects, child and caregiver characteristics, household size and asset index, as well as child risk and time preferences (columns 1 to 5).

Result 1: Children whose caregivers suffer from PTSD symptoms are robustly less likely to contribute to the public good.

3.5.2 Prenatal stress and child digit ratio

While this result is interesting in itself, it does not say much about the mechanisms through which traumatized caregivers may transmit a preference for lower cooperation to the child. The literature on the intergenerational transmission of preferences predicts that caregivers transmit their own preferences both genetically

Table 3.3 – Caregiver PTSD predicts lower child cooperation

	Public Good Contribution				
	(1)	(2)	(3)	(4)	(5)
Caregiver PTSD	-0.072** (0.035)	-0.070** (0.033)	-0.087** (0.036)	-0.091** (0.036)	-0.089** (0.038)
Child controls	N	Y	Y	Y	Y
Caregiver and peer controls	N	N	Y	Y	Y
Sub-county FE	N	N	N	Y	Y
Child risk and time preferences	N	N	N	N	Y
Household size and assets index	N	N	N	N	Y
Observations	442	442	440	440	440
R^2	0.005	0.011	0.038	0.041	0.042

Notes: Standard errors corrected for school level clustering (42) are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***. Child controls: Age, Female, Age×Female; Caregiver controls: Age, Female, Age×Female, Acholi, Education; Peer controls: Proportion of excluded peers, Class dummies. See Table 3.A.6 for the full list of coefficients.

(i.e. directly transmitting their own hard-wired preference set) and through environmental influences (i.e. more or less voluntarily “teaching” the child to free-ride or cooperate). We hypothesize that a third mechanism may be at play: the epigenetic effect of prenatal stress. To proxy for this, we introduce a standardized measure of the 2D:4D digit ratio—an indicator of maternal distress during early fetal development. As digit ratios are expected to become smaller in response to increased maternal distress during early fetal development (i.e. increased fetal testosterone to estradiol ratios), we use the negative standardized digit ratio as relative measure of prenatal stress:

$$DigitRatio_{is} = - \frac{2D : 4D_{is} - \overline{2D : 4D}}{\sigma} \quad (3.1)$$

where $2D : 4D_{is}$ is the digit ratio of individual i ; $\overline{2D : 4D}$ is the mean digit ratio in the sample; σ is the standard error; and the negative sign produces a positive relationship between our proxy and actual prenatal stress—for ease of interpretation.

Yet, for our marker of prenatal stress to plausibly explain part of the relationship we observe in Table 3.3, we need to find evidence in our data that it is affected by victimization during the conflict. Through a series of paired t-tests we compare the digit ratio of (84) biological mothers to that of their offspring (Table 3.4, row

1), and of their female offspring only (43) (Table 3.4, row 2). Moreover, we identify same-mother siblings born between 1990 and 1996, a relatively non-violent period in Pader district. We test the hypothesis that (43) siblings born prior to the intensification of war violence have a significantly different digit ratio (Table 3.4, row 3), and verify its robustness by looking at (26) same-sex siblings (Table 3.4, row 4). Children born during the conflict have significantly smaller digit ratios than their relatives born before the conflict.¹⁹

Table 3.4 – Children born during the conflict exhibit lower digit ratios than their relatives

	Observations		2D:4D		
	Child	Control	Child	Cont.	Diff.
Biological mothers	84	84	0.942	0.956	-0.013*** (0.005)
Biological mothers (female child)	43	43	0.940	0.953	-0.013** (0.006)
Same-mother sibling	43	43	0.936	0.948	-0.013* (0.008)
Same-mother sibling (same sex)	26	26	0.941	0.950	-0.009 (0.008)

Notes: * $p < 0.10$, ** $p < 0.05$, ***.

This result is in line with previous evidence that prenatal stress results in smaller digit ratios (Ward and Weisz, 1980; vom Saal et al., 1990; Lilley et al., 2010). In particular, we observe a reduction in the digit ratio of approximately 1/4 of a standard deviation. Next we test whether PTSD of the caregiver actually predicts lower digit ratio in children using the following model:

$$2D : 4D_{is} = \alpha + \beta \text{CaregiverPTSD}_{is} + \gamma' \mathbf{X}_{is} + \zeta_s + \varepsilon_{is} \quad (3.2)$$

where $\text{CaregiverPTSD}_{is}$ is a measure of the PTSD symptoms of the caregiver, \mathbf{X}_{is} is a vector of child characteristics including age, gender, and their interaction, and caregiver characteristics including age, gender, their interaction, ethnicity

¹⁹This involved a separate step of data collection: in 2014 we traced back as many mothers and older siblings of the sampled children as possible given the time gap, and measured their digit ratios.

and education. ζ_s represents sub-county fixed effects, and ε_{is} is a stochastic error term.

Table 3.5 shows that PTSD in caregivers is associated child 2D:4D digit ratios that are lower by approximately 0.3 of a standard deviation (columns 1 and 2). As a robustness check, columns 3 and 4 show instead a traumatization factor derived from the same PTSD questions: one standard deviation increase in traumatization results in a lower digit ratio by about 0.1 standard deviations (columns 3 and 4).

Table 3.5 – Caregiver trauma predicts lower digit ratio in children

	Standardized digit ratio			
	(1)	(2)	(3)	(4)
PTSD	-0.296** (0.111)	-0.289** (0.123)		
Trauma factor			-0.116** (0.051)	-0.109* (0.058)
Child controls	N	Y	N	Y
Caregiver controls	N	Y	N	Y
Sub-county FE	N	Y	N	Y
Observations	442	440	442	440
R^2	0.021	0.043	0.013	0.035

Notes: Standard errors corrected for school level clustering (42) are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***. Child controls: Age, Female, Age×Female; Caregiver controls: Age, Female, Age × Female, Acholi, Education.

Result 2: Children born during the conflict exhibit significantly lower digit ratios than their biological mothers and same-mother siblings. Also, caregiver trauma predicts lower digit ratios in children.

We then proceed to test whether genetic, environmental, and epigenetic mechanisms explain away the relationship of Table 3.3. We first introduce as additional explanatory variable the caregiver’s choice in the public goods game to capture the effects of nature and nurture (columns 1 and 2), and later our marker of prenatal stress (columns 3 and 4). Column 1 and 2 of Table 3.6, show that while the coefficient of PTSD is lower, it is still significant. Columns 3 and 4 of Table 3.6 show that once the 2D:4D digit ratio is introduced, the direct effect is further

reduced and becomes insignificant. Moreover the adjusted R^2 actually increases once PTSD is removed from the model.

Table 3.6 – Caregiver cooperation and prenatal stress explain the effect of PTSD on child cooperation

	Public Good Contribution			
	(1)	(2)	(3)	(4)
Standardized digit ratio (-)			-0.069** (0.026)	-0.066** (0.026)
Caregiver public good contribution	0.150*** (0.040)	0.152*** (0.045)	0.141*** (0.042)	0.142*** (0.047)
PTSD	-0.056 (0.035)	-0.077** (0.036)	-0.036 (0.038)	-0.059 (0.038)
Child controls	N	Y	N	Y
Caregiver and peer controls	N	Y	N	Y
Sub-county FE	N	Y	N	Y
Observations	442	442	442	442
R^2	0.028	0.033	0.047	0.051

Notes: Standard errors corrected for school level clustering (42) are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***. Child controls: Age, Female, Age×Female; Caregiver controls: Age, Female, Age×Female, Acholi, Education; Peer controls: Proportion of excluded peers, Class dummies. See Table 3.A.6 for the full list of coefficients.

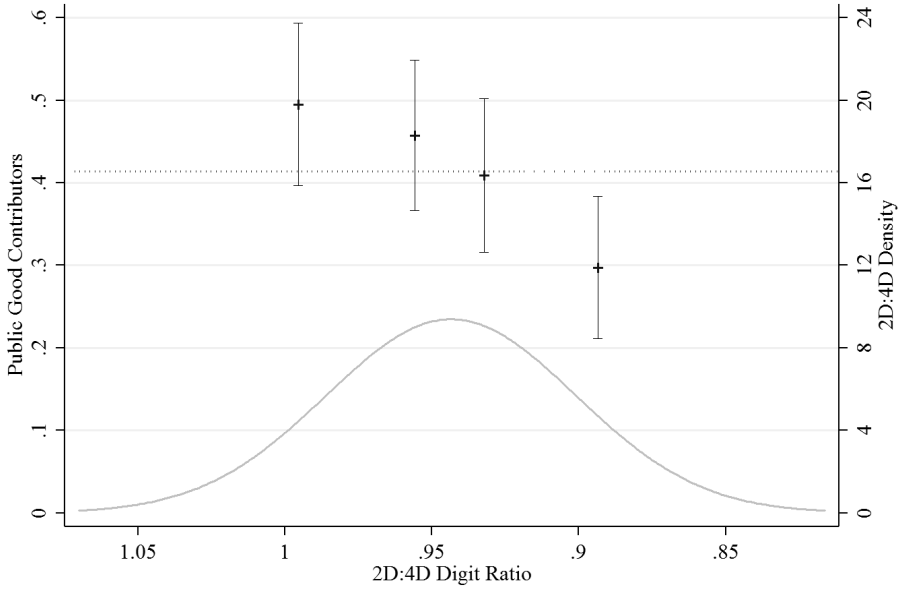
3.5.3 Child digit ratio and public good contribution

Having established a negative relationship between caregiver PTSD and child digit ratio, we continue by analyzing the relationship between children’s digit ratios and their cooperation in the public goods game graphically. From Buser (2012) we expect it to be negative. Indeed, Figure 3.3 illustrates that the prevalence of public good contributions decreases among children with lower digit ratio—as prenatal stress increases.

Parametrically, we show this relationship by estimating a specification with only the prenatal stress proxy as a regressor:

$$Cooperation_{is} = \alpha + \beta DigitRatio_{is} + \varepsilon_{is} \quad (3.3)$$

where $Cooperation_i$ is a dummy taking value of 1 if individual i played the group card in the public goods game; $DigitRatio_{is}$ is our measure of prenatal stress. Standard errors are clustered for 42 schools.

Figure 3.3 – Digit ratio quartiles and prevalence of public good contributions

We find that our marker of prenatal distress is negatively correlated with the child's probability of contribution to the public good (Table 3.7, column 1). Yet, this relationship could potentially be driven by other child characteristics. First, of course, the preferences of children may be driven by those of their caregivers through environmental as well as genetic mechanisms (Dohmen et al., 2012). We find a strong relationship between the social preferences of the caregiver and the child, but not between the risk preferences of the caregiver and the social preferences of the child. Children are about 15% points more likely to contribute if their main caregiver contributes to the public good in a separate game.²⁰ Second, prosocial preferences develop throughout childhood and adolescence, and become increasingly gender-dependent with approaching adulthood (Eisenberg et al.,

²⁰For a sub-sample of 154 parents, we verify that the parents' digit ratio is not driving our result—i.e. the genetic component does not foreshadow the epigenetic effect. We find no effect of parental digit ratios on the cooperation of children (see Table 3.A.4).

2006). Controlling for age, gender, their interaction, for school grade and for tribe does not, however, alter the result (see Table 3.7, column 3). Third, the probability of contributing to the public good may also be influenced by the social environment in which a child is growing up. Both family and peer demographics have been linked to children's prosocial behavior, though the evidence is often inconclusive (Eisenberg et al., 2006). To account for possible environmental influences, we control for caregiver characteristics including the level of education, gender, age, and the interaction of the latter two. We also control for the proportion of children who participated in the public goods game, but were excluded from the analysis due to their age (column 4).²¹ The inclusion of these additional covariates does not affect the main result (see Table 3.7, column 3), nor does the inclusion of spatial fixed effects (see Table 3.7, column 5). We also control for family size and wealth, both of which have been associated with children's prosocial preferences (Zaff et al., 2003), as well as for the child's risk and time preferences, to reach the following fully specified linear probability model in column 6:²²

$$Cooperation_{is} = \alpha + \beta DigitRatio_{is} + \gamma' \mathbf{X}_{is} + \delta_s + \zeta_s + \eta' \mathbf{Z}_{is} + \varepsilon_{is} \quad (3.4)$$

²¹We initially sampled 4 second-graders, 4 fourth-graders and 4 sixth-graders from each of the 42 schools, resulting in a total sample of 504 children. All 12 children from each school would participate in the public goods game, but only those born after the re-escalation of violence in Pader district in 1998 were included in our analysis, resulting in a removal of between 0 and 5 children per school. Since prosocial preferences increase with age, younger children playing the game in a group with a high share of older children may be more likely to contribute to the public good due to both peer effects and rational expectations about the behavior of the older children.

²²Literature highlights several trade-offs between linear probability (LPM) and probit models. First, compared to a probit, the LPM does not estimate the structural parameters, but this chapter is mostly concerned with marginal effects (intuitively interpretable with LPM). Second, LPM error terms are heteroskedastic by construct; we thus use cluster robust standard errors, which are heteroskedasticity-consistent. Finally, Hoxby and Oaxaca (2006) show that the potential bias of LPM increases with the fraction of predicted probabilities that lie outside the (unconstrained) unit interval. In our main specification, the predicted probabilities lie between 0.003 and 0.815; we thus expect our estimations to be largely unbiased and consistent. In fact, marginal effects calculated through a probit very closely resemble those of our selected LPM (see Table 3.A.5, column 1).

where most notation is the same as above. Additionally δ_s represents the fraction of children in a school excluded from our analysis, and Z_{is} is a vector of covariates including household size, assets, time preferences and risk preferences of child i . Standard errors are still clustered for 42 schools.

The additional covariates included in vector Z_{is} are potentially endogenous to our model. Since, they do not alter the main result (see Table 3.7, column 6), nor is their correlation with the probability to contribute to the public good statistically significant (see Table 3.A.6), we exclude them from any further analysis.

Throughout several specifications – controlling for an increasingly extensive set of covariates – the main result remains stable: One standard deviation drop in the digit ratio decreases the child’s probability of contribution to the public good by around 7% points, or around 17% of the mean prevalence.

Result 3: Prenatal stress reduces the taste for cooperation. One standard deviation drop in the digit ratio decreases the child’s probability of contribution to the public good by 17%

3.5.4 Alternative mechanisms

Next, we investigate potential alternative mechanisms. First, Section 3.2 highlights the role of postnatal war exposure in shaping individual preferences. The children in our sample were at most 8 years of age at the end of hostilities, but postnatal witnessing of conflict-related violence may still have affected their taste for cooperation (Bauer et al., 2014). Second, the literature discussed in Section 3.2 predicts that prenatal stress may capture the effect of early life deprivation. Height, for instance, is an anthropometric indicator of early-life experiences comparable to longitudinal measures such as height and weight at birth (Currie and Vogl, 2013). Similarly, low birth weight is associated with later-life low BMI (Walker et al., 2002), and severe deprivation at an early stage has persistent effects on cognitive ability (Beckett et al., 2006; Figlio et al., 2014). Both these postnatal mechanisms could be correlated with both prenatal stress, and possibly

Table 3.7 – Prenatal stress reduces cooperation

	Public Good Contribution					
	(1)	(2)	(3)	(4)	(5)	(6)
Standardized digit ratio (-)	-0.078*** (0.024)	-0.072*** (0.025)	-0.070*** (0.025)	-0.070*** (0.025)	-0.070*** (0.025)	-0.072*** (0.025)
Caregiver public good contribution	0.144*** (0.041)	0.144*** (0.041)	0.141*** (0.041)	0.145*** (0.045)	0.145*** (0.047)	0.146*** (0.047)
Caregiver risk averse		0.001 (0.054)	-0.002 (0.055)	0.012 (0.056)	0.014 (0.055)	0.013 (0.055)
Child controls	N	N	Y	Y	Y	Y
Caregiver and peer controls	N	N	N	Y	Y	Y
Sub-county FE	N	N	N	N	Y	Y
Child risk and time preferences	N	N	N	N	N	Y
Household size and assets index	N	N	N	N	N	Y
Observations	442	442	442	440	440	440
R ²	0.025	0.046	0.050	0.075	0.078	0.080

Notes: Standard errors corrected for school level clustering (42) are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child controls: Age, Female, Age \times Female; Caregiver controls: Age, Female, Age \times Female, Acholi, Education; Peer controls: Proportion of excluded peers, Class dummies. See Table 3.A.6 for the full list of coefficients.

public good contributions. This could therefore result in a spurious correlation between the 2D:4D digit ratio and cooperation.

We find that in our sample postnatal war exposure does not significantly affect cooperation when controlling for child, caregiver and peer characteristics (Table 3.8, columns 1-2), and does not wash out the effect of prenatal stress. On the other hand, we find that height-for-age is positively associated with cooperation: one standard deviation increase in the height-for-age increases the likelihood of contribution to the public good by between 3% points and 6% points in the simple and fully specified models respectively. BMI-for-age and IQ-for-age do not enter significantly (Table 3.8, columns 3-4). Nonetheless, prenatal stress remains significant even when controlling for these potential alternative mechanisms together: its coefficient is stable and robust to controlling for markers of early life deprivation (column 5).

Table 3.8 – Alternative mechanisms: postnatal war exposure and early life deprivation

	Public Good Contribution				
	(1)	(2)	(3)	(4)	(5)
Standardized digit ratio (-)	-0.076*** (0.024)	-0.075*** (0.025)	-0.077*** (0.025)	-0.073*** (0.026)	-0.072*** (0.026)
Postnatal war exposure	0.024 (0.021)	-0.005 (0.031)			0.007 (0.032)
Height-for-age			0.025* (0.015)	0.058*** (0.020)	0.058*** (0.020)
BMI-for-age			-0.015 (0.025)	-0.017 (0.024)	-0.018 (0.024)
IQ-for-age			-0.012 (0.021)	-0.010 (0.024)	-0.009 (0.025)
Child controls	N	Y	N	Y	Y
Caregiver and peer controls	N	Y	N	Y	Y
Sub-county FE	N	Y	N	Y	Y
Observations	440	438	440	438	436
R^2	0.027	0.057	0.030	0.073	0.074

Notes: Standard errors corrected for school level clustering (42) are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***. Child controls: Age, Female, Age×Female; Caregiver controls: Age, Female, Age×Female, Acholi, Education; Peer controls: Proportion of excluded peers, Class dummies.

Result 4: The relationship between prenatal stress and cooperation is stable and robust to controlling for early life deprivation markers and postnatal war exposure.

On top of this, we conduct a final series of robustness checks. The main result holds when we limit the sample to only blood related caregiver-child pairs, to parent-child pairs, and to mother-child pairs (see Table 3.A.7). The result also holds when estimated using a probit model, when including school and enumerator fixed effects, and when clustering standard errors on both school and year-of-birth levels (see Table 3.A.5, columns 1-4). We also split our sample at the median of the digit ratio, finding that caregivers of children with low 2D:4D ratio (high prenatal stress) are not only more likely to suffer from PTSD, but also less likely to contribute to the public good (see Table 3.A.8). The difference in PTSD rates is expected, but the difference in cooperation rates could introduce a selection bias into our results. To account for any such observable bias, we match participants using kernel matching over all our child and caregiver control characteristics, including caregiver cooperation, and perform a propensity score matching analysis on the common support area. The resulting average treatment effect is again robustly significant (see Table 3.A.5, column 5).

3.5.5 Causality and unobserved selection

If certain types of caregivers – with a certain preference for cooperation or risk before the war – self-selected into or out of traumatic events, they may have passed their preferences on to the children through genetic and environmental channels. Even though we control for the preferences of caregivers, this may not suffice to net out the selection bias. One way to test whether the patterns found in Table 3.A.8 are explained mostly by endogenous preference formation or by selection bias is to look at the relationship between the caregiver’s digit ratio and their traumatization. Table 3.9 shows that, while caregiver war exposure is a strong predictor of their traumatization,²³ caregiver digit ratios are not. In other words,

²³We find a robustly positive and significant relationship between war exposure and trauma among caregivers, with a higher prevalence of PTSD among women. A one standard deviation

while Table 3.5 shows that caregiver traumatization is negatively correlated with child digit ratios, caregiver digit ratios are uncorrelated with their likelihood of being traumatized.

Table 3.9 – Conflict exposure predicts caregiver trauma, caregiver digit ratio does not

	PTSD			Trauma factor		
	(1)	(2)	(3)	(4)	(5)	(6)
War exposure	0.0815*** (0.0221)	0.102*** (0.0225)	0.138*** (0.0454)	0.182*** (0.0527)	0.230*** (0.0492)	0.271*** (0.0960)
Female		0.284*** (0.0509)	0.332*** (0.0859)		0.563*** (0.106)	0.769*** (0.156)
Caregiver digit ratio			-0.0855 (1.184)			0.906 (2.220)
Caregiver controls	N	Y	Y	N	Y	Y
Sub-county FE	N	Y	Y	N	Y	Y
Observations	440	438	154	440	438	154
R^2	0.028	0.105	0.157	0.033	0.129	0.229

Notes: * $p < 0.10$, ** $p < 0.05$, ***. Caregiver controls: Age, Female, Age×Female, Acholi, Education.

Again, this does not allow us to exclude all forms of self-selection. We are unable to conclusively verify the randomness of the behavior of armed groups, or the influence of other pre-war characteristics. Our causal interpretation of the results may thus suffer from potential bias due to the omission of such unobservables. Following Blattman and Annan (2010), we provide a graphical benchmark of the sensibility of our results to exogeneity assumptions (Imbens, 2003; Harada, 2013). The curve in Figure 3.4 represents the locus of partial correlation points of a hypothetical pseudo-unobservable with our assignment and outcome variables, that would reduce estimated effect below significance at the 5% level.²⁴ The selective contour is conservative, as it only reduces the coefficient size by about one quarter.²⁵ Yet, the potentially confounding mechanisms we identify and discuss above (i.e. caregiver cooperation preferences, early life deprivation, and postnatal

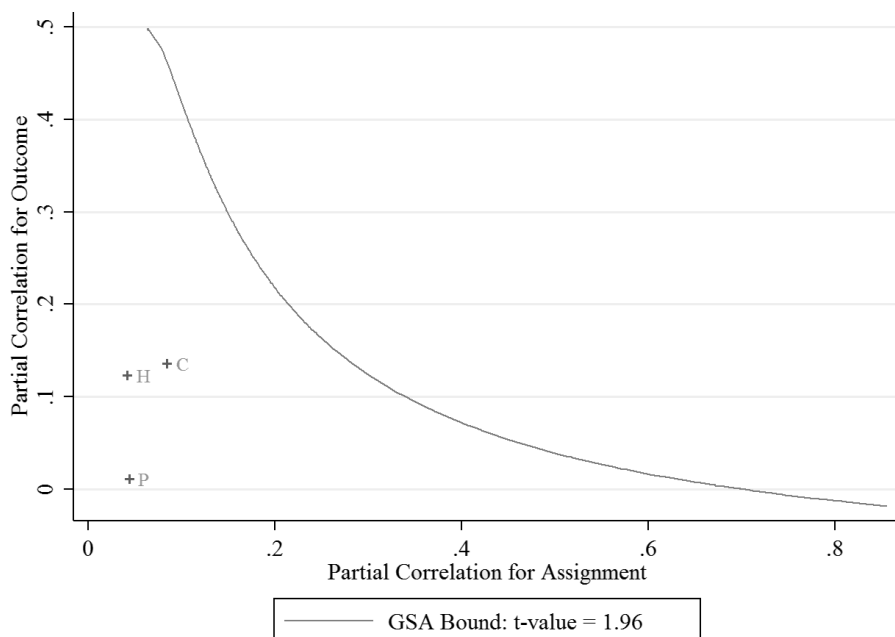
increase in our war exposure index increases the likelihood of PTSD by almost 10%, and scores on the trauma factor by 20% of a standard deviation.

²⁴H: Height-for-age; C: Caregiver cooperation; P: Postnatal war exposure

²⁵Blattman and Annan (2010), for instance, plot a contour that decreases the effect of the assignment by one half. This would be equivalent to shifting the contour in Figure 3.4 outwards.

war exposure) lie far below the selected threshold. To rule out the effect of prenatal stress, any unobserved covariate not considered in our analysis would require a partial correlation with both the treatment and the assignment well above the curve of Figure 3.4.

Figure 3.4 – Sensitivity to the unconfoundedness assumption



3.6 Conclusions

“The womb may be more important than the home”, wrote the late David J. Barker (1990) in his seminal work on the fetal origins of adult disease. Barker’s hypothesis has spawned a large volume of literature exploring its economic implications. This study builds upon the fetal origins literature, and tests the hypothesis that prenatal events may not only alter later-life individual abilities and health trajectories

(Almond and Currie, 2011), but also social preferences. In particular, we look at the impact of prenatal stress on the preferences for cooperation among children born during an armed conflict. We play a dichotomous one-shot public goods game in Pader, a district in Northern Uganda, with children born during the 1998-2006 period of intense fighting between government forces and the Lord's Resistance Army (LRA). Our identification strategy exploits variations in the 2D:4D digit ratio—a marker of in utero hormone exposure negatively associated with high maternal distress during early fetal development. It also exploits the assumption that war exposure and traumatization were conditionally unconfounded in our context (Doom and Vlassenroot, 1999; Jackson, 2002; Blattman, 2009; Blattman and Annan, 2010). We find that a rise in our marker of prenatal distress robustly reduces the child's probability of contribution to the public good. The estimated effect is quantitatively large, stable, and robust to controlling for alternative mechanisms such as early life deprivation, caregiver public good contribution preferences, and war exposure after birth. We also show that it is robust to relaxing the assumption of unconfoundedness and self-selection. In other words, we show that preferences are endogenously determined right from the womb.

Our results thus support three separate findings from previous studies. Firstly – and perhaps obviously – violent conflict exposure is traumatizing. Secondly, a mother's traumatization during pregnancy affects the hormonal balance of the fetus as observable through its 2D:4D digit ratio. Thirdly – and most importantly – in utero hormonal balance affects later-life social preferences. By analyzing these three relationships concurrently in a post-conflict context – where violence has differentially impacted large portions of the population – we find evidence supporting the entire causal chain: from conflict in one generation to economic behavior in the next one. Prenatal stress triggers adaptive mechanisms that go far beyond the well-established relationship between postnatal war exposure and preferences. The socio-economic consequences of conflict may thus be reaching much further than previously thought, and the womb may well be far more crucial than Barker ever imagined.

Appendix 3.A: Tables

Table 3.A.1 – Spatial autocorrelation of war exposure and trauma

	War exposure	PTSD
	(1)	(2)
Moran's I	-0.045	-0.030
$E(I)$	-0.024	-0.024
$sd(I)$	0.056	0.055
z	-0.364	-0.099
p -value	0.358	0.461
Observations	42	42

Notes: p -values are based on a 1-tail test.

Table 3.A.2 – Intraclass correlation of war exposure and trauma

	War exposure	PTSD
	(1)	(2)
ρ	0.043	0.020
Observations	440	442
Groups	42	42

Notes: ρ is the fraction of variance due to school-level random effects from a GLS regression.

Table 3.A.3 – Sensitivity of results to the exclusion of potentially biased measures

	Public Good Contribution			
	All	$\Delta < 5\%$	$\Delta < 3\%$	$\Delta < 1\%$
	(1)	(2)	(3)	(4)
Standardized digit ratio (-)	-0.078*** (0.025)	-0.108*** (0.037)	-0.100** (0.044)	0.099 (0.280)
Child controls	Y	Y	Y	Y
Caregiver and peer controls	Y	Y	Y	Y
Sub-county FE	Y	Y	Y	Y
Observations	440	189	131	39
R^2	0.046	0.082	0.115	0.409

Notes: * $p < 0.10$, ** $p < 0.05$, ***. Child controls: Age, Female, Age \times Female; Caregiver controls: Age, Female, Age \times Female, Acholi, Education; Peer controls: Proportion of excluded peers, Class dummies.

Table 3.A.4 – Robustness of the main result to controlling for parental digit ratios

	Public Good Contribution			
	(1)	(2)	(3)	(4)
Standardized digit ratio (-)			-0.139*** (0.036)	-0.125*** (0.040)
Caregiver digit ratio (biological parents only)	-0.048 (0.041)	-0.077* (0.041)	-0.060 (0.037)	-0.079* (0.040)
Child controls	N	Y	N	Y
Caregiver and peer controls	N	Y	N	Y
Sub-county FE	N	Y	N	Y
Observations	154	154	154	154
R^2	0.010	0.123	0.065	0.162

Notes: * $p < 0.10$, ** $p < 0.05$, ***. Child controls: Age, Female, Age×Female; Caregiver controls: Age, Female, Age×Female, Acholi, Education; Peer controls: Proportion of excluded peers, Class dummies.

Table 3.A.5 – Robustness of the main result to alternative specifications

	Public Good Contribution				
	Probit model	School FE	Enumerator FE	Two-way clustering	Kernel PSM
	(1)	(2)	(3)	(4)	(5)
Standardized digit ratio (-)	-0.083*** (0.028)	-0.065*** (0.023)	-0.052** (0.023)	-0.076*** (0.025)	-0.111** (0.052)
Child controls	Y	Y	Y	Y	Y
Caregiver and peer controls	Y	Y	Y	Y	Y
Sub-county FE	Y	N	Y	Y	Y
School level clustered SE	42	N	42	42	N
Year of birth clustered SE	N	N	N	9	N
Observations	440	440	440	440	425
R^2	0.043	0.137	0.172	0.057	

Notes: * $p < 0.10$, ** $p < 0.05$, ***. Child controls: Age, Female, Age × Female; Caregiver controls: Age, Female, Age×Female, Acholi, Education; Peer controls: Proportion of excluded peers, Class dummies. Standard errors in (5) are bootstrapped with 100 iterations.

Table 3.A.6 – Prenatal stress reduces cooperation (all coefficients)

	Public Good Contribution				
	(1)	(2)	(3)	(4)	(5)
Standardized digit ratio (-)	-0.072*** (0.025)	-0.070*** (0.025)	-0.070*** (0.025)	-0.070*** (0.025)	-0.072*** (0.025)
Caregiver PG contribution	0.144*** (0.041)	0.141*** (0.041)	0.145*** (0.045)	0.145*** (0.047)	0.146*** (0.047)
Caregiver risk averse	0.001 (0.054)	-0.002 (0.055)	0.012 (0.056)	0.014 (0.055)	0.013 (0.055)
Child age		0.003 (0.015)	0.020 (0.020)	0.018 (0.020)	0.018 (0.020)
Child female		-0.178 (0.243)	-0.185 (0.246)	-0.191 (0.256)	-0.202 (0.258)
Child age × female		0.016 (0.021)	0.017 (0.021)	0.017 (0.022)	0.018 (0.022)
Caregiver age			-0.004 (0.003)	-0.004 (0.003)	-0.004 (0.003)
Caregiver female			-0.306* (0.174)	-0.303* (0.173)	-0.311* (0.174)
Caregiver age × female			0.009** (0.004)	0.009** (0.004)	0.009** (0.004)
Caregiver Acholi			0.010 (0.119)	-0.004 (0.142)	-0.000 (0.148)
Caregiver education			0.015 (0.037)	0.016 (0.038)	0.013 (0.039)
Proportion of excluded peers			-0.466* (0.272)	-0.508 (0.330)	-0.514 (0.342)
P2			0.124 (0.078)	0.121 (0.078)	0.122 (0.078)
P6			0.017 (0.062)	0.020 (0.063)	0.017 (0.064)
Lira Palwo sub-county				-0.036 (0.076)	-0.045 (0.079)
Lukole sub-county				-0.034 (0.077)	-0.039 (0.079)
Parabongo sub-county				-0.032 (0.083)	-0.040 (0.085)
Patongo sub-county				-0.069 (0.139)	-0.072 (0.137)
Awere sub-county				-0.093 (0.073)	-0.096 (0.078)
Pajule sub-county				0.018 (0.101)	0.009 (0.102)
Puranga sub-county				-0.025 (0.087)	-0.026 (0.092)
Child risk preferences					0.007 (0.046)
Child time preferences					-0.016 (0.057)
Household size					0.006 (0.009)
Asset index					0.008 (0.020)
Observations	442	442	440	440	440
R ²	0.046	0.050	0.075	0.078	0.080

Notes: Standard errors corrected for school level clustering (42) are in parentheses. * $p < 0.10$, ** $p < 0.05$, ***. Atanga sub-county is taken as reference category, and therefore omitted.

Table 3.A.7 – Sensitivity of results to caregiver relationship to the child

	Public Good Contribution			
	All	Blood-related	Parents	Mothers
	(1)	(2)	(3)	(4)
Standardized digit ratio (-)	-0.076*** (0.025)	-0.075*** (0.025)	-0.081*** (0.028)	-0.089** (0.033)
Child controls	Y	Y	Y	Y
Caregiver and peer controls	Y	Y	Y	Y
Sub-county FE	Y	Y	Y	Y
Observations	440	438	351	189
R^2	0.057	0.056	0.053	0.105

Notes: Standard errors corrected for school level clustering (42) are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Child controls: Age, Female, Age×Female; Caregiver controls: Age, Female (excluding column 4), Age×Female (excluding column 4), Acholi, Education; Peer controls: Proportion of excluded peers, Class dummies.

Table 3.A.8 – Split-sample balance (caregivers)

Variable	High child 2D:4D (Low prenatal stress)		Low child 2D:4D (High prenatal stress)		Diff.	Std. err.
	Obs.	Mean	Obs.	Mean		
2D:4D	76	0.957	78	0.956	-0.001	0.006
Public good contribution	221	0.561	221	0.452	-0.109**	0.047
Age	220	42.3	220	41.8	-0.5	1.1
Female	221	0.489	221	0.579	0.090*	0.047
Education level	221	0.864	221	0.814	-0.050	0.076
Risk preferences	221	0.299	221	0.326	0.027	0.044
War exposure	220	0.746	220	0.758	0.012	0.016
PTSD (dummy)	221	0.317	221	0.466	0.149***	0.046
PTSD (factor)	221	-0.142	221	0.142	0.284***	0.094
Christian	221	0.977	221	0.968	-0.009	0.015
Acholi	221	0.995	221	0.986	-0.009	0.009
Household size	221	7.92	221	8.26	0.34	0.28
Asset index	221	-0.028	221	0.028	0.056	0.095

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 3.B: Figures

Figure 3.B.1 – Average war exposure per school

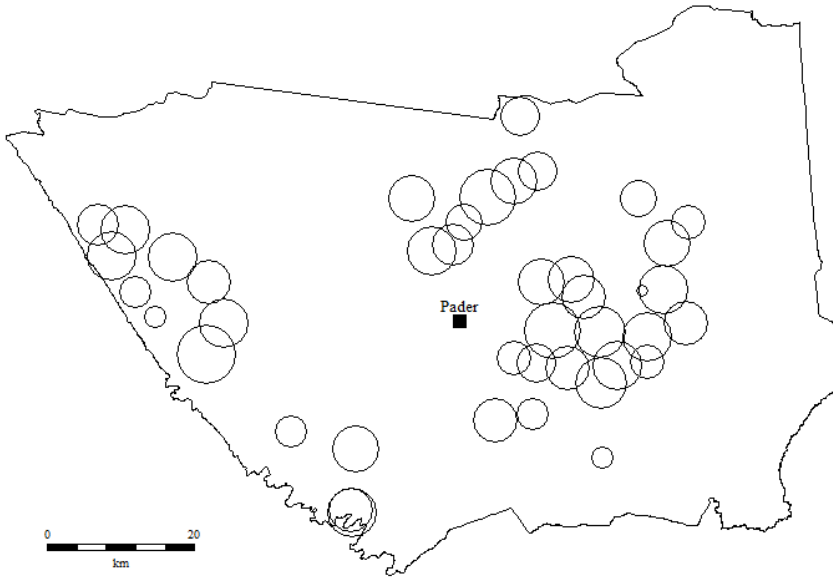
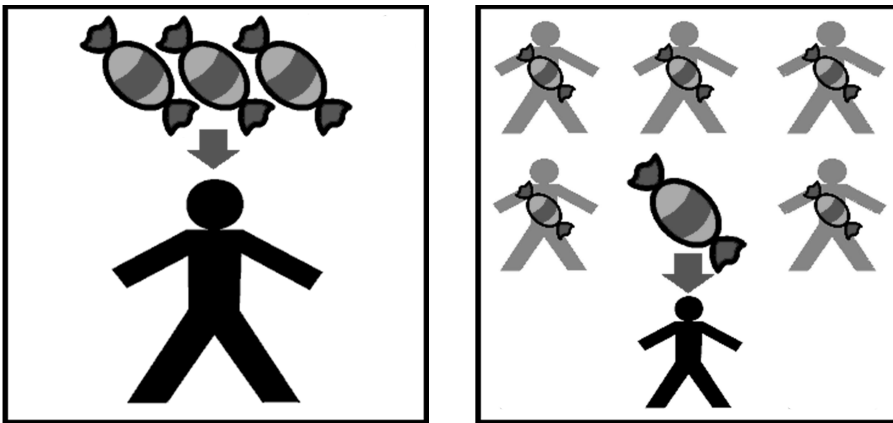


Figure 3.B.2 – Private and group cards in the public goods game for children



Appendix 3.C: Variable definitions

2D:4D. A child level measure of the relative length of the index finger of the right hand with respect to the ring finger (in cm).

Standardized digit ratio (-). A child level variable derived by standardizing the 2D:4D (z -score). The negative sign is added for ease of interpretation.

Public good contribution (child and caregiver). An individual level dummy for both child and caregiver representing the individual choice in the public goods game: “group card” or “private card”. The choice “group card” takes value 1, 0 otherwise.

Age (child and caregiver). Age of respondent (child and caregiver) i in years, rounded down to the last birthday.

Female (child and caregiver). An individual level dummy taking value of 1 if respondent (child and caregiver) is female, 0 otherwise.

Class dummies (P2, P4, P6). Child-level dummies indicating the current school grade of child i .

Height-for-age. The height of child i standardized for his age class (de Onis et al., 2007).

BMI-for-age. The body-mass-index of child i standardized for his age class (de Onis et al., 2007).

IQ-for-age. The IQ of child i , measured using standard Raven’s matrices and standardized for his age class (in sample).

Time preferences. A child level dummy taking value of 1 if child prefers to receive two candies at the end of the survey rather than one half way, 0 otherwise.

Risk preferences (child). A dummy equal to 1 if the child consistently chooses for the riskier option in a two dichotomous lottery choices; 0 otherwise.

Risk preferences (caregiver). A dummy equal to 0 if the caregiver chose the least risky one of several lottery options; 1 otherwise.

War exposure (child and caregiver). A caregiver level victimization index derived from answers to 23 war witnessing questions (see Macksoud, 1992). The postnatal war exposure of child i is proxied by the war exposure of the caregiver multiplied by the fraction of violent conflict events that took place after the birth of child i .

Education level. A caregiver level variable indicating the number of completed years of education.

PTSD. A caregiver level measure of post-traumatic stress disorder, calculated using the PCL-civilian checklist.

Christian. A dummy taking value of 1 if the caregiver is Christian by religion, 0 otherwise.

Acholi. A dummy taking value of 1 if the caregiver is ethnically Acholi, 0 otherwise.

Household size. The number of people sharing the same roof and sharing the same pot.

Asset index. A principal factor (see Sahn and Stifel, 2003) of assets possessed by the caregiver's household.

Chapter 4

Institutional Change and the Dynamics of Social Capital

Experimental Evidence from Uganda

Abstract: We explore how the introduction of formal insurance affects the within-village dynamics of social capital in south-western Uganda. As a proxy for social capital we use contributions in a public goods game. Consistent with existing evidence, our data suggests that formal insurance crowds out social capital. However, and surprisingly, overall contributions to the common pot do not go down because insurance adopters reduce their contributions. Instead, social capital erodes because the uninsured, or the local poor, lower their contributions. We discuss what might explain this finding.

This chapter is based on:

Cecchi, F., J. Duchoslav, and E. Bulte (2016). Formal Insurance and the Dynamics of Social Capital: Experimental Evidence from Uganda. *Journal of African Economies* 25(3), 418–438.

4.1 Introduction

Institutions are constraints devised by humans to shape human interaction (North, 1990, p. 3). This broad definition captures both formal and informal institutions, where the former are often treated as exogenous constraints, enforced by an outside party (possibly “the state”), and the latter are endogenous constraints—self-enforcing rules representing the sub-game perfect equilibrium of a repeated game (e.g. Greif, 1993; Aoki, 2001a,b). While recent empirical work has identified institutions as a key determinant of economic performance (e.g. Rodrik et al., 2004), much remains unknown about how institutions evolve over time, or about the interaction between different (types of) institutions.

In this article, we ask how the introduction of a new formal institution affects the performance of a preexisting informal one. This topic is gaining importance as, globally, systems of formal institutions are expanding. For example, global value chains are penetrating further into societies hitherto oriented towards subsistence activities; the formalization of land rights increasingly affects customary institutional tenure arrangements; and the expanding reach of the state and formal court system is altering informal judicial institutions. The current wave of experimentation with (index) insurance products in environments characterized by informal sharing arrangements represents another example. While some theoretical work exists to analyze how (exogenous) changes in formal institutions affect the equilibrium of repeated games (e.g. Aoki, 2001b), empirical work in this domain remains very scarce. Roland (2004) describes the interaction between slow- and fast-moving institutions and argues that, depending on the context, specific institutions may be complements or substitutes (see also Williamson, 2009; Alesina and Giuliano, 2015). Institutional innovations in one domain may therefore crowd out, or solidify, existing institutional arrangements.¹

In this chapter, we seek to enhance our understanding of the impact of a specific formal institutional innovation on local cooperation (“social capital”) using experimental methods. We focus on the provision of insurance—a domain

¹For recent experimental work along these lines, refer to Chandrasekhar et al. (2016).

relatively well-studied by economists. Rural producers in developing countries are exposed to various shocks and typically are better off when pooling their risks (in particular when the co-variation of risks across individuals is modest or absent). Since, until recently, prohibitive transaction costs typically precluded the writing of formal insurance contracts, rural producers by and large depended on informal insurance arrangements to secure their livelihoods (e.g. Townsend, 1994; Udry, 1994; De Weerd and Fafchamps, 2011). Such sharing could take various forms, including redistribution within friendship or kinship networks (e.g. Fafchamps and Gubert, 2007; Alger and Weibull, 2010), or transfers in patron-client relationships (Richards, 1996). In recent years, however, various agencies have experimented with the provision of formal insurance for rural households, based on written contracts and possibly outside enforcement. This development was facilitated by improvements in communication technology, which lowered transaction costs, but also by the creation of “new” insurance products such as index insurance.

Insofar as formal and informal insurance are “substitutes”, one might expect that the introduction of formal insurance possibilities will crowd out informal insurance arrangements. A small literature seeks to empirically test this hypothesis, providing some support for it. Dercon and Krishnan (2003) find that public transfers in the form of food aid crowd out informal sharing in rural Ethiopia.² Landmann et al. (2012) use data from the Philippines to show that formal insurance lowers voluntary transfers among members in social networks. Klohn and Strupat (2013) examine the link between the provision of formal health insurance and informal transfers in Ghana and find that formal insurance reduces both the probability of making transfers and amounts transferred. Lin et al. (2014) confirm the crowding out thesis in a laboratory setting.³

The main objective of this chapter is to empirically explore how the introduction

²Also see Bähre (2011), who examines the relation between formal financial arrangements and personal networks in post-apartheid South Africa, finding that increasing redistribution created frictions within networks.

³A related literature concerns the effects of migration (remittances) as a mechanism to provide insurance. For example, Morten (2016) establishes that temporary migration decreases informal risk sharing.

of formal health insurance affects within-village cooperation and social capital in a sample of Ugandan villages. We first compare the behavior of villagers with and without access to formal health insurance and obtain an intention-to-treat effect at the village level. We proceed by disentangling behavior of adopters and non-adopters. Unlike existing work, we consider a voluntary insurance program with imperfect uptake—allowing us to probe intra-village implications. Moreover, we do not use informal transfers as our measure of social capital. Instead, our main dependent variable is based on behavior in public goods (PG) games. The PG game captures the ability of communities to coordinate on first-best outcomes and represents a measure of social capital at the village level (e.g. Fearon et al., 2009; Voors et al., 2011). Following Klohn and Strupat (2013) and Morten (2016), one may hypothesise that formally insured village members depend less on their fellow villagers to sustain their livelihoods and, as a consequence, invest less in social relations.⁴ If so, the result will be erosion of social capital, translating into lower contributions to the local public good.

In areas with access to formal insurance, we find that uptake is far from uniform and skewed towards the sub-sample of wealthier villagers. We also find that average contributions to the public good are lower in areas with access to formal insurance. These results corroborate simple game theoretical predictions (outlined below) and are consistent with the empirical findings mentioned above. Next, we zoom in on within-village social capital dynamics, documenting patterns in the data that are harder to explain. Lower aggregate contributions to the public good are not due to the withdrawal of insurance adopters. Instead, reduced average contributions are fully explained by declining contributions of non-adopters. Moreover, we find that non-adopters only contribute less when matched with adopters. Following sociological theories of social embeddedness (Granovetter, 1985) and the anthropological “weapons of the weak” perspective formulated by Scott (1985), we speculate that sending behavior by non-adopters

⁴Of course there are other reasons why the introduction of formal insurance (or storage) possibilities may cause adopters to opt out of informal institutions or networks. Klohn and Strupat (2013) discuss that formal institutions may crowd out altruistic behavior (see Bowles, 2008) or may reduce (the bite of) social sanctions associated with exit (see Grimm et al., 2017).

serves as a signal. Specifically, our empirical results are consistent with the story that non-adopters signal their discontent caused by the feeling of being left behind. If the wealthy adopt formal insurance, they may opt out of informal sharing networks in the future, leaving non-adopters worse off. However, we cannot rule out several alternative explanations, so the exact mechanism linking non-adoption to low contributions should be explored more fully in future research.

The rest of this chapter is organized as follows. In Section 4.2, we outline and illustrate the conventional economic perspective on the evolution of informal institutions in response to changes in the environment. In Section 4.3, we provide context regarding the case study and describe the insurance intervention as implemented by a local NGO. In Section 4.4, we summarize our data, outline our identification strategy and formulate four research questions. Section 4.5 contains our empirical results, which we speculatively seek to interpret in Section 4.6. The conclusions ensue.

4.2 The interaction between formal and informal institutions

We outline a simple economic perspective on the interaction between formal and informal institutions—an issue fitting in an emerging literature on the co-evolution of institutions and culture (see Alesina and Giuliano, 2015, for a recent overview). The economic perspective lends itself naturally to game-theoretic analysis, and economists have developed a coherent framework in which informal institutions are seen as the equilibrium outcomes of repeated interactions—outcomes that are persistent over time, shaping expectations and gradually evolving into norms of appropriate behavior. See, for example, Aoki (2001b) for a treatment of institutions as endogenous and self-enforcing equilibrium outcomes of a repeated game.⁵

⁵We emphasize that non-economists have also contributed important ideas to the issue of co-evolving institutions and informal (cultural, traditional) norms and arrangements. For example, and focusing on the interaction between traditional society and “the state”, the sociologist Galvan (2004) describes a process referred to as “institutional syncretism”. This

Consider the case of informal insurance via sharing as an example. Economists have studied the evolution of such arrangements as alternative insurance opportunities emerge. Informal insurance via gifts and transfers occurs within networks of family members (or friends) because of altruism and may involve support to deal with persistent shocks (chronic illness or disability— see De Weerdt and Fafchamps, 2011). Informal insurance may also be motivated by expected reciprocity among individuals (or households) in a context of repeated interaction (e.g. Kimball, 1988; Coate and Ravallion, 1993; Ligon et al., 2002; De Weerdt and Fafchamps, 2011). The theoretical literature has emphasized that such non-altruistic sharing arrangements should be self-enforcing; individuals are willing to help others facing a temporary shock because of the credible promise of reciprocity in the future. Participation constraints limit the extent of risk sharing that is possible, as villagers balance the short-term benefit from renegeing on sharing obligations against the loss of foregoing the (expected) potential benefits from sharing in all future periods.

How does the emergence of formal insurance affect a binding participation constraint? An individual who is expelled from the insurance network now has the option to adopt formal insurance—paying a fixed fee every period and receiving a payment in case of a health shock. Because an actuarially fair insurance product improves the autarky outcome of adopters in the future, the set of self-enforcing informal insurance transfers decreases. Intuitively, because individuals have access to a substitute insurance product tomorrow (other than transfers from peers), it is more tempting to renege on obligations today and opt out of the network. Only a menu of small transfers can now be supported as the new equilibrium outcome, and insurance options for non-adopters unambiguously deteriorate (see Coate and Ravallion, 1993; Ligon et al., 2002).

The main message is that substitute insurance mechanisms lower the value attached to informal sharing arrangements (for those able to adopt the substitute). Substitutes could be formal insurance, as above, but alternatives exist. For

involves the reworking of state-imposed institutions from below by peasants to make these new (and sometimes alien) institutions consistent with tradition.

example, Ligon et al. (2002) focus on the implications of self-insurance via storage, which also increases the value of autarky. As above, individuals who were previously (almost) indifferent between participating in the sharing network and autarky will now renege when they have the option to store, and subsequently opt out of the network. While their welfare improves as a result, the utility of individuals remaining in the network decreases unambiguously. They are worse off because the network shrinks and loses part of its ability to absorb shocks. Ligon et al. (2002) demonstrate that introducing the possibility of storage may even reduce overall welfare. Migration-induced remittances may provide yet an alternative substitute mechanism to protect households from health shocks, and Morten (2016) indeed demonstrates that migration tends to decrease insurance provided via informal networks.

Theory does not necessarily suggest that the provision of formal insurance will always crowd out informal sharing. For example, Dercon et al. (2014) demonstrate that informal (within-group) sharing and index insurance are complements, rather than substitutes—so that the provision of insurance crowds in, rather than crowds out, informal sharing arrangements. The reason is that the basis risk implied by index insurance creates demand for additional informal insurance. Dercon et al. (2014) argue that “the benefits of encouraging groups to combine these products with increased risk sharing among members might be large, resulting in substantially larger take-up rates”.⁶ Nevertheless, in the context of idiosyncratic risk and indemnity-based payments, such as the context we study below, the economic perspective proposes that the introduction of formal insurance crowds out informal insurance.

Importantly, informal insurance is only one specific dimension of social capital. In reality, villagers will typically interact in different social spheres and have to address a range of social dilemmas. The ability of groups of subjects to overcome social dilemmas provides a measure of that group’s social capital. In the words of sociologist Granovetter (1985), economic transactions are “socially embedded”.

⁶de Janvry et al. (2014) also argue in favor of selling insurance to groups rather than individuals, to overcome coordination failures and to internalize positive externalities.

Aoki (2001a) writes that “the economic transaction domain is embedded in a social exchange domain in which the same members repeatedly interact socially and invest in, and enjoy returns from, social capital”. This perspective can be worked out formally using game theory, by linking multiple games and expanding the payoff structure. For example, Aoki (2001a,b) demonstrates that linking an economic exchange and social interaction game may expand the set of equilibria that can be supported in the exchange game.⁷ If so, the reverse is also true—de-linking institutions should contract the set of feasible equilibria in other domains of exchange. Specifically, societies will be less able to overcome social dilemmas (e.g. provide public goods) after substituting informal insurance between villagers by formal insurance provided by an outside agent. Partial formalization of exchange in societies may curtail cooperation in informal domains of interaction.

In what follows we test whether formal insurance affects voluntary contributions of villagers to the local public good—distinguishing between those adopting insurance and those not adopting insurance. To empirically probe this issue, we organized behavioral games in a sample of rural Ugandan communities.

4.3 The case study: health insurance in Uganda

Approximately half the population in Uganda is below 14 years of age, and this population is expected to double in size in the next 20 years (UN DESA, 2012). To help sustain strong post-war economic recovery during the last decade, Uganda focuses on health interventions and the accumulation of human capital (World Bank, 2011).

We focus on Kitagata sub-county, located in the South-Western Sub-region of Uganda. A local not-for-profit organization called Save for Health Uganda (SHU)

⁷Aoki (2001a,b) writes that interlinkages between institutions may be “understood in terms of a general-equilibrium-like feedback among the strategies of individual agents across different domains of the economy. Then an economy as a whole may be viewed as a robust and coherent overall cluster of mutually interdependent institutions”.

has recently implemented a health micro-insurance project across four parishes in Kitagata, with the ultimate mid-term goal of covering the entire sub-county.⁸ In what follows, we refer to households from these parishes as having “access” to formal health insurance—i.e. in these parishes households actively decided whether to adopt insurance, or not. Each village in the covered parish is encouraged and facilitated to create a so-called Community Health Financing (CHF) scheme, which provides health insurance to member families. Members receive an insurance card, allowing access to all services provided by contracted facilities.⁹ These services include transport to the hospital in case of delivery, ante-natal care, out-patient services, admission services and surgery (SHU, 2013).

Although there was slight variation with respect to the starting date of village level schemes, SHU required villages within the same parish to begin simultaneously, ensuring relatively homogeneous up-take rates. When we collected our data in August 2012, all the randomly selected villages from areas with access to the insurance had a running scheme, to which approximately 53% of the families had subscribed. The great majority of the villagers in access areas (some 95%) indicated to be familiar with CHF schemes. In parishes without access to the formal insurance, around 56% of the people interviewed had heard about the schemes, but were aware that it was not (yet) available to them.¹⁰ On average, member households were expected to pay just above 26,000 USh (or 10 USD) to cover the yearly insurance premium, with some variation depending on village and family size. Premiums were not sufficient to cover insurance costs, which are subsidized by international donors and sponsors. Nonetheless, only 18% of the participating families had been able to pay the full yearly premium to that date, with average payment rates hovering around 34%.¹¹ Families that have not fully

⁸Since then, Kitagata Sub-county has been administratively divided into two sub-counties: Kitagata and Kasaana. SHU, however, keeps working in both sub-counties with a single program.

⁹Contracted facilities include one private hospital and two public healthcare facilities.

¹⁰Notably, SHU had already undertaken informational and sensitization meetings in two parishes in no access areas by the time of our fieldwork, and while the scheme was neither active nor running, the insurance was scheduled to start there in the near future.

¹¹Dekker and Wilms (2010) find that participants in a similar private health-insurance scheme in five rural and two urban communities in Uganda face comparable difficulties in paying the premium: only 37% of participating households were able to pay.

paid the premium do not yet qualify for compensation; so many families currently have one foot in the formal system but also need to continue investing in informal arrangements.

We believe that there is considerable demand for formal health insurance schemes, but that paying for the premium remains a challenge for many families. The latter is confirmed by many open-ended exit interviews after the public goods game. While we did not collect hard evidence on pre-existing informal arrangements of mutual insurance, anecdotes complement the (qualitative) literature on local informal arrangements (Taylor et al., 1996; McDonald et al., 1999) and stress the importance of relatives, neighbors and friends in providing financial support in times of hardship.¹² We interpret this context as one of being in institutional flux, in which a considerable share of the local population appears “in limbo” between alternative insurance modalities.

4.4 Experimental identification, data and design

The health insurance scheme was implemented according to a pipeline approach. While the ultimate aim is to establish a running CHF in every parish, logistical constraints forced the implementing agency to gradually roll out their intervention so that some parishes are treated earlier than others. We refer to households living in parishes where the intervention had not yet taken place as not having access to formal insurance. Strictly speaking, the selection of (early) access parishes was arbitrary and not based on an explicit randomization procedure. But, not surprisingly, access and no-access households are very similar across almost all dimensions we measured. Abundant anecdotal evidence suggests that access parishes are comparable to no-access parishes and that there are more differences between villages within the same parish than between parishes. Moreover, villages with and without access are not significantly different in terms of total number

¹²The open-ended questionnaire highlighted that people help each other in times of illness, contributing to hospital bills and helping the family financially through donations of food and money, or by transporting sick persons to the hospital. In the words of one respondent, “here the story is that people help each other [...] when they are sick, and in case of unlucky events”.

of households, distance and time from a major road and from the Sub-County headquarters (see Table 4.1).¹³

Table 4.1 – Group means of socio-economic variables

Variables	No-access parishes		Access parishes		Diff.	SE
	N	Mean	N	Mean		
Village size (households)	23	65.13	21	58.05	-7.08	4.27
Time to main road	23	6.13	21	7.24	1.11	1.56
Distance from main road	23	2.33	21	2.47	0.14	0.66
Time to town	23	15.70	21	14.67	-1.03	3.18
Distance from town	23	5.50	21	4.31	-1.19	0.89
Female participant	230	0.47	210	0.42	-0.05	0.05
Single	216	0.05	207	0.03	-0.02	0.02
Married/Engaged	216	0.79	207	0.79	0.00	0.04
Widowed	216	0.15	207	0.16	0.01	0.02
Male household head	230	0.75	210	0.76	-0.01	0.04
Age household head	216	46.61	199	50.23	3.61**	1.54
Education household head	230	5.31	209	4.95	-0.36	0.41
Household size	230	6.13	210	6.05	-0.08	0.23
Mothers in house	230	0.74	210	0.79	0.05	0.41
Watch TV weekly	230	0.87	210	0.88	0.01	0.03
Read newspaper weekly	230	0.24	210	0.24	0.00	0.04
Radio	230	0.91	210	0.91	0.00	0.03
Phone	230	0.84	210	0.78	-0.06	0.04
Bicycle	230	0.45	210	0.42	-0.03	0.05
Motorbike	230	0.15	210	0.11	-0.04	0.03
Television	230	0.04	210	0.04	0.00	0.02
Car	230	0.04	210	0.04	0.00	0.02
Generator	230	0.02	210	0.03	0.01	0.01
Wealth factor	230	0.04	210	-0.05	-0.09	0.08
House features index	230	1.40	210	1.35	0.05	0.08
Common assets index	230	2.20	210	2.11	0.05	0.08

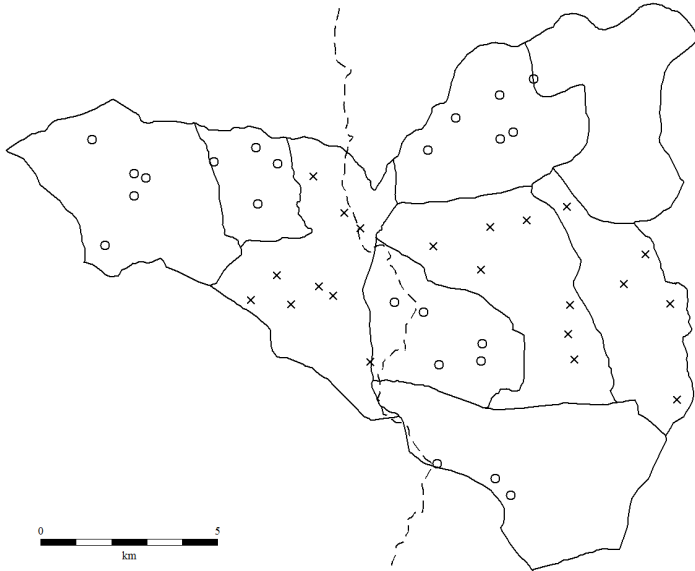
Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We randomly selected 21 villages from three access parishes, and another 23 villages from five neighboring parishes (in what follows: no-access villages). Four of these parishes are located in the same sub-county as the access parishes (Kitagata), and one parish containing five villages is located in neighboring

¹³Both distance (km) and time (minutes) were measured following the route most commonly used by locals. We used a motorbike taxi as reference for time, as it represents the most commonly used motorized vehicle by locals, as well as the means of transport typically used by SHU extension agents.

Mitooma sub-county, bordering the largest parish with an active insurance scheme (see Figure 4.1 for a spatial representation of the distribution of access and no-access villages). In each village, we first constructed a census of all households and then randomly selected 10 households per village. After a household survey, we randomly selected one adult family member to participate in our lab-in-field experiment.¹⁴ We have no relation with the NGO offering the insurance product, and to avoid gift exchange and demand effects, we did not emphasize the insurance program in our sessions.

Figure 4.1 – Distribution of villages with and without access to the insurance



Notes: Cross represents surveyed villages in parishes with access, circles represent surveyed villages in parishes without access. The dotted line represents the main road crossing the Kitagata sub-county.

¹⁴Enumerators made a list of each eligible family member before randomly selecting one name through a transparent ballot. Households were informed that the selected participant would be the only accepted household representative in the experiment and that failure to comply with the rule would result in the exclusion of the household from the study. Furthermore, 44.5% of selected participants were female (see Table 4.1) and compliance averaged 93%. The plausible non-random attrition bias is discussed in Section 4.5.

Table 4.1 summarizes the household data, distinguishing between access and no-access households. We first establish that access and no-access respondents are indistinguishable according to most socio-economic dimensions—the two groups appear balanced. This is consistent with information given to us by the implementing NGO—they did not use a specific targeting rule when prioritizing villages to enroll in the program. Age of the household head is the only variable that differs, but this difference is small and presumably caused by chance. A key variable for our purposes is the so-called wealth factor. This variable is constructed by factoring out a wealth index from a series of dummy variables related to asset ownership (i.e. phone, radio, television, generator, bicycle, motorbike and car) and house features (iron roofing, brick walls and cement floors) locally perceived as primary indicators of wealth. The average asset index score is not significantly different in access and no-access villages. Also, for the full range of socio-economic variables, bias-corrected variance estimates within parishes are greater than those across parishes. For definitions of the remaining variables, refer to Appendix 4.A.

In terms of analysis, we first explain variation in adoption of insurance in access villages. Surveys identified lack of financial resources as the most frequent reason for opting-out of the insurance scheme. We therefore expect wealthy households to be more likely to join the scheme—non-random adoption of the new technology based on income or wealth. On the other hand, because health status is also likely to vary with wealth, demand for insurance may also be greater among the non-wealthy. Ultimately, the distribution of effective demand for formal insurance across social groups is an empirical question and using the sub-sample of treated parishes (villages), we estimate the following probit models:

$$Pr(I_i = 1|W_i) = \alpha + \beta W_i + \gamma W_i^2 + \varepsilon_i \quad (4.1)$$

and

$$Pr(I_i = 1|P_i) = \alpha + \beta P_i + \gamma' \mathbf{X}_i + \varepsilon_i \quad (4.1')$$

where I_i is a dummy for insurance uptake by household i , W is a standardized

factor of wealth (based on a range of assets, discussed above), P is a poverty dummy, \mathbf{X} is a vector of individual and household characteristics and ε_i is a random error term. The poverty dummy takes a value of one when a household does not own more than one of the three most common assets (phone, radio and bicycle) and not more than one improved house feature (iron roof, brick wall and cement floor).¹⁵ In words, we explain adoption status by variables measuring wealth or poverty, and our first research question reads as follows:

RQ 1: Are wealthy households more likely to pick up health insurance than poor ones?

Next, we are interested in explaining the effect of adoption of insurance on village-level social capital. To construct a measure of social capital, we follow Fearon et al. (2009) and organized a lab-in-field experiment, or a standard PG game with five participants. A PG game captures the ability of groups of respondents to overcome free riding incentives. It likely also picks up altruism and trust (as respondents want to avoid outcomes where they feel others have taken advantage of them), as well as the extent to which norms of sharing and cooperation are internalized by villagers. These issues speak to different dimensions of (cognitive) social capital. Our measure of social capital is thus quite broad and extends beyond a willingness to share risk and income informally (as would be measured more cleanly and saliently by, for example, the solidarity game proposed by Selten and Ockenfels (1998)). Intuitively, one might expect that picking up effects across domains of social interaction are more demanding than picking up impact within the same domain (say: protecting individual respondents from idiosyncratic shocks by either formal insurance or sharing). In that sense, our test based on the PG game is perhaps more challenging than tests based on solidarity game data.

¹⁵The chosen poverty line corresponds to a wealth level that is one standard deviation below the average wealth index in our sample. This yields a poverty headcount of 17.8% that matches the region's poverty headcount based on DHS data (see Daniels, 2011).

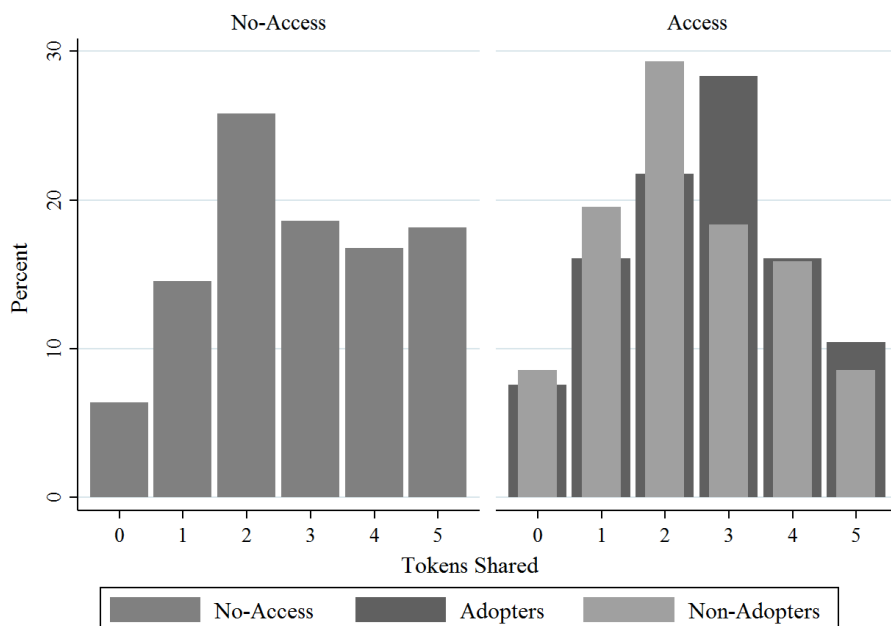
We randomly divided our 10 participants from each village into two groups of five individuals, carefully explained the rules of the game, and played a trial round. While individual contributions were not revealed, the people knew with whom they were matched for the five rounds of the PG game (and presumably whether these peers had adopted the formal insurance, or not). We gave five tokens to each participant, who could then anonymously contribute any integer amount into the common pot. After participants made their contributions, we doubled the amount in the pot and split this amount equally across all group members. As is well known, the joint surplus is maximized when all participants contribute the full amount to the pot, but the privately optimal contribution (the Nash equilibrium) in the context of a finitely repeated game is to give nothing. A large body of evidence from other contexts (and other games) suggests that many respondents are not entirely selfish, and willing to contribute positive amounts.

After the trial round, participants were informed that they would play the game an unknown number of times (up to a maximum of five) and that their pay-out at the end of the game would be based on a randomly selected round (determined by a simple lottery).¹⁶ Participants were sitting at a distance of about 5 m from each other, in a circle, facing outwards (not facing each other). Decisions about how many tokens to share were shown privately to the experimenters by placing the shared tokens in the closed left hand, while keeping the rest in the closed right hand. In turns, experimenters would approach participants and open their left fist rapidly to count and record the number of shared tokens. It was impossible for participants to observe each other. Payoffs for participants averaged around USD2.20—or about two-third of daily rural household incomes in the Western region, as estimated by the Ugandan Bureau of Statistics.¹⁷

¹⁶In the absence of group reshuffling, game dynamics of reputation, retaliation and learning are expected to drive contribution decisions in repeated games (Rand et al., 2009). Our intent, however, is to proxy real life dynamics. For this reason, our analysis is based solely on the choices performed during the first round of each game version. Successive iterations were performed so that defection is not necessarily the only Nash equilibrium present: players may cooperate on an equilibrium path, provided that they are sufficiently interested in future outcomes (Dal Bó, 2005). For the same reason, the exact number of repetitions was not revealed to participants.

¹⁷<http://www.ubos.org/UNHS0910/chapter7.Average%20Monthly%20Household%20Income.html>. The payoffs after the trial round were not revealed so as not to influence subsequent play in

Figure 4.2 – Sharing decisions of participants with and without access to formal insurance



If formal insurance crowds out informal insurance and reduces cooperation at the local level, we expect aggregate (or average) contributions to the common pot to be lower for respondents with access to the formal insurance scheme. We seek to pick up an intention-to-treat effect (as did Klohn and Strupat, 2013) and estimate the following two models:

the game. On average, participants without access to the formal insurance shared 2.89 tokens, against the 2.60 tokens shared by adopters and 2.39 by non-adopters with access. Similarly, only 25.5% of the participants with access made a “high contribution” to the common pot (four or five tokens), against 34.8% of those without access (see Figure 4.2). Our empirical analysis is based on contributions in the first round. Contributions in subsequent rounds are less likely to be independent observations, because players may respond to each other as the game unfolds.

$$S_i = \alpha + \beta I_i + \varepsilon_i \quad (4.2)$$

and

$$Pr(SH_i = 1|I_i) = \alpha + \beta I_i + \varepsilon \quad (4.3)$$

where S represents the number of tokens shared in the common pot, I is again a dummy with value 1 for respondents living in access villages; β measures an intention to treat effect and SH is a dummy for having shared a high amount (i.e. four or five tokens), so this model is estimated using a probit specification. In some models, we control for a vector of observables, and in all regression models we will cluster standard errors at the village level. In words, by estimating (4.2) and (4.3) we ask:

RQ 2: Is the possibility to access a formal insurance associated with lower average contributions to the public good in a PG game?

We now analyze contributions to the common pot more closely. The conventional economic perspective on the evolution of informal insurance suggests that formal insurance provides a substitute mechanism for the wealthy, who could subsequently choose to opt out of existing informal arrangements and cease cooperation with fellow villagers. We estimate the following “naïve” models, which resemble equations (4.2) and (4.3) above but distinguish between adopters and non-adopters:

$$S_i = \alpha + \beta A_i + \gamma NA_i + \varepsilon \quad (4.4)$$

and

$$Pr(SH_i = 1|A_i, NA_i) = \alpha + \beta A_i + \gamma NA_i + \varepsilon \quad (4.5)$$

where A is a dummy with value 1 for actual adopters (in access villages), and NA is a dummy with value 1 for non-adopters (in access villages). Respondents without access are the omitted category. Note that models ((4.4) and (4.5)) are naïve and

possibly produce misleading results because of the endogeneity of adopter status. Households self-select in the insurance scheme, and variables explaining this choice may also drive behavior in the public good game. With this (important) caveat in mind, our third research question is:

RQ 3: Assuming that the introduction of formal insurance is associated with reduced aggregate contributions to the common pot in a PG game, who is responsible for these lower contributions: adopters, non-adopters, or both?

As mentioned, correlations obtained in our regression framework may be spurious. To attenuate concerns about omitted variables, we expand the empirical analysis. First, we explain adoption status (based on observations in access parishes only) and, based on the resulting regression coefficients, we predict the probability of non-adoption for all households (including those in no-access villages). This analysis enables us to compare public good contributions for (predicted) adopters and non-adopters across villages with and without access to insurance. We are interested in establishing whether predicted non-adopters without access behave similarly to predicted non-adopters with access, and estimate

$$Pr(SH_i = 1 | Pr(n - a)) = \alpha + \beta Pr(n - a)_i + \varepsilon_i \quad (4.6)$$

and

$$Pr(SH_i = 1 | Pr(n - a), I) = \alpha + \beta Pr(n - a)_i + \gamma I_i + \delta I_i \times Pr(n - a) + \varepsilon_i \quad (4.6')$$

where $0 \leq Pr(n - a) \leq 1$ is the predicted probability of non-adoption, and $I \times Pr(n - a)$ is the model with interaction term between $Pr(n - a)$ and access to insurance (I).

Second, we also estimate a propensity score matching (PSM) model, matching respondents from access and no-access areas. To probe whether contributions by predicted non-adopters are similar, or not, we then compare contributions of non-adopters, with credibly similar respondents in no access-areas, based on nearest neighbor matching.

RQ 4: Do predicted non-adopters with access to formal insurance behave the same as predicted non-adopters without access—is behavior of non-adopters driven by population characteristics or by insurance?

Finally, to examine the mechanism explaining our empirical results we will also estimate a placebo model based on inter-group variation of group composition. We outline this step below, after first addressing research questions 1–4.

4.5 Empirical results

We first test the prediction that wealthier households are more likely to purchase formal health insurance than relatively poor households. Focusing on parishes where insurance was offered, we explain variation in adoption across households. Results are reported in Table 4.2. Columns (1) and (2) illustrate a positive correlation between wealth and adoption. This relation may possibly be non-linear (Column (1)), but a linear relation ensues when we omit two outliers from the sample (see Column (2)). A one standard deviation increase in wealth increases the probability of adopting formal insurance by about 12%. The finding that relatively wealthy households are more likely to adopt formal insurance is confirmed in Columns (3) and (4), where we replace the wealth variables with a poverty dummy. We report probit marginal effects, so poor households are 20% less likely to adopt health insurance than non-poor households in access villages.¹⁸

We next ask how the introduction of formal insurance affects play in the PG game. We first use the number of tokens shared as the dependent variable and examine how this quantity varies with access status. Our access dummy captures whether or not a respondent’s household resides in a village that has an active

¹⁸This finding is confirmed by abundant anecdotal evidence. Many respondents cite “financial problems” as the main reason for not purchasing the insurance. They claim that their income is too low in comparison to the fee and would have joined the scheme “if only they had more money”. Others state that they would join if there was a “reduction on the fee that SHU asks from them”, or stress that they “joined but paid half of the money”, and thus could not benefit from the insurance coverage.

Table 4.2 – Wealth and the adoption of health insurance

	Insurance adoption			
	(1)	(2)	(3)	(4)
	Probit	Probit	Probit	Probit
Wealth factor	0.122** (0.050)	0.116** (0.054)		
Wealth factor squared	-0.111** (0.048)	-0.070 (0.074)		
Poverty dummy			-0.226** (0.098)	-0.198** (0.093)
Additional controls	No	No	No	Yes
Cluster robust s.e.	21	21	21	21
<i>N</i>	210	208	210	185
Correct predictions at $p = 0.5$	63%	63%	62%	70%

Notes: Cluster robust standard errors in parentheses; probit marginal effects. Wealth factor based on vector of household level asset dummies. Model (2) removes the highest outlier from the sample. Probit marginal effects for Model (3); Poverty dummy takes the value of unity when household owns not more than one of the three most common assets (phone, radio, bicycle) and not more than one improved house feature (iron roof, brick wall, cement floor). Additional controls include age household head, male household head, education household head, single, household size, female participant, mothers in house, watch TV weekly, read newspaper weekly. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

CHF scheme, regardless of its own adoption status. Of 440 households interviewed, only 409 correct household representatives showed up for the afternoon (PG) game session. The other 31 participants were substituted by a random villager that was not a member of any of the interviewed households. We do not find detectable self-selection bias into the experiment and therefore estimate the model using OLS, Poisson and Probit.¹⁹ In Column (1), we find that respondents with access to the formal insurance share on average fewer tokens than their counterparts, consistent with findings by Klohn and Strupat (2013). Using a Poisson model yields similarly significant results.

In Columns (2)–(4), we attempt to go further and identify who is responsible for the lower PG contributions—adopters or non-adopters. While one may expect that adopters will reduce their contribution to the PG—as they have less use for social capital in the future—our empirical results suggest something different. Lower public good contributions are driven by relatively lower contributions of

¹⁹We also ran a Heckman selection model. While some observed characteristics affect the likelihood of showing up (i.e. if the household head is male, or if the randomly selected respondent is female), the coefficient on the inverse Mills ratio is statistically insignificant.

non-adopters. On average, non-adopters with access to the formal insurance contribute about 0.4 tokens less than respondents without access, or over one-quarter of a standard deviation. In Column (3), we distinguish between households that have paid their premium (i.e. eligible to benefit from the insurance) and those that have only paid in part. Again, non-participants drive the reduction in contributions. In Column (4), we show the same result using a Poisson specification instead.

Similarly, we try to explain the incidence of high contributions (i.e. the contribution of four or five tokens). There are fewer high contributors in access areas (Column (5)), which is driven by non-adopters (Columns (6) and (7)). They are 33% less likely to be high contributors, again representing about one-quarter of a standard deviation.

Table 4.3 – Sharing in the public goods game

		Tokens shared (S_i)				High contributor (SH_i)		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
		OLS	OLS	OLS	Poisson	Probit	Probit	Probit
Insurance access		-0.281*				-0.093**		
		(0.144)				(0.045)		
Adopter			-0.188				0.082	
			(0.185)				(0.057)	
Full payment				0.008	0.003			0.049
				(0.352)	(0.125)			(0.132)
Partial payment				-0.234	-0.087			-0.112*
				(0.192)	(0.072)			(0.054)
Non-adopter			-0.402*	-0.402*	-0.155**		-0.101**	-0.101**
			(0.226)	(0.226)	(0.091)		(0.046)	(0.046)
Cluster	robust	44	44	44	44	44	44	44
s.e.								
N		409	409	409	409	409	409	409

Notes: Cluster robust standard errors in parentheses; probit marginal effects in (5)-(7). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Other factors may explain both adoption and contribution rates. Table 4.2, for instance, shows that non-adoption is primarily explained by wealth—poor households are significantly less likely to adopt insurance. If poorer households systematically contribute less to the common pot, the correlations in Table 4.3 may be spurious. We next use information about the propensity to be a non-adopter

(based on regression results in Column (4), Table 4.2). This enables us to compare “predicted non-adopters” in access areas to “predicted non-adopters” in no-access areas.

Results are reported in Table 4.4. In Column (1), we only use data from villages with access to the insurance product and demonstrate that predicted non-adopters are less likely to share a high fraction of their tokens in the PG game. These results, of course, echo those in Table 4.3. However, when estimating the same model for the sub-sample of respondents with no-access, we find that the sign of the coefficient of interest changes—from negative to positive (Column (2)). Predicted non-adopters without access are more likely to contribute four or five tokens to the common pot. This may reflect the greater importance attached to social capital by relatively poor households. Upon pooling the data and including an interaction term (predicted non-adoption probability multiplied by insurance access), we capture these results in one specification. According to Columns (3) and (5), predicted non-adopters in no-access areas are more likely to contribute more to the common pot, but this is not true for predicted non-adopters in access villages—the interaction effect dominates the level effect of predicted non-adoption. Results in (3) are robust to clustering at the parish level (8 clusters) instead of the village level (44 clusters), using wild bootstrap inference to correct for the small number of clusters (Cameron et al., 2008).

As an alternative approach, in Columns (4) and (6) we use PSM to compare non-adopters to similar counterparts in no-access areas. We thus exclude the 106 adopters from this analysis. The average treatment effect obtained through nearest neighbor matching is consistent with previous findings. Matching is based on the same variables as in Column (4) of Table 4.2. A logit model is used to calculate propensity scores. Balancing properties are satisfied for all matching variables, in each of the five blocks of equal score range. Also, the mean propensity score in each block is not different for non-adopters (in this setting, the “treated”) and no-access participants (“controls”). The predicted scores range between 0.044 and 0.688 for non-treated and between 0.045 and 0.602 for controls, so the common support region ranges between 0.045 and 0.602 (suggesting adequate overlap between

Table 4.4 – Predicted non-adoption and sharing in the public goods game

	High contributor (SH_i)				Tokens shared (S_i)	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	Poisson	Probit	Probit
$Pr(n - a)$	-0.325*	0.309*	0.305*		0.691	
	(0.171)	(0.165)	(0.162)		(0.523)	
Insurance access			0.247**		0.424	
			(0.110)		(0.351)	
Insurance access $\times Pr(n - a)$			-0.635***		-1.258*	
			(0.235)		(0.678)	
Non-adopter with access				-0.105*		-0.402**
				(0.058)		(0.186)
Cluster robust s.e.	21	23	44	No	44	No
N	185	184	369	303	369	303

Notes: Cluster robust standard errors in parentheses; probit marginal effects in (1)–(3). Average treatment effect on the treated (ATT) using nearest neighbor matching with random draw. The propensity score in (4) and (6) is calculated on poverty dummy, age household head, male household head, education household head, single, household size, female participant, mothers in house, watch TV weekly, read newspaper weekly (see Table 4.2, Column (4)); the same variables are used to predict non-adoption in Columns (1)–(3) and (5). Clustering at the parish level in (3), and correcting for the small number of clusters following Cameron et al. (2008), yields a p-value of 0.020 for the interaction term, significant at the 5% level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

groups). Restricting the analysis to the common support region ($N = 303$) yields an ATT of -0.105 in (4), and of -0.349 in (6). The former coefficient is significantly different from zero at the 10% level, the latter at the 5% level. Kernel matching with bootstrapped standard errors yields similar results to Table 4.4.

We interpret these findings as tentative support for the idea that our results in Table 4.3 are not driven by spurious correlation. Non-adopters do not contribute less in public good games because they are poor, but because of the interaction between poverty and formal insurance availability.

4.6 Towards a speculative interpretation

Our empirical results indicate that (i) the average contribution to the common pot is lower in villages with access to insurance than in no-access villages and (ii) that non-adopters are responsible for the fall in contributions. These results are consistent with a theory based on signal sending. Non-adopters may feel “left behind” after the relatively wealthy have purchased formal insurance: they fear

that the quality of the informal sharing network is compromised if the wealthy opt out, and signal their discontent by foregoing cooperation in other domains of social interaction.

Such a story is consistent with work by the anthropologist Scott (1985), who studied the consequences of mechanization in rice farming in a Malay village. This process invited consolidation of farms (e.g. by inviting landowners to start reneging on long-term tenure arrangements) and dramatically reduced demand for hired labor. The result was a “proletarianization” of small farmers and landless laborers—the creation of an underclass of society members whose well-being was increasingly inconsequential for the upper strata of society (and whose interests were increasingly ignored). This is the unraveling of social networks caused by selective exit of the privileged (i.e. the adopters). Scott then demonstrates that “losers” in one sphere of interaction (e.g. the labor market) are fully aware of their situation, resent it and seek to remind the adopters of their historical responsibilities and social obligations by retaliation in other spheres of interaction—including everyday social life.

Scott (1985) investigates such strategies in detail. Peasants recognize their limited (economic, political and symbolic) power and typically prefer non-rebellious and non-revolutionary acts of resistance. Typical “weapons of the weak” consist of low-key recalcitrance, foot dragging, dissimulation, false compliance, pilfering, feigned ignorance, chicanery, slander, arson and acts of minor sabotage. Resistance strategies range from gossip and character assassination to strikes (such as when machines break down and landowners suddenly need labor to harvest their fields) and boycotts of social activities such as weddings and political rallies organized by those well-off. More in general, resistance signals a reduced overall tendency to cooperate. This indicates a deterioration of local levels of social capital, undermining the ability of communities to overcome social dilemmas and coordinate on first-best outcomes. Even though the adverse consequences of insurance for non-adopters are more indirect and less immediate than the consequences of mechanization for unskilled workers, we believe our findings neatly fit this narrative.

To probe the plausibility of this particular causal mechanism explaining our results, we considered a placebo treatment. Random assignment of villagers to groups implied that the share of adopters varied across groups, and some groups did not contain any adopters. In such groups, non-adopters cannot signal their discontent by contributing little to the common pot. If our hypothesis is correct, non-adopters should contribute more to the common pot in group without adopters. To test this, we focus on behavior of non-adopters (from Treatment villages). There are 82 of such non-adopters in our total sample. We regress common pot contributions of these non-adopters on a dummy variable indicating the absence of adopters in their experimental group. We also estimate models based on the full sample of observations from treatment villages and regress common pot contributions on the interaction of non-adopters and the group dummy indicating the absence of adopters.

Results are presented in Table 4.5 and reveal that non-adopters give less in the presence of adopters. Comparing group sessions with and without adopters, we find that non-adopters give almost 1.7 tokens more when there is no need to signal discontent (Column (1)). Based on the full sample of respondents from treatment villages, we observe that our subjects contribute on average 0.88 fewer tokens when they are matched with non-adopters. However, when they are a non-adopter themselves, they contribute 2.55 tokens more (so that, net, non-adopters again donate 1.7 tokens more when they play the public goods game with fellow non-adopters).

However, we hasten to point out that our results cannot “prove” the weapons of the weak theory, as alternative explanations may fit the data as well. For example, the way the insurance product was marketed could have encouraged individualistic thinking (e.g. by suggesting to senior household members that they alone were ultimately responsible for the well-being of their families and that this product would help them meet their responsibilities). If so, lower public good contributions by non-adopters might be a response to a widening aspiration gap. Alternatively, the uptake of insurance could be viewed by non-adopters as a signal of wealth on the part of the takers, and of their own relative poverty in an environment

Table 4.5 – Signal sending as the causal mechanism?

	Tokens shared non-adopters only (1) OLS	Share high non-adopters only (2) Probit	Tokens shared treatment villages (3) OLS	Share high treatment villages (4) Probit
Zero others adopt	1.671* (0.841)	0.439 (0.278)	-0.877** (0.346)	-0.301*** (0.0358)
Non-adopter			-0.366* (0.213)	-0.0523 (0.0526)
Zero others adopt × non-adopter			2.547*** (0.908)	0.852*** (0.0255)
<i>N</i>	82	82	188	188
(Pseudo-) <i>R</i> ²	0.050	0.027	0.048	0.046

Notes: Robust standard errors in parentheses; probit marginal effects in (2) and (4). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

where relative wealth is typically only imperfectly observed. If so, lower public good contributions could be due to the activation of a norm about richer people contributing more to local public goods, and poorer people contributing less (e.g. Ensminger, 2004). A third explanation could be that non-adopters seek to cement relationships with (remaining) non-adopters and re-allocate their spending accordingly. Disentangling the merits of these alternative explanations is left for future research.

4.7 Discussion and conclusions

Social capital is a key determinant of the ability of communities to overcome social dilemmas, and fostering social capital and cooperation has become a policy goal in and of itself. For example, Fearon et al. (2009) demonstrate that outside innovations may facilitate investments in social capital. However, not all interventions have such benign effects. Specifically, interventions which benefit a sub-sample of the population may weaken mutual dependencies and erode social capital. Our study provides evidence to support the latter—our data are consistent with a broad perspective on social interactions, where reduced inter-dependencies

in one domain (informal insurance) invite behavioral changes in other domains (public good contributions).

Formal insurance provides a substitute for informal insurance, crowding out social capital and informal cooperation. Anecdotally, we find evidence of such dynamics in our open-ended exit questionnaire. Respondents lament that since the introduction of the formal insurance, some of the informal risk sharing mechanisms have deteriorated. A non-adopter states that, in contrast to insurance holders, when she fell sick she had to “totally cater for herself”. Another respondent claims that someone “had joined SHU, and therefore refused to help a neighbor who had a sick son”. In the words of other respondents, the insurance scheme “spoils care for those who can’t pay for the scheme”, and “if you are not a member you are not catered for”. But the deterioration of mutual assistance works both ways. Various respondents emphasize that (non-adopting) neighbors refuse to take insurance holders to the hospital (even if transport costs are typically not part of the insurance package). Another respondent mentions the case of a pregnant woman whom “people refused to help, because she is in the insurance”. Apparently, non-adopters are not passively undergoing their plight—they take action and signal their discontent by ceasing cooperation in other domains, in an effort to “discipline” adopters.

Indeed, one important new finding of our study is that the erosion of social capital is not primarily caused by adopting community members—those respondents who may opt out of the insurance network. Instead, the deterioration of social capital in our field experiments is explained by a behavioral response of non-adopters. Non-adopters contribute less to the public good, but only when they are matched with adopters.

Appendix 4.A: Detailed variable definitions

Female participant. A dummy equal to 1 if selected game participant is female, 0 otherwise.

Single. A dummy equal to 1 if the household head is single, 0 otherwise.

Married/Engaged. A dummy equal to 1 if the household head is married or engaged, 0 otherwise.

Widowed. A dummy equal to 1 if the household head is widowed, 0 otherwise.

Age household head. Age class of the household head, in 20-year blocks from adulthood (18).

Male household head. A dummy equal to 1 if the household head is male, 0 otherwise.

Education household head. A dummy equal to 1 if the household head has completed primary school.

Household size. Number of people that usually eat at least one daily meal together with the household head, acknowledging its authority and living with the rest of the household.

Mothers in house. A dummy equal to 1 if the household comprises at least one pregnant woman or mother taking care of a child at present time, 0 otherwise.

Watch TV weekly. A dummy takes value of 1 if the household head declares to watch TV at least once a week, zero otherwise.

Read newspaper weekly. A dummy equal to 1 if the household head declares to read the newspaper at least once a week, 0 otherwise.

Radio. A dummy equal to 1 if the household possesses any type of radio, 0 otherwise.

Phone. A dummy equal to 1 if the household possesses any type of phone, 0 otherwise.

Bicycle. A dummy equal to 1 if the household possesses any type of bicycle, 0 otherwise.

Motorbike. A dummy equal to 1 if the household possesses any type of motorbike, 0 otherwise.

Television. A dummy equal to 1 if the household possesses any type of television, 0 otherwise.

Car. A dummy equal to 1 if the household possesses any type of car, 0 otherwise.

Generator. A dummy equal to 1 if the household possesses any type of electricity generator, 0 otherwise.

Wealth factor. A principal factor obtained by factor analysis following Sahn and Stifel (2003), from the above-mentioned asset list (radio to generator) possessed by the household.

House features index. The sum of three dummy variables representing improved house features, i.e. iron roof, brick wall and cement floor.

Common assets index. The sum of three dummy variables representing the three most common assets, i.e. radio, phone and bicycle.

Poverty dummy. A dummy equal to 1 if the household scores 1 or less in the House features index and scores 1 or less in the Common assets index, zero otherwise.

Adopter. A dummy equal to 1 if a household is residing in a parish (village) with an active and running CHF insurance scheme, and adopted the insurance, 0 otherwise.

Non-adopter. A dummy equal to 1 if a household is residing in a parish (village) with an active and running CHF insurance scheme, and knowingly renounced participation, 0 otherwise.

Full payment. A dummy equal to 1 if a household that adopted the CHF insurance scheme has fully paid the due premium, 0 otherwise.

Partial payment. A dummy equal to 1 if a household that adopted the CHF insurance scheme has only partially paid the due premium, 0 otherwise.

Insurance access. The sum of Adopter and Non-adopter. It takes value 1 if the household has access to the formal insurance, regardless of its adoption status; zero otherwise.

S. The number of tokens contributed to the common pot of the public goods game by a given participant, minimum is 0 and maximum is 5.

SH. A dummy equal to 1 if the participant has contributed 4 or 5 tokens to the common pot of the PG game—(i.e. one standard deviation or more above the average), 0 otherwise.

Non-adoption probability $Pr(n-a)$. The probability of adoption as predicted for respondents from access areas, based on the individual and household level socio-economic characteristics of Column (4), Table 4.2.

Chapter 5

Towards Prosocial Policy

Evaluating the Impact of Performance-Based Financing in Ugandan Health Facilities

Abstract: We estimate the effect of introducing performance-based financing on the allocative efficiency and quality of healthcare provision. We use a unique panel of output and expenditure data from small private not-for-profit healthcare facilities in Uganda in combination with a patient satisfaction survey focused on the perceived quality of healthcare. In the observed facilities, performance-based financing increases output and allocative efficiency by more than 25% with no apparent reduction in quality. The result is robust to more traditional estimation techniques.

This chapter is based on:

Duchoslav, J. and F. Cecchi (2017). Evaluating the Impact of Performance-Based Financing in Healthcare Provision: A Case Study in Uganda. Working paper.

5.1 Introduction

Improvements in healthcare provision form one of the cornerstones of economic development: 3 of the 8 Millennium Development Goals were directly related to healthcare, and the new Sustainable Development Goals call for increased health financing in developing countries. There is certainly room for improvement in this respect. According to the World Bank (2016), low-income countries spent 7 times less money per capita on health than middle-income ones, and 26 times less than non-OECD high-income countries in 2014. However, when expressed as a percentage of GDP, health expenditures were roughly the same in low-income countries (5.7%) as in middle- and high-income non-OECD countries (5.8% and 5.5% respectively).¹ If low-income countries are to increase spending on health, they must either divert it from elsewhere, or wait for economic growth to increase the total available funds. Alternatively, they can try to do more with less—to increase allocative efficiency.

Performance-based financing (PBF) is becoming a popular measure to do just that (Hecht et al., 2004; Brenzel, 2009; Eldridge and Palmer, 2009; Honda, 2013). As the term suggests, PBF makes the amount of funding a healthcare provider receives conditional on performance, most commonly defined in terms of output. Healthcare providers and their staff should respond positively to such incentive schemes as long as the marginal revenue offsets the marginal costs associated with higher effort and increased output (Rusa et al., 2009). While this mechanism is straightforward in theory, the evidence that it works well in practice is inconclusive, and mostly focuses on partial outcomes (Eijkenaar et al., 2013). Moreover, Scheffler (2010) points out that most evaluations lack proper control groups, or confound the effect of performance incentives with additional elements of the program that might impact practices. Even the most rigorously designed randomized control trials (RCTs) often prove to be unworkable, or prone to last-minute political and organizational interferences that weaken the validity of

¹High-income non-OECD countries include for example Argentina, Russia or Saudi Arabia. OECD countries, whose healthcare spending figures are somewhat inflated by the United States, spent 12.6% of their GDP – or 149 times more per capita – than their low-income counterparts.

the findings, especially in the context of developing countries (Soeters et al., 2006). In 2012, an editorial of the Bulletin of the World Health Organization about PBF in low- and middle-income countries advocated the use of other robust evaluation designs if conducting an RCT is not feasible. One option is an interrupted time series, in which outcome data are collected at regular intervals during baseline and post-intervention periods (Fretheim et al., 2012). We take this approach in the present chapter, using a panel dataset from a network of private not-for-profit health facilities in Uganda to estimate the effects of introducing PBF on healthcare output and – by extension – on the allocative efficiency of healthcare service delivery.

In Section 5.2 we provide the background for the study, focusing on the latest literature on PBF and a description of the healthcare system in Uganda. Section 5.3 describes the data, while the methodological approach is outlined in Section 5.4. The results are presented in Section 5.5 and discussed in Section 5.6. Section 5.7 concludes.

5.2 Background

5.2.1 PBF: a policy to improve healthcare provision

The growing evidence that the health of its population is an important determinant of a country's economic growth (Bloom and Canning, 2000; Weil, 2007) has provided an additional argument – besides the ethical ones – for the need for functional and accessible healthcare provision. This in turn is essentially a function of structural inputs (including people, infrastructure, knowledge, drugs, material, equipment, and technology) and the processes transforming these inputs into outputs (Eichler, 2006). Though usually thought of as complementary, the right processes can – to an extent – make up for the lack of inputs (Peabody et al., 2004). By improving the transforming processes, more output can be produced using the same limited inputs.

In recent years, PBF has become one of the favorite ways to stimulate such

improvements (Hecht et al., 2004; Brenzel, 2009; Eldridge and Palmer, 2009; Honda, 2013). However, while the principal-agent problem has been successfully reduced by conditioning payment on performance in many other contexts, it is rather difficult in processes with such multi-dimensional output as healthcare, the production of which is complicated to quantify.

Even if performance is understood in its most limited sense as output, the many different types of output produced by a healthcare provider have to be taken into account when assessing its performance – either individually or according to some conversion logic. Expanding the notion of performance to include the quality of produced output naturally complicates the matter even further. Some PBF schemes have nonetheless made attempts to condition funding on performance in quality indicators (Perrot et al., 2010). In Basinga et al. (2011) for example, quality is measured as an index of observable structural and process measures, and enters the payment formula as a multiplicative factor that may lower the output-based payment. Soeters et al. (2011) instead try to ensure quality maintenance through comprehensive agreements with providers and regulators, and measure it through patient-perceived quality surveys that do not directly influence bonuses.

Besides the various ways in which they define and measure performance, PBF schemes differ along three major lines (Perrot et al., 2010):

1. performance targets and associated payments may apply to individual employees or to entire facilities;
2. conditional payments may make up the entire funding of the facility (or salary of the individual) or there may be a fixed component; and
3. payments may be conditional on fixed targets or incremental.

Each PBF program is thus quite unique, and operates in a unique setting. It is therefore hard – if not impossible – to agree on best practice (Ireland et al., 2011). In our case, incentive payments are determined at the facility level, and make up only a fraction of total facility income with capped incremental bonuses. The allocation of the bonus payments is at the discretion of the in-charge of the facility, and typically redistributed to employees.

Existing literature identifies several potential pitfalls of PBF. Oxman and Fretheim (2008) warn against the danger of widening the already existing gap between poorly- and well-performing facilities. Other concerns include the risk of increased gaming, i.e. systematic reporting bias (Lu, 1999), target-led distortions resulting in the production of services with negative marginal value (Wynia, 2009), and cherry-picking of patients who are most suited to achieve targets (Ireland et al., 2011). Finally, the bureaucratization of healthcare delivery, which is – to an extent – necessary to implement a PBF program, may end up crowding out intrinsic motivation (Frey and Jegen, 2001), inducing a decline in physician professionalism and morale (Wharam et al., 2009).

Despite these issues, some of which we touch upon in the discussion, PBF is receiving increasing attention from academics as well as policy-makers. Several studies document positive effects of PBF on partial outcomes (Meessen et al., 2006, 2007; Soeters et al., 2011; Basinga et al., 2011; Sekabaraga et al., 2011; Bonfrer et al., 2014), while others find no lasting effects (Banerjee et al., 2008; Morgan, 2010). The evidence to date is thus inconclusive (Eijkenaar et al., 2013), with even the most rigorous studies suffering from various methodological shortcomings (Fretheim et al., 2012).

5.2.2 Setting

The health sector in Uganda is characterized by a high degree of fragmentation with a mixture of public, private not-for-profit and private for-profit healthcare providers (Björkman and Svensson, 2009), with government health spending representing only 30% of the total (Hunt, 2010). Although the Ugandan Ministry of Health takes non-governmental providers into account in its planning, providing partial funding to some of them, private health facilities and practices account for half of Uganda's reported healthcare output and operate independently of public ones (Government of Uganda - Ministry of Health, 2010). The policies governing healthcare services in the country are consequently as diverse as the service providers. In this complex situation, the Ugandan Ministry of Health piloted a large-scale PBF program for healthcare providers, which mostly turned

out to be a failure (Morgan, 2010).

In this chapter, we focus on one of the largest not-for-profit private healthcare providers in the country – the Uganda Catholic Medical Bureau (UCMB). The UCMB runs an extensive country-wide network of hospitals and health centers accounting for over a third of private healthcare facilities in Uganda (Government of Uganda - Ministry of Health, 2010). The structure of the administration of the UCMB healthcare facilities mirrors that of the Catholic church itself: each of the 15 dioceses has its own health office, which is responsible for the operation of the health units within its territory. The central UCMB office coordinates the diocesan health offices, sets policy on the national level, monitors and evaluates the diocesan offices and the individual health units, and represents them nationally and internationally.

In 2008, the UCMB selected the diocesan health office in Jinja – one of the smaller dioceses – to pilot a PBF scheme in its six mid-sized health centers² to test the practical feasibility of this new approach before possibly extending the scheme to all its health units. The Diocese of Jinja was selected for its manageable size, for its long cooperation with an international NGO which helped set the scheme up, and for its historically good performance.³

The scheme was introduced in the 2008/2009 fiscal year. For each health unit, it set an output target based on its performance in the previous 3 years. The targets are defined in terms of the Standard Unit of Output (SUO) – a weighted average of the most commonly performed procedures and services (see Section 5.3 for full details). Starting in fiscal year 2009/2010, each of the participating facilities would receive an incremental bonus, conditional on reaching its pre-specified target. On

²There are several levels of healthcare facilities in Uganda: dispensaries and aid posts (HC I), health centers (HC II-IV from the smallest to the largest ones), and hospitals (HC V) (Björkman and Svensson, 2009). The pilot PBF scheme targets solely the mid-level HCs III and IV. We consequently limit our analysis to units of the same size.

³Due to the historically good performance of the pilot facilities, parallel trends cannot be assumed, precluding the possibility of using standard difference-in-differences approaches when estimating the effect of PBF. We instead use a panel-data structural break approach (see Section 5.4).

average, the bonus would increase the income of the facilities – about a third of which comes from user fees and the remainder from government grants – by about 5%. The targets were set sufficiently low for all units to reach them every year, effectively making the scheme a fully incremental one. Although the targets are set and the bonuses paid at the unit level, the heads of the units typically use the extra income to top up the salaries of their employees proportionally to their hours worked, thus bringing individual incentives in line with those at the facility level.

5.3 Data

To gauge the effect of PBF on the allocative efficiency of healthcare delivery, we use a range of output and input measurements collected by the UCMB from all its health units, which amount to a panel spanning up to 246 mid-sized health units over a period of thirteen years (fiscal year 2001/2002 – fiscal year 2013/2014). The last six fiscal years follow the introduction of PBF in the treated centers.

Output is measured using the Standard Unit of Output (*SUO*) – a unit in which the PBF targets are defined. The *SUO* is constructed from common services typically provided by small health facilities, taking into account their relative input requirements in terms of cost and time:

$$SUO = \text{outpatients} + 5 \times \text{inpatients} + 2 \times \text{deliveries} + 0.3 \times \text{ANC} + 0.2 \times \text{immunizations} \quad (5.1)$$

where *outpatients*, *inpatients*, *deliveries*, *ANC*, and *immunizations* are the numbers of outpatient visits, inpatient admissions, deliveries, ante-natal care visits (including family planning) and immunizations respectively. Total *expenditures* are measured in millions of 2014 Ugandan Shillings (UGX). Panel descriptive statistics for *SUO* – as well as for the other two main factors of production, i.e. the total number of staff (*staff*) and capital proxied by bed capacity (*beds*) – are presented in Table 5.1. Figure 5.1 shows the trends in output of PBF and control facilities. It plots a two-year moving average of output for PBF and control

Table 5.1 – Panel descriptive statistics

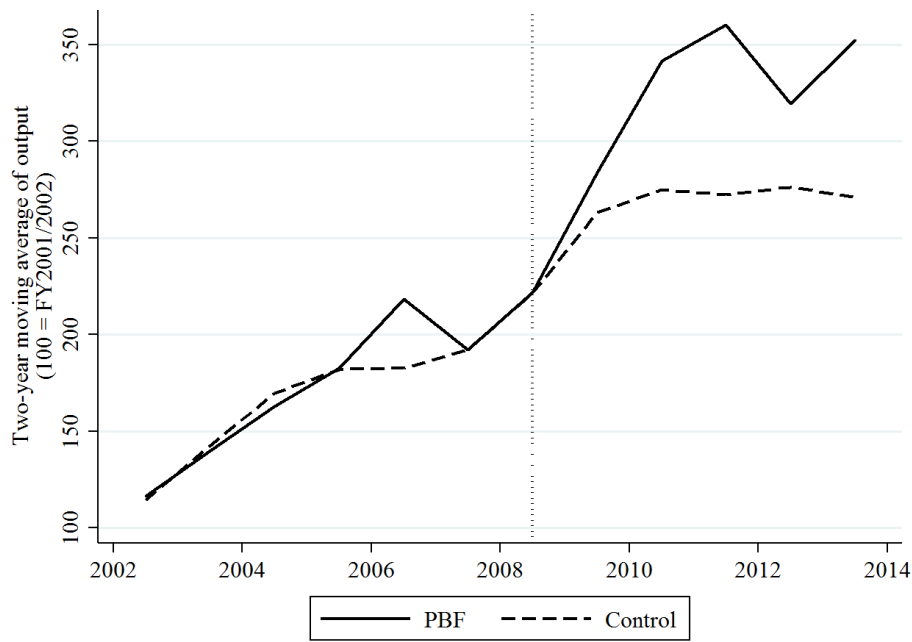
<i>year</i>	Obs.	PBF	<i>SUO</i>	<i>expenditures</i>	<i>staff</i>	<i>beds</i>
2001/02	79	N	10,471 (8,753)	78.1 (95.3)	11.7 (7.4)	19.8 (16.7)
2002/03	60	N	10,892 (10,309)	79.9 (56.9)	12.0 (8.0)	22.6 (16.8)
2003/04	120	N	11,982 (8,760)	86.7 (88.1)	11.7 (7.8)	19.4 (15.6)
2004/05	122	N	13,587 (9,543)	93.9 (91.2)	12.1 (8.3)	20.7 (16.0)
2005/06	145	N	12,798 (9,917)	86.4 (89.3)	10.8 (7.4)	19.8 (15.7)
2006/07	151	N	13,180 (10,012)	79.1 (75.6)	11.0 (7.7)	20.3 (15.7)
2007/08	170	N	12,956 (10,729)	72.3 (77.2)	10.4 (7.4)	19.5 (16.9)
2008/09	167	Y	15,533 (12,370)	75.9 (79.9)	9.8 (6.6)	18.3 (15.2)
2009/10	184	Y	18,361 (14,918)	97.7 (127.0)	10.7 (6.7)	21.3 (16.4)
2010/11	184	Y	15,635 (12,963)	91.1 (105.2)	11.0 (7.6)	20.8 (16.0)
2011/12	173	Y	14,660 (11,803)	91.3 (135.4)	11.2 (8.3)	22.0 (16.0)
2012/13	184	Y	15,063 (11,799)	85.0 (111.6)	11.5 (6.9)	21.9 (16.1)
2013/14	164	Y	15,042 (10,295)	92.4 (116.1)	12.1 (7.7)	22.3 (15.8)

Notes: Standard deviations in parentheses.

facilities after standardizing each facility's output at 100 at the beginning of the observed period.⁴ The dotted vertical line marks the introduction of the PBF scheme.

⁴Using a moving average instead of a simple one smooths the plots, especially in the case of the PBF facilities, which are few in number. Standardizing the output of each facility to 100 in fiscal year 2001/2002 prevents gaps in the data from affecting the average.

Figure 5.1 – Output trends



5.4 Empirical strategy

When asked to estimate the effect of the introduction of a new technology – such as PBF – on productivity and efficiency, a development microeconomist or a student of industrial organization might instinctively turn to estimating the parameters of a production function (with an additional technology-shifting term indicating whether or not PBF is in place) by means of regression analysis or some more general parametric approach such as the generalized method of moments (GMM). Yet, most such analyses of healthcare production typically involve data envelopment analysis (DEA) or stochastic frontier analysis (SFA) approaches (Hollingsworth, 2008). Each approach has its advantages as well as drawbacks.

DEA does not require one to make any distributional assumptions nor to impose a functional form on the production technology, but it is very sensitive to outliers and measurement error due to its non-stochastic nature. This is a problem especially in the context of developing countries where data are often sketchy at the best. SFA and various regression methods are considerably less sensitive to outliers and measurement error than DEA, but require one to impose assumptions on the distribution of efficiency or the error term and about the form of the production function. DEA and SFA are commonplace in the field of productivity and efficiency analysis, while the results of standard regression and GMM approaches can be more readily interpreted by a traditionally trained microeconomist. Moreover, GMM allows for healthcare production to be modeled as the dynamic process that it is (Scott and Coote, 2010), bringing the econometric model closer to reality.

In what follows, we start with a simple pooled ordinary least squares (OLS) model, gradually relaxing various assumptions to arrive to a dynamic system GMM model which fully accounts for the dynamics inherent to the healthcare production process. We then compare the results to those obtained from DEA and SFA methods.

5.4.1 Panel regression analysis and dynamic system GMM

We first estimate a simple pooled OLS model, regressing output on PBF and the equivalent of a translog production function with expenditures, number of staff and capital (proxied for bed capacity) as factors of production. We then correct a bias due to the presence of facility-level fixed effects by estimating a fixed-effects (FE) model, and we include fiscal year fixed effects to produce the following fully specified model:

$$\ln SUO_{it} = \alpha + \beta PBF_{it} + \gamma' \mathbf{factors}_{it} + \eta_i + \tau_t + \varepsilon_{it} \quad (5.2)$$

where PBF_{it} is a dummy equal to one if PBF is in place in the given facility and fiscal year. $\mathbf{factors}_{it}$ is a vector representing a translog production function with

$expenditures_{it}$ (total expenditures in millions of UGX), $staff_{it}$ (total number of staff), $beds_{it}$ (bed capacity proxying for capital) as production factors. η_i are facility-level fixed effects, τ_t are fiscal year fixed effects, and ε_{it} is a stochastic error term.⁵

Although we technically estimate the effect of PBF on output rather than on allocative efficiency, it has no consequence for the magnitude and statistical significance of the estimated effect β since all factors of production enter on the right-hand side of the regression equation. Substituting allocative efficiency for output on the left-hand side would only affect the values of γ , leaving β unaffected. We choose the present specification for the ease of its interpretation as a translog production function.

Since it can be expected that performance in one period is heavily influenced by performance in previous time periods (Scott and Coote, 2010), we reestimate the model using a three-step feasible generalized least squares estimator (FGLS) to correct for autocorrelation in the data.⁶

The static models developed so far may be further biased by imbalance of the panel—48% of observations are missing. This would not be a major issue if the observations were missing at random. It is however quite conceivable that better-administered units enter the dataset early on, potentially biasing down any time trend. Indeed, the unbalancedness of the panel stems mainly from an expansion of the dataset through time (there are 79 observations in 2001 against 164 in 2014), and to a much lesser extent through attrition and time-series gaps.

⁵The full model is thus specified as follows:

$$\begin{aligned} \ln SUO_{it} = & \alpha + \beta PBF_{it} + \gamma_1 \ln expenditures_{it} + \gamma_2 \ln staff_{it} + \gamma_3 \ln beds_{it} \\ & + \gamma_4 (\ln expenditures)_{it}^2 + \gamma_5 (\ln staff)_{it}^2 + \gamma_6 (\ln beds)_{it}^2 \\ & + \gamma_7 (\ln expenditures \times \ln staff)_{it} + \gamma_8 (\ln expenditures \times \ln beds)_{it} \\ & + \gamma_5 (\ln staff \times \ln beds)_{it} + \eta_i + \tau_t + \varepsilon_{it} \end{aligned}$$

⁶We test for serial autocorrelation following Wooldridge (2002). We verify the assumption of stationarity using a series of augmented Dickey-Fuller (ADF) tests such as the Fisher-type test suggested by Choi (2001) and that of Im et al. (2003).

We verify the possibility of non-random censoring using the procedure proposed by Nijman and Verbeek (1992) to test for non-random attrition in panels. The non-random imbalance of the panel makes it imperative to introduce dynamics into the model. Our final estimator is therefore a panel-robust two-step system GMM, which allows for correcting both autocorrelation and imbalances in the panel.⁷

Dynamic panel data analysis helps reduce – if not resolve – key econometric problems often arising from empirical studies that use conventional cross-sectional or time-series datasets. The large number of data points increases the efficiency of econometric estimates and, by utilizing information on both the inter-temporal dynamics and the individuality of the entities being investigated, it better controls for the effects of missing or unobserved variables (Hsiao, 2003). Also, by following facilities over a 13-year time span, we can construct a proper recursive structure to study the before-after effect, addressing concerns over the short-lived nature of PBF-induced increases in performance (Oxman and Fretheim, 2008; Maynard, 2012).

5.4.2 Data envelopment analysis

To verify the robustness of the results, we also estimate the effect of PBF on productivity following the more traditional approaches of DEA- and SFA-based two-stage procedures. The DEA approach, where efficiency scores are obtained non-parametrically through linear programming methods and regressed on facility characteristics, is popular because its non-stochastic first stage does not require one to make any assumptions about the functional form of production technology. Instead, the efficiency factor θ is defined as the output of each facility – or decision-making unit (DMU) in the parlance of DEA – relative to the output of a

⁷In order to eliminate selection bias (and serial autocorrelation), we follow Roodman (2009), using a panel-robust two-step Blundel-Bond system GMM estimator with forward orthogonal-deviations, both first differences and levels of the independent variables as standard instruments, and a GMM-style instrument that collapses all available lags of the lagged dependent variable for each time period into one moment.

virtual facility with the same levels of input, which in turn is a linear combination of the most efficient facilities in the dataset. As such, θ is a facility's current output expressed as a fraction of its maximum potential output given the current levels of inputs and maximum efficiency. Formally, we solve the following linear program for each facility in each fiscal year:

$$\begin{aligned}
 & \min_{\theta, \lambda} \theta, \\
 & \text{s.t. } -\frac{1}{\theta} SUO_i + \mathbf{SUO}'\lambda \geq 0, \\
 & \quad \mathbf{x}_i - \mathbf{X}'\lambda \geq 0, \\
 & \quad \mathbf{N}'\lambda = 1, \\
 & \quad \lambda \geq 0
 \end{aligned} \tag{5.3}$$

where $0 \leq \theta \leq 1$, SUO_i is the output of the i -th facility, \mathbf{x}_i is a vector of its inputs (*expenditures, beds, staff*), \mathbf{SUO} is a vector of outputs of all the N facilities in a given year, λ is an $N \times 1$ vector of constant weights, and \mathbf{N} is a $N \times 1$ vector of ones.

In the second stage, we estimate the effects of the introduction of PBF on θ , using the double-bootstrapped maximum likelihood truncated regression procedure proposed by Simar and Wilson (2007) to correct for biases arising from within-facility correlation of efficiency:

$$\theta_{it} = \alpha + \beta PBF_{it} + \delta_i + \tau_t + \varepsilon_{it} \tag{5.4}$$

where PBF_{it} equals 1 if the i -th facility had a PBF program in place in fiscal year t and 0 otherwise, δ_i are spatial fixed effects on the diocese level, and τ_t are fiscal year fixed effects.⁸

⁸To estimate the second stage double-bootstrapped maximum likelihood truncated regression, we adapt Stata code written by Wolszczak-Derlacz and Parteka (2011) and kindly provided by the authors.

5.4.3 Stochastic frontier analysis

The deterministic nature of DEA gives the first stage of this approach the advantage of not necessitating any functional or distributional assumptions, but it also renders it extremely sensitive to outliers. The issue is further aggravated in developing country contexts like ours, where data collection methods are often suboptimal, making the data exceptionally noisy. SFA – the most common alternative to DEA – is considerably less sensitive to outliers and noise in the data, but does require an explicitly defined production. We use the same translog production function as in (5.2):

$$\ln SUO_{it} = \alpha + \gamma' \mathbf{factors}_{it} + (v_{it} - u_{it}) \quad (5.5)$$

where u_{it} is measure of inefficiency (a Euclidean distance from the estimated production frontier) with a truncated normal distribution and v_{it} a normally distributed stochastic error term. Following Battese and Coelli (1995), we use a maximum likelihood random effects model to estimate the impact of PBF on u_{it} :

$$u_{it} = \zeta + \beta PBF_{it} + \delta_i + \tau_t + w_{it} \quad (5.6)$$

where w_{it} is a normally distributed stochastic error term and the rest of the notation is the same as above, so that:

$$\ln SUO_{it} = \alpha + \gamma' \mathbf{factors}_{it} + (v_{it} - (\zeta + \beta PBF_{it} + \delta_i + \tau_t + w_{it})) \quad (5.7)$$

5.5 Results

To estimate the effects of introducing PBF on healthcare output and, as we control for factors of production, also on the allocative efficiency of healthcare service delivery, we start with a naïve pooled OLS model (Table 5.2, column 1), which yields a large, positive and statistically very significant effect of PBF on output, but is potentially biased as it ignores the panel structure of the dataset. We therefore

proceed with estimating a fixed effects model (Table 5.2, column 2),⁹ additionally controlling for a possible time trend by including fiscal year fixed effects (Table 5.2, column 3).¹⁰ To correct for first-order autocorrelation,¹¹ we also estimate the model using an FGLS estimator (Table 5.2, column 4).

The static models produce large, significant effects of PBF on output and productivity. However, the panel is heavily unbalanced and autocorrelated.¹² To address these issues, our final estimator is a system GMM (Table 5.2, column 5). Moving from a naïve static pooled OLS model to a fully specified dynamic GMM one, assumptions are progressively relaxed and various problems corrected for, and the magnitude of the estimated effect of the introduction of PBF on healthcare output decreases drastically from a hardly believable value of 73% in model (1) to a more realistic 27% in model (5). Its statistical significance also decreases in the process, but always remains below the 5% level.

Result: The introduction of PBF increases output through improved allocative efficiency of healthcare provision by 27%, or over a third of a standard deviation.

As a robustness check, we also estimate the effect of PBF on productivity using the more traditional two-stage DEA and SFA approaches. The linear program

⁹A Wald test confirms the presence of facility-level fixed effects ($F(215, 1677) = 5.07$ is significant at the 1% level). A cluster-consistent Hausman-type test following Arellano (1993) reveals that a random effects model would be inconsistent ($\chi^2(10) = 28.251$ is significant at the 1% level).

¹⁰A Wald test confirms the joint significance of fiscal year fixed effects ($F(12, 215) = 12.65$ is significant at the 1% level).

¹¹Following Wooldridge (2002), we reject the null hypothesis of no first order autocorrelation in the panel ($F(1, 192) = 61.532$ is significant at the 1% level). We verify the assumption of stationarity through a series of ADF tests, all of which reject the null hypothesis of non-stationarity at the 1% significance level.

¹²Applying the procedure proposed by Nijman and Verbeek (1992), we find that the censoring of the data is non-random (see Table 5.A.3 in Appendix Appendix 5.A: for details).

¹³To avoid proliferation of instruments in the system GMM model, we take the spatial fixed effects one level up within their nested structure, using diocese- instead of facility-level fixed effects.

Table 5.2 – PBF effects on output and efficiency

	(1)	(2)	(3)	(4)	(5)
	OLS	FE	FE	FGLS	sysGMM
Dep. var.	$\ln SUO$	$\ln SUO$	$\ln SUO$	$\ln SUO$	$\ln SUO$
PBF	0.734*** (0.129)	0.737*** (0.123)	0.492*** (0.131)	0.509*** (0.095)	0.266** (0.130)
$\ln SUO_{it-1}$					0.377*** (0.053)
Translog inputs	Y	Y	Y	Y	Y
Facility FE	N	Y	Y	Y	N
Diocese FE ¹³	N	N	N	N	Y
Time FE	N	N	Y	Y	Y
SE clustered by facility	Y	Y	Y	N	N
Mean dep. var.	9.321	9.321	9.321	9.321	9.372
Std. dev. dep. var.	(0.699)	(0.699)	(0.699)	(0.699)	(0.684)
Obs.	1903	1903	1903	1896	1572
R^2	0.594	0.276	0.384		
Wald $\chi^2(230)$				11063***	
Instruments					52
$AR(2) z$					0.76
Sargan $\chi^2(8)$					7.30
Hansen $\chi^2(8)$					7.64
Diff. in Hansen $\chi^2(1)$					1.66

Notes: Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Inputs: *expenditures*, *staff*, *capacity*. See Table 5.A.1 in Appendix Appendix 5.A: for full results.

in the first DEA stage produces largely varied efficiency scores θ across the sample of facilities and years with mean value $\mu_\theta = 0.439$ and standard deviation $\sigma_\theta = 0.240$. In words, the health facilities have on average been producing only 44% of the output that they could potentially produce, given their inputs and the efficiency of the best-performing units.¹⁴ In the second double-bootstrapped maximum-likelihood stage, we estimate that the introduction of PBF increases efficiency θ by 20.1% points (i.e. by 45.8%) – or almost one standard deviation (Table 5.3, column 1).

Though estimated simultaneously, the SFA approach can effectively be thought of as a two-stage process which first estimates a production frontier corrected for inefficiency u , and then regresses u on PBF. We estimate that the introduction

¹⁴It is important to note that the magnitude of θ can easily be biased down by outliers. This does not however affect the relative changes due to technology shifters such as PBF.

Table 5.3 – PBF effects on output and efficiency – robustness

Dep. var.	(1)	(2)	
	DEA 2 nd stage	1 st stage	SFA 2 nd stage
	θ	$\ln SUO$	u
<i>PBF</i>	0.201*** (0.038)		-0.476** (0.241)
Translog inputs	Y	Y	Y
Time FE	Y	N	Y
Diocese FE	Y	N	Y
Mean dep. var.	0.439	9.321	0.817
Std. dev. dep. var.	(0.240)	(0.699)	(0.228)
Obs.	1903	1903	
Wald χ^2 (df)	1045.71 (31)	2877.92 (9)	

Notes: (Cluster) robust (bootstrapped) standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See Table 5.A.2 in Appendix Appendix 5.A: for full results.

of PBF reduces u by 0.476 (Table 5.3, column 2), which – by a simple arithmetic manipulation – is an equivalent of a 61.0% increase in θ .

The SFA procedure is more robust to noise in the data than DEA, but both techniques are likely to bias the coefficient estimates upwards due to their static nature. Indeed, they produce statistically significant results similar in direction and magnitude to those from regression estimations, but considerably larger in magnitude than estimates from system GMM.

5.6 Discussion

As demonstrated in Section 5.5, PBF led to a statistically and economically significant increase of allocative efficiency – and thus output – in the healthcare facilities where it was introduced. A dynamic model results into a much lower estimate (an increase of around 25%) than static models (an increase between 50% and 75%). This is consistent with Scott and Coote (2010), who assert that healthcare provision is an inherently dynamic system, with current output and efficiency predicting future output and efficiency. In a panel setting, static models which ignore this positive autocorrelation will overestimate the effects of

any positive external shock such as an introduction of a PBF scheme. Dynamic estimators like system GMM can overcome this issue and are thus more suitable for evaluating the impact of such shocks.

Although much smaller in magnitude than static estimation methods might suggest, a 27% increase in output is far from modest, especially in light of the fact that it is due to a financial incentive scheme worth only about 5% the total income of the participating facilities. It is quite plausible that such a sudden rise in output could come at the expense of its quality. To see whether this might be an issue, we conducted a two-wave patient satisfaction survey in the PBF facilities and in a sub-sample of those without PBF.¹⁵ Both survey waves took place after PBF had been implemented (in 2012 and 2014), so we cannot infer much about the direct effect of the introduction of PBF on the quality of services. The data nonetheless clearly show that perceived quality of healthcare did not decline between the two waves. In fact, patient satisfaction increased in both PBF and control facilities, and there is even weak evidence that it rose faster in the PBF units (see Appendix Appendix 5.B: for the full analysis).

Existing literature raises several other concerns about the potential pitfalls of introducing a PBF scheme:

Lu (1999) points out that incentivized output targets may encourage facility administrators to over-report output. While this is in general possible, we are confident that such misreporting is unlikely in our case. Firstly, the PBF scheme only adds a marginal amount of money to the budget. Most of the facilities' income is derived from fixed government subsidies and fee-for-service user payments. Any doctoring up of data would require compensating the parallel

¹⁵While patient satisfaction is subjective, several studies endorse its validity as an instrument for measuring quality of healthcare (Davies and Ware, 1988; Andaleeb, 2001; Johansson et al., 2002) and it has recently been placed at the core of policy recommendations regarding PBF in the United States (Ryan, 2009; Wharam et al., 2009). Leonard (2008) shows that satisfaction is jointly produced with quality during the course of a consultation and that patients respond to increased quality by being more likely to be satisfied. Moreover, patient satisfaction reflects both process quality and clinical quality (Marley et al., 2004), making it a good measure of the overall quality of healthcare delivery.

increase in user fees, which makes this type of fraudulent reporting as unlikely as it is expensive. Secondly, UCMB datasets have long been used at the national level as an advocacy device, requiring high levels of reliability, as well as an internal means of performance comparison across facilities, inducing all facilities to minimize unintended under-reporting regardless of the financing mechanism at play. Finally, unlike other healthcare providers in Uganda, UCMB engages in monthly data verification and monitoring activities throughout the country. To spot any plausible fallacies in the reporting, it also activated additional post-episode-of-care verifications at the community level for PBF facilities.

Another potential issue arises from the way in which the targets are aggregated from partial output indicators. If the relative weight given to individual indicators does not reflect their actual cost in terms of factors of production, an introduction of weighted output targets such as the present one defined in terms of the SUO could distort the balance of provided services, potentially resulting in the production services with negative marginal value (Wynia, 2009). This however seems unlikely in a developing-country context where the supply of healthcare services often falls short of demand. Indeed, the three output indicators which constitute the largest part of total SUO production in observed facilities – the number of *outpatients* (52% of total SUO production), *inpatients* (36%) and *immunizations* (8%) – all rose at roughly the same rate following the introduction of PBF (see Table 5.A.4 in Appendix Appendix 5.A:).

Furthermore, Frey and Jegen (2001) and Wharam et al. (2009) are concerned that the bureaucratization of healthcare delivery, which is to some extent necessary to implement a PBF scheme, may crowd out intrinsic motivation, inducing a decline in the professionalism and morale of medical staff. We do not have any quantitative data on staff motivation, but in-depth interviews we conducted with managerial as well as rank-and-file staff in the PBF facilities suggest quite the opposite—the additional salary bonuses which became possible thanks to PBF made staff feel appreciated for their work, motivating them to increase their effort.

Finally, Oxman and Fretheim (2008) warn against the danger of widening the already existing gap between poorly- and well-performing facilities. This could

become an issue in cases where PBF is used to reallocate an existing pot of funds in what could be described as a zero-sum game. In our case, however, the performance bonuses come from an external source, precluding the necessity to decrease funding of under-performing facilities in order to top up that of the more successful ones. Any improvements should therefore be Pareto-efficient in our case.¹⁶ It is nevertheless impossible to verify this in our present study given the small number of treated facilities. A larger trial – if not an implementation across the board – of the scheme would be needed in order to assert that no facility is worse off as a result of PBF.

5.7 Conclusion

Using a panel of output and expenditure data from small private not-for-profit healthcare facilities in Uganda, we estimate the contribution of performance-based financing towards achieving greater allocative efficiency. We rely on a dynamic panel estimation, verifying its robustness through more traditional two-stage DEA and SFA procedures to find that healthcare providers respond strongly to targets by increasing output through improved allocative efficiency. In our case, output rose by over 25%. The result, though lower than those from more traditional static models which are likely to be biased upwards, is economically significant, and the increased efficiency does not come at the expense of quality. This shows the potential of PBF in motivating healthcare staff to put additional effort into their work, thus improving the performance of underfunded healthcare systems in developing countries.

¹⁶This highlights a trade-off between Pareto-efficiency and financial sustainability of many PBF schemes—a topic that is very policy-relevant, but beyond the scope of this chapter.

Appendix 5.A: Tables

Table 5.A.1 – PBF effects on output – full regression specifications

	(1)	(2)	(3)	(4)	(5)
	OLS	FE	FE	FGLS	sysGMM
Dep. var.	$\ln SUO$	$\ln SUO$	$\ln SUO$	$\ln SUO$	$\ln SUO$
PBF	0.734***	0.737***	0.492***	0.509***	0.266**
	(0.129)	(0.123)	(0.131)	(0.0945)	(0.130)
$\ln SUO_{it-1}$					0.377***
					(0.053)
$\ln expenditures$	0.655***	0.402**	0.488**	0.500***	0.594***
	(0.149)	(0.158)	(0.195)	(0.073)	(0.111)
$\ln staff$	0.237	-0.334	-0.336*	-0.351***	0.077
	(0.202)	(0.210)	(0.184)	(0.108)	(0.143)
$\ln capacity$	0.011	-0.032	-0.074	-0.085	0.058
	(0.154)	(0.172)	(0.170)	(0.085)	(0.115)
$(\ln expenditures)^2$	-0.035	-0.047*	-0.054*	-0.055***	-0.046***
	(0.025)	(0.026)	(0.029)	(0.012)	(0.017)
$(\ln staff)^2$	0.097**	0.016	0.044	0.060*	-0.005
	(0.047)	(0.050)	(0.051)	(0.035)	(0.049)
$(\ln capacity)^2$	0.029	0.031	-0.001	0.018	-0.013
	(0.032)	(0.034)	(0.031)	(0.016)	(0.021)
$\ln expenditures \times \ln staff$	-0.046	0.110*	0.065	0.060*	0.013
	(0.056)	(0.062)	(0.059)	(0.036)	(0.053)
$\ln expenditures \times \ln capacity$	0.031	0.022	0.036	0.026	0.008
	(0.043)	(0.035)	(0.035)	(0.021)	(0.028)
$\ln staff \times \ln capacity$	-0.051	0.018	0.041	0.015	0.029
	(0.053)	(0.054)	(0.056)	(0.031)	(0.044)
Time FE	N	N	Y	Y	Y
Diocese FE	N	N	N	N	Y
SE clustered by facility	Y	Y	Y	N	N
Mean dep. var.	9.321	9.321	9.321	9.321	9.372
Std. dev. dep. var.	(0.699)	(0.699)	(0.699)	(0.699)	(0.684)
Obs.	1903	1903	1903	1896	1572
R^2	0.594	0.276	0.384		
Wald $\chi^2(230)$				11063***	
Instruments					52
$AR(2) z$					0.76
Sargan $\chi^2(8)$					7.30
Hansen $\chi^2(8)$					7.64
Diff. in Hansen $\chi^2(1)$					1.66

Notes: Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5.A.2 – PBF effects on output – full frontier analysis specifications

Dep. var.	(2) SFA	
	1 st stage $\ln SUO$	2 nd stage u
<i>PBF</i>		-0.476** (0.241)
$\ln expenditures$	0.670*** (0.094)	
$\ln staff$	0.094 (0.082)	
$\ln beds$	0.170 (0.125)	
$(\ln expenditures)^2$	-0.041** (0.016)	
$(\ln staff)^2$	0.110** (0.046)	
$(\ln beds)^2$	-0.013 (0.015)	
$\ln expenditures \times \ln staff$	-0.053 (0.044)	
$\ln expenditures \times \ln beds$	0.057** (0.029)	
$\ln staff \times \ln beds$	-0.045 (0.045)	
Time FE	N	Y
Diocese FE	N	Y
Mean dep. var.	9.321	0.817
Std. dev. dep. var.	(0.699)	(0.228)
Obs.	1903	
Wald $\chi^2(9)$	2877.92	

Notes: SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5.A.3 – Nijman-Verbeek tests for non-random censoring

	(1)	(2)	(3)	(4)
	FGLS	FGLS	sysGMM	sysGMM
Dep. var.	$\ln SUO$	$\ln SUO$	$\ln SUO$	$\ln SUO$
<i>PBF</i>	0.505*** (0.092)	0.421*** (0.123)	0.268** (0.131)	0.264** (0.130)
In previous FY	0.034** (0.017)		-0.007 (0.072)	
Number of FY		0.008* (0.004)		-0.002 (0.005)
Translog inputs	Y	Y	Y	Y
SE clustered by facility	Y	N	N	N
SE clustered by diocese	N	Y	Y	Y
Mean dep. var.	9.337	9.321	9.372	9.372
Std. dev. dep. var.	(0.691)	(0.699)	(0.684)	(0.684)
<i>N</i>	1817	1896	1572	1572
Wald χ^2	11118***	3452***		
Wald χ^2 df	229	41		
Instruments			53	53
<i>AR</i> (2) <i>z</i>			0.75	0.74
Sargan χ^2 (8)			7.35	7.66
Hansen χ^2 (8)			7.78	8.04
Diff. in Hansen χ^2 (1)			1.91	2.29

Notes: Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5.A.4 – PBF effects on individual output indicators

	(1) sysGMM	(2) sysGMM	(3) sysGMM	(4) sysGMM
Dep. var.	$\ln SUO$	$\ln outpatients$	$\ln inpatients$	$\ln immunizations$
PBF	0.266** (0.130)	0.276*** (0.100)	0.245 (0.197)	0.320*** (0.120)
Dep. var. $it-1$	0.377*** (0.053)	0.405*** (0.042)	0.475*** (0.074)	0.531*** (0.060)
Dep. var. $it-2$			0.103** (0.043)	
Translog inputs	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Diocese FE	Y	Y	Y	Y
Mean dep. var.	9.319	8.705	6.513	8.333
Std. dev. dep. var.	(0.705)	(0.689)	(1.078)	(0.869)
Obs.	1572	1572	1380	1572
Instruments	52	52	51	52
$AR(2) z$	0.76	0.72	0.19	-0.22
Sargan χ^2 (df)	7.30 (8)	5.15 (8)	4.02 (6)	11.43 (8)
Hansen χ^2 (df)	7.64 (8)	4.01 (8)	3.55 (6)	9.55 (8)
Diff. in Hansen $\chi^2(1)$	1.66	0.35	0.02	0.06

Notes: SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Inputs: *expenditures*, *staff*, *capacity*. An additional lag of the dependent variable ($\ln inpatients_{it-1}$) is needed in model (3) to satisfy identification requirements.

Appendix 5.B: Quality

To complement the input and output panel data, we also collected information about the quality of healthcare delivery in a sub-sample of the facilities. The data was only collected after the introduction of the PBF scheme, and it is therefore not suited for analyzing the impact of PBF on healthcare quality. It can however help us address some of the reservations one might have about an introduction of a scheme that only incentivizes quantity, not quality. The relevant analysis is presented in this appendix.

5.B.1 Data

While our sample contains facilities belonging to the same size class, we further use the output and expenditure data along with several other characteristics to match the six PBF facilities with six similar ones receiving fixed funding, identified among other facilities in linguistically and culturally affine areas. The matches are based on a propensity score calculated from a set of indicators not constituting a building block of the SUO, as measured in the last year prior to the introduction of the scheme.¹⁷ By matching on values collected before the implementation of PBF, we ensure that the propensity scores are not influenced by any potential confounding effects of the intervention.

To gauge the perceived quality of the services in these facilities, we adapted the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey (Centers for Medicare and Medicaid Services, 2011) to the Ugandan environment and administered it in two waves in 2012 and in 2014. While patient satisfaction is subjective, several studies endorse its validity as an instrument for measuring quality of healthcare (Davies and Ware, 1988; Andaleeb, 2001; Johansson et al., 2002) and it has recently been placed at the core of policy

¹⁷The matching characteristics include the facilities' income and expenditures, catchment population, number of staff, bed capacity, number of carried out diagnostic procedures and minor surgical operations, average length of stay of admitted patients, number of fatalities, and the availability of mental counseling.

recommendations regarding PBF in the United States (Ryan, 2009; Wharam et al., 2009). Leonard (2008) shows that satisfaction is jointly produced with quality during the course of a consultation and that patients respond to increased quality by being more likely to be satisfied. Moreover, patient satisfaction reflects both process quality and clinical quality (Marley et al., 2004), making it a good measure of the overall quality of healthcare delivery. To address concerns that perceived quality might be highly dependent on the relative differences vis-à-vis the nearest available alternatives, each of the units was further matched with the nearest similarly-sized public facility as well as a village half-way between the catchment areas of each private-public pair, resulting in a final sample of 24 facilities and 12 neighboring villages.

In the villages, we randomly selected 12 adult respondents from a previously recorded household census. At the facilities, accidental sampling was used instead, interviewing the first 10 people exiting the facility on an unannounced day. In total, 384 interviews were carried out in each wave, which – excluding respondents who had not visited the catholic facilities in the three years prior to the interview – resulted in 440 incidentally truncated interviews. From answers to a set of questions regarding various aspects of perceived quality of service, we factor out an index of perceived quality (*quality*), scaled from 0 (most dissatisfied) to 1 (most satisfied). Other personal-level confounding characteristics measured through the survey include a physical health index (*health*), an asset index (*assets*) approximating wealth,¹⁸ education (*primary*) – a dummy equal to one if the respondent completed primary education, sex (*female*) – a dummy equal to one if the respondent is female, and age (*age*) in years. Finally, *govt* is the mean perceived quality of the matched state-owned facility measured in the same way as *quality*.¹⁹ As can be seen in Table 5.B.1, the sample was well balanced across all the confounding characteristics (except for *govt*) in 2012, but completely

¹⁸The physical health index is based on the SF-12 Health Survey as proposed by Ware et al. (1996). The asset index is obtained by principal factor analysis following Sahn and Stifel (2003)

¹⁹We use the mean perceived quality of the main public facility rather than the quality perceived by each respondent, de facto controlling for the relative reputation of the public competitor.

Table 5.B.1 – Survey sample balance

	Control		PBF		Diff.	t
	Mean	Std. dev.	Mean	Std. dev.		
2012						
<i>quality</i>	0.877	0.098	0.830	0.134	-0.047	2.7825***
<i>health</i>	0.560	0.278	0.499	0.286	-0.061	1.5326
<i>assets</i>	0.267	0.186	0.268	0.200	0.001	-0.0420
<i>primary</i>	0.383	0.489	0.455	0.500	0.072	-1.0294
<i>female</i>	0.713	0.455	0.718	0.452	0.005	-0.0851
<i>age</i>	37.438	19.008	35.857	14.685	-1.581	0.653
<i>govt</i>	0.407	0.005	0.429	0.007	0.022	2.5974**
2014						
<i>quality</i>	0.904	0.101	0.903	0.110	-0.001	0.1028
<i>health</i>	0.607	0.280	0.624	0.278	0.017	-0.4787
<i>assets</i>	0.274	0.176	0.227	0.164	-0.047	2.1539**
<i>primary</i>	0.431	0.498	0.489	0.502	0.057	-0.8851
<i>female</i>	0.734	0.444	0.832	0.375	0.098	-1.8556*
<i>age</i>	41.633	16.562	35.827	15.580	-6.106	2.9377***
<i>govt</i>	0.362	0.004	0.390	0.006	0.028	3.8334***

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

unbalanced in 2014. We will address these issues below.

5.B.2 Empirical strategy

It is evident from the previous section that we only have data on perceived quality of 12 selected facilities from two survey waves which took place several years after the introduction of PBF in 6 of the units. Lacking a baseline survey, we clearly cannot estimate the effect of the introduction of PBF on the perceived quality of healthcare provision in a difference-in-differences (DID) setting. We can, however, use the DID approach to compare the trends in the PBF and non-PBF facilities after the new financing system had been introduced and after the management and staff of the facilities had presumably gotten accustomed to it.

We already outlined above our choice to investigate quality through a HCAHPS-type survey rather than through observable structural and process measures. A meaningful measure of perceived quality of the health facilities can only be obtained from respondents who had recently received treatment there.

This can potentially introduce a serious self-selection bias. Before proceeding to a full DID model specification for comparing the trends in perceived quality of healthcare delivery in PBF and non-PBF facilities, we therefore first estimate a bivariate sample-selection model as proposed by Heckman (1979) to check for such bias. We then estimate the PBF trend effects using a simple DID model with cluster-robust standard errors on the sub-sample of non-incidentally truncated interviews, gradually adding covariates until reaching the following fully specified model:

$$quality_{it} = \alpha + \beta_1 PBF_i + \beta_2 (PBF_i \times FU_t) + \beta_3 FU_t + \gamma' \mathbf{X}_i + \delta' \mathbf{Z}_{it} + \zeta_{govt_i} + \eta_i + \varepsilon_{it} \quad (5.8)$$

where FU is a dummy equal to one for all observations from the 2014 follow-up, \mathbf{X} is a vector of production factors proxying the size of the private facility (*expenditures*, *staff* and *beds*), \mathbf{Z} is a vector of respondent characteristics comprising *health* (a physical health index), *assets* (an asset index approximating wealth), *primary* (a dummy equal to one if the respondent completed primary education), *female* (a dummy equal to one if the respondent is female), and *age* (respondent's age in years). *govt* is the mean perceived quality of the nearby state-owned health facility, η are location type fixed effects indicating whether the interview took place at a private facility, a public facility, or a village in between the two, and ε is a stochastic error term.

Finally, to account for possible biases resulting from the imbalance in the 2014 sample, we further estimate a kernel propensity score DID model using \mathbf{Z} and *govt* to compute the propensity scores.

5.B.3 Results

In this section, we compare the trends in the perceived quality of healthcare provision in PBF and non-PBF facilities following the introduction of the new financing system in the former. Before proceeding to a full DID model specification, we first estimate a Heckman bivariate sample-selection model to check for self-selection into visiting the private facilities. While some of the

observed respondent characteristics affect the likelihood of visiting the private facility within 3 years prior to the interview, the coefficient on the inverse Mills ratio is statistically insignificant (Table 5.B.2, column 1). In other words, the self-selection bias in the sample does not affect the coefficient estimates in the second stage. It is therefore safe to estimate the PBF trend effects using a simple DID model with cluster-robust standard errors on the sub-sample of non-incidentally truncated interviews.

Table 5.B.2 – PBF effects on trend in perceived quality of healthcare—regressions

		(1) Heckman	(2) OLS	(3) OLS	(4) OLS
Dep. var.	<i>private</i>	<i>quality</i>	<i>quality</i>	<i>quality</i>	<i>quality</i>
$PBF \times FU$		0.037 (0.022)	0.038** (0.016)	0.035** (0.016)	0.038** (0.015)
PBF		-0.045*** (0.017)	-0.047 (0.026)	-0.040 (0.026)	-0.035 (0.026)
FU		0.031* (0.016)	0.030** (0.011)	0.028* (0.013)	0.031* (0.014)
Facility size	N	Y	Y	Y	Y
Respondent characteristics	Y	N	N	Y	Y
Competitor quality	Y	N	N	N	Y
Location type FE	N	N	N	N	Y
SE clustered by facility pair	N	N	Y	Y	Y
Mean dep. var.		0.880	0.880	0.880	0.880
Std. dev. dep. var.		(0.115)	(0.116)	(0.115)	(0.115)
N	743	430	440	430	430
Mills λ		-0.027 (0.050)			
R^2			0.069	0.082	0.085

Notes: Robust SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Starting with a naïve specification which only includes the DID terms and size controls (Table 5.B.2, column 2), we gradually include respondent characteristics (Table 5.B.2, column 3), the mean level of perceived quality of the nearest similarly-sized public facility (Table 5.B.2, column 4), and controls for the location and timing of the interview (Table 5.B.2, column 5). To address the imbalance in the 2014 sample, we also estimate a kernel propensity score DID model using respondent characteristics and public competitor quality to compute the propensity scores and controlling for the remaining covariates from the full DID

OLS model (Table 5.B.3).

Table 5.B.3 – PBF effects on trend in perceived quality of healthcare—PSM

	2012			2014			DID
	Control	PBF	Diff.	Control	PBF	Diff.	
<i>quality</i>	0.851	0.889	-0.038	0.928	0.920	-0.008	0.030
SE			(0.030)			(0.023)	(0.019)

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The estimates of perceived quality are stable across the specifications. Perceived quality of healthcare increased between 2012 and 2014 in the PBF as well as non-PBF facilities (see coefficients on FU in Table 5.B.2). There is also some evidence that the increase was much more pronounced in the case of PBF facilities (see coefficients on $PBF \times FU$ in Table 5.B.2), though the statistical significance of this relationship falls below the 10% level when we account for self-selection into attending a private facility by means of a PSM DID estimation (see column DID in Table 5.B.3). It is nevertheless clear that PBF did not lead to a decline in patient satisfaction between the two survey waves.

Chapter 6

Discussion and Conclusions

6.1 General discussion

In his *Wealth of Nations*, Adam Smith (1776) laid the foundations of modern economics by noting that individuals cooperate not out of regard for each other's wellbeing, but for their self-love. In doing so, Smith did not, however, reject the idea that people care about the welfare of others. Quite to the contrary, he wrote in *The Theory of Moral Sentiments*:

That we often derive sorrow from the sorrows of others, is a matter of fact too obvious to require any instances to prove it; for this sentiment, like all the other original passions of human nature, is by no means confined to the virtuous or the humane, though they perhaps may feel it with the most exquisite sensibility. The greatest ruffian, the most hardened violator of the laws of society, is not altogether without it.
(Smith, 1761, p.2)

Economists have thus been well aware of the basic premise of prosocial preferences by the time their discipline began its split from philosophy. For a long time however, they tended to treat prosocial preferences – like any other preferences – as traits which may form important components of economic models, but ones which are exogenous to these models, and whose emergence and formation should be the concern of other scientific domains (Friedman, 1962, p.13).

It is therefore not surprising that the first systematic attempts to understand how preferences are shaped were not made by economists, but by psychologists. Perhaps most notably, the (radical) behaviorists such as B. F. Skinner (1938) set out to analyze the sources of human preferences in a quantitative way. Considering the inner causes of behavior too vague to be precisely observed, Skinner instead turned to the environment as an important behavioral determinant:

The practice of looking inside the organism for an explanation of behavior has tended to obscure the variables which are immediately available for a scientific analysis. These variables lie outside the organism, in its immediate environment and in its environmental history. They have a physical status to which the usual techniques

of science are adapted, and they make it possible to explain behavior as other subjects are explained in science. These independent variables are of many sorts and their relations to behavior are often subtle and complex, but we cannot hope to give an adequate account of behavior without analyzing them. (Skinner, 1953, p.31)

Significant progress in understanding the physiological workings of the human mind has been made since Skinner's time, thanks largely to advances in various fields of neuroscience (see e.g. Camerer et al., 2005; Keverne and Curley, 2008). Considering the ever more rapid changes in both our physical and social environment, Skinner's focus on finding the sources of human behavior outside the body and mind is nevertheless as relevant now as it was in the heyday of radical behaviorism. Indeed, economics has taken some clues from this approach in recent years, investigating the effects of environmental shocks on a range of economic outcomes (see e.g. Almond, 2006; Dell et al., 2009; Burke et al., 2015). While certainly very informative, most of this literature only brushes over the behavioral mechanisms which lead to the outcomes that are its main focus.¹ By investigating the effects of changes in the physical and social environment on individual prosocial preferences, this thesis fills a part of the gap.

The word "environment" as it is used throughout this thesis should be understood in its broad sense as the aggregate of surrounding things, conditions, and influences including the social and cultural forces that shape the life of a person or a population (Dictionary.com, 2017), rather than in the stricter ecological sense, in which it is commonly used. It encompasses purely physical conditions such as ambient temperature (Chapter 2), the purely social ones in the form of institutional arrangements (Chapters 4 and 5) as well as their intersection, where social conditions – armed conflict – affect the physical conditions faced by unborn children (Chapter 3).

There are almost certainly other environmental conditions that affect prosocial (as well as other) preferences. Rather than to investigate them all, this thesis is

¹Notable exceptions include, among others, Bowles (1998), Weinstock (2008), Voors et al. (2012) or Cecchi et al. (2016).

intended to illustrate their diversity. In doing so, I hope to achieve two long-term goals. First, this thesis should serve as a springboard for further research into the ways in which environmental changes shift individual preferences and thus people's behavior. Second, its findings, as well as those of related future research, should be taken into account by policymakers whose decisions affect our environment. In what follows, I review the main lessons learned from the research that underpins this thesis, their policy implications, and the potential directions of future scientific inquiry that they show.

6.2 Prenatal temperature shocks

The work of the late physician and epidemiologist David J. Barker (1990), who proposed that adult disease may often originate in adverse environmental conditions during gestation, has inspired a large volume of research which explores its economic implications. In Chapter 2, I build upon the fetal origins literature, testing the hypothesis that the consequences of prenatal events may extend beyond effects on later-life individual abilities and health trajectories (Almond and Currie, 2011), also affecting social preferences. In particular, I investigate the impact of temperature shocks faced by pregnant women on their children's later-life prosocial preferences. I find that exposure to higher than usual ambient temperatures during gestation reduces a child's probability of contribution to the public good, with the negative effect lasting into adulthood.

This finding empirically links a chain of already established relationships. Ambient temperature shocks during gestation have been shown to negatively affect birth weight (Wells and Cole, 2002; Lawlor et al., 2005; Deschênes et al., 2009). Low weight at birth predicts – among many other later-life outcomes – relatively lower cognitive abilities (Hack et al., 2005), and those in turn reduce the taste for cooperation (Moore et al., 1998; Zhang et al., 2015). Taken together, these relationships suggest that a connection between prenatal ambient temperature and prosocial preferences should exist. Nonetheless, this thesis is, to the best of my knowledge, the first to provide explicit evidence of this link, albeit in a reduced

form. The next logical step is the development and estimation of a structural form model. Such a study will, by its nature, require a much more complete dataset than mine. At the bare minimum, it will need to make use of medical birth records, and more precise measurements of cognitive ability.

More generally, there is ample scope for research into the exact nature of the mechanisms which are at play here. There is consensus on the fact that temperature shocks negatively affect birth outcomes, but not on the physiological processes involved. Reduced fetal nutrient intake caused either directly by altered maternal nutrient intake or indirectly by impaired placental function is often mentioned as a potential mechanism, as are hormonal imbalances caused by heat shocks. There is some evidence that these mechanisms are at play in other mammals such as sheep (Hansen, 2009; Bloomfield et al., 2003), but studies of their role in human gestation are rare and inconclusive (Yajnik, 2002; Smith and Nicholson, 2007). The mechanisms linking poor birth outcomes with cognitive ability are perhaps even less explored (Aarnoudse-Moens et al., 2009). Considering the ethical issues which render most experimental study designs off limits to fetal and perinatal pathology, there may be scope for medical science to get methodologically inspired by applied economics and its frequent use of innovative econometric techniques and natural experiments.

Regardless of the underlying processes, the finding that prenatal temperature shocks reduce later-life cooperation fits into a broader and growing body of literature which links ambient temperature to various economically important outcomes. Temperature has been found to influence the level of economic production (Dell et al., 2009; Burke et al., 2015), the onset of conflict (Hsiang et al., 2013), mental health (Adhvaryu et al., 2015), and even human reproductive behavior—with consequences for physical health and educational outcomes of the offspring (Wilde et al., 2017). These findings add weight to the (already important) issues typically associated with climate change such as decreasing biodiversity, sea level rise, or food and water security, and should be considered alongside them when assessing the costs and benefits of policies to tackle climate change.

6.3 Prenatal stress

Chapter 3 continues in the same vein as Chapter 2, building upon the fetal origins literature. It investigates how prenatal stress induced by random violence affects the preferences for cooperation among children born during an armed conflict, exploiting variations in the 2D:4D digit ratio—a marker of in utero hormone exposure negatively associated with high maternal distress during early fetal development. In this chapter, I show that prenatal stress reduces the probability that children contribute to the public good.

Similarly to the case of prenatal temperature shocks, there are empirically established relationships which together suggest that a causal link between prenatal exposure to conflict and later-life prosocial preferences should exist. Firstly, exposure to violent conflict is a traumatizing experience (de Jong, 2002; Lopes Cardozo et al., 2004; Miller et al., 2008). Secondly, a mother's traumatization during pregnancy affects the hormonal balance of the fetus (Ward and Weisz, 1980; vom Saal et al., 1990) as observable through its 2D:4D digit ratio (Manning et al., 2003; Lutchmaya et al., 2004; Lilley et al., 2010; Zheng and Cohn, 2011). Thirdly – and most importantly – in utero hormonal balance affects later-life social preferences (Buser, 2012; Brañas-Garza et al., 2013). To the best of my knowledge, these three relationships have not been analyzed concurrently before. By doing so, this research provides evidence which, for the first time, supports the entire causal chain from conflict in one generation to economic behavior in the next one.

This finding has several ramifications for both theoretical and applied economic research. Moving away from the belief that preferences are hardwired (Stigler and Becker, 1977), most economists now accept the idea that they are molded by the environment in which an individual operates as well as by his genetic code. Several models of preference transmission have been developed (Bisin and Verdier, 2001; Bisin et al., 2004; Bulte and Horan, 2011; Doepke and Zilibotti, 2017), and the debate now centers more on the relative importance of nurture and nature (see e.g. Cesarini et al., 2008, 2009). The notion that prenatal stress affects later-life

preferences introduces an additional channel of preference transmission—one that lies between the nature-nurture duality. Further research should test whether these effects persist throughout life (like those of prenatal temperature shocks discussed in Chapter 2), and verify their generalizability beyond the Ugandan context. If confirmed in other settings, the epigenetic path should be considered in future inquiries into intergenerational transmission of preferences.

The policy implications of the idea that conflict events may shape the preferences of a generation yet unborn are twofold. First, the definition of a victim of war may need to be expanded to include those who were traumatized before birth. In practice, this means that organizations which intervene in post-conflict situations to provide psychosocial support or help (re-)build positive intra-community relationships may need to expand their target population as well as the time horizon of their interventions. Second, the design of such interventions should take into account the heterogeneous effects that exposure to conflict has on different victims. While prenatal conflict exposure reduces the taste for cooperation, postnatal exposure fosters it, at least within one's social circles (Voors et al., 2012; Gilligan et al., 2014). Recognizing these differences may help policymakers fine-tune post-conflict interventions to better suit the composition of their target population.

6.4 Institutional change

In Chapter 4, I move away from the physical aspects of human environment, focusing instead on the social ones. I study the effects of a sudden introduction of a formal institution on individual cooperative behavior within informal arrangements. In particular, I look at how an NGO intervention which helped create a mutual health insurance affected cooperative behavior in a public goods game. In contrast with cases where outside institutional innovations increased cooperation (Fearon et al., 2009), but in line with evidence that formal insurance crowds out informal sharing (Klohn and Strupat, 2013), I find that the introduction of formal insurance reduces contributions to the public good. This reduction in

cooperation levels is, however, not due to the adopters of the formal insurance who may now have less need for informal reciprocal networks, and who therefore (partially) withdraw from them. It is instead the non-adopters who become less cooperative towards the adopters.

The results are in line with sociological theories of social embeddedness (see e.g. Polanyi, 1944; Granovetter, 1985) and the ‘weapons of the weak’ perspective advanced by the political scientist and anthropologist James Scott (1985). Viewed through this lens, the behavior of those who did not adopt the insurance towards those who did can be interpreted as a signal of the former’s displeasure with the new institution in which they do not want to – or more likely cannot afford to – participate, and which may be undermining informal risk-sharing arrangements. While anecdotal evidence from the field supports this interpretation, several alternative explanations which also fit the data cannot be ruled out. For example, if the uptake of insurance is understood as an indication of wealth on the part of the adopters, lower public good contributions by non-adopters could be due to the activation of a norm about people contributing to local public goods according to their means relative to each other (Ensminger, 2004). Alternatively, non-adopters may reallocate their contributions to other remaining non-adopters in expectation of these relationships taking on a more important role now that the adopters have an alternative to informal risk-sharing arrangements. Finally, it is possible that the way the insurance product was marketed encouraged individualistic thinking, widening the aspiration gap of those who could not afford to purchase it. In response, non-adopters may have simply reallocated the means at their disposal from contributing to the public good towards trying to close this gap. Disentangling the merits of these alternative explanations should be the subject of future research.

The fact that a well-intended institutional innovation – such as the introduction of mutual health insurance to a previously uninsured community – can negatively impact social capital and cooperation, has policy implications that are perhaps obvious, but important to point out. It epitomizes two common shortcomings of many development interventions. First, even interventions with clearly defined

and well justified objectives – such as addressing the failure of markets to provide functional insurance – can have undesirable consequences unforeseen by the policymakers. It is important that they pay closer attention to the potential externalities of their interventions, taking into account lessons learned from previous successes and especially from previous failures. Second, too often do development interventions conform to the funding cycle rather than to the needs of their beneficiaries. The intervention on which this case study is based was projected to last 5 years, for example, with no plans for future funding. The immediate benefits of donor-subsidized insurance may well have outweighed the associated social costs, but the former were likely to end with the intervention, while the latter may linger on for much longer. To establish how long the social costs persist would necessitate a follow-up study, which would ironically require a new round of funding. If this situation is to be ameliorated, it is important that development organizations plan their projects for the long term, follow up on their beneficiaries even after the intervention ended, and be ready to come back in case of unforeseen complications. For their part, donors should not only encourage such behavior, but make their funding conditional on it.

6.5 Extrinsic motivation

Chapters 2, 3, and 4 illustrate how changes in both the physical and social environment can be detrimental to individual prosocial behavior. Perhaps the most obvious way to reverse such deterioration of prosociality (to the extent that it is desirable—an assertion that few would dispute) would be to address its root causes, that is to reverse the environmental changes themselves. This might not, however, be the most practicable solution. While there is broad scientific – and increasingly also political – consensus on the need to slow down and ideally reverse climate change, the complexity of the task makes it one with which humanity will likely keep grappling for a while. The elimination of armed conflict is an even more elusive goal if history is any guidance. Finally, reversing institutional changes may often not be desirable at all, considering that their intended outcomes are often generally beneficial to society. This is of course not to say that research

focused on tackling climate change, preventing conflict, and reducing the negative externalities of institutional change should be discouraged. It does however mean that such inquiry is well beyond the scope of this thesis.

Chapter 5 instead illustrates one of the many ways in which prosocial behavior can be incentivized with a relatively simple and easily implementable policy. In this chapter, I evaluate the impact of introducing performance-based financial incentives on staff effort and, consequently, on allocative efficiency and output in healthcare provision. I show that in the case under investigation, financial incentives conditioned on output and worth roughly 5% of total expenditures increased staff effort to the extent that output rose by over 25%, without any detectable drop in the quality of the provided services. This not only shows the potential of performance based financing to improve the performance of underfunded healthcare systems in developing countries, but also that extrinsic motivation can be used to foster behavior which benefits the society rather than just the individual.

The simplicity and potentially high returns on investment of incentivizing prosocial behavior with financial (or other) rewards make it an appealing policy tool. However, there are circumstances in which putting a price tag on prosocial behavior can actually discourage people from engaging in it (see e.g. Titmuss, 1970; Ariely et al., 2009; Lacetera and Macis, 2010a,b; Gneezy et al., 2011). In fact, even interventions similar to the one described here have been unsuccessful in many cases (see Eijkenaar et al., 2013, for an overview). More research into the conditions under which extrinsic motivation succeeds in fostering prosocial behavior is therefore needed. Until more progress is made in this direction, policymakers should tread carefully when setting up incentive structures to promote prosociality. In this respect, small scale pilots with impact evaluation integrated into their design are an excellent way to establish best practices without making big, expensive mistakes.

6.6 Concluding remarks

This thesis identifies several ways in which environmental changes affect intrinsic prosocial preferences, and outlines a possible direction for ameliorating any such negative effects. While it provides some new insight into the complexity of the formation of social preferences, it leaves many questions unanswered, and raises several new ones.

As pointed out in the preceding 4 sections of this chapter, many of the mechanisms that are behind the relationships described in this thesis are poorly understood, and deserve the attention of the scientific community. Moreover, it is likely that this thesis only scratches the surface when it comes to environmental effects on preferences. The full range of environmental factors involved is unclear. Could pollution, for example, affect our behavior? Or could it be shaped by the political system in which we grow up? Questions also remain about the timing of these effects. I show that prenatal temperature shocks reduce cooperativeness, but what is their effect in early life or in adulthood? Finally, what other preferences might be affected? I focus on prosociality as one of the building blocks of human society, but if prosocial preferences can be affected by environmental change, why not risk and time preferences, which are also fundamental to our everyday decisions and – in their aggregate – to the economy at large? The scope for further research is immense, and much of it will have to cut across disciplines if it is to fully answer these questions.

The subject of this thesis – social preferences – is of fundamental interest to economics. However, I start by borrowing the words of a theoretical physicist about the necessity to work together, and go on to draw on the findings from many other disciplines ranging from social anthropology and psychology to neurology and endocrinology. While writing the thesis, I benefited from conversations with biologists, sociologists, physicians, chemical, electrical and agricultural engineers, anthropologists, agronomists, psychologists, climate scientists, nutritionists, and – of course – economists. To me, there is no better illustration of the importance of interdisciplinary communication. In fact, if I could start all over again, I would

strive to not only communicate, but also collaborate more with people outside of my own discipline. With this in mind, I find it fitting to end by quoting another physicist, the great Isaac Newton (1675): “If I have seen further, it is by standing on the shoulders of giants.”

References

- Aarnoudse-Moens, C. S., N. Weisglas-Kuperus, J. B. van Goudoever, and J. Oosterlaan (2009). Meta-Analysis of Neurobehavioral Outcomes in Very Preterm and/or Very Low Birth Weight Children. *Pediatrics* 124(2), 717–728.
- Adhvaryu, A., J. Fenske, N. Kala, and A. Nyshadham (2015). Fetal Origins of Mental Health: Evidence from Africa. CSAE Working Paper WPS/2015-15.
- Alesina, A. and P. Giuliano (2015). Culture and Institutions. *Journal of Economic Literature* 53(4), 898–944.
- Alger, I. and J. W. Weibull (2010). Kinship, Incentives, and Evolution. *American Economic Review* 100(September), 1725–1758.
- Almond, D. (2006). Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population. *Journal of Political Economy* 114(4), 672–712.
- Almond, D. and J. Currie (2011). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives* 25(3), 153–172.
- Almond, D. and B. Mazumder (2011). Health Capital and the Prenatal Environment: The Effect of Ramadan Observance during Pregnancy. *American Economic Journal: Applied Economics* 3(4), 56–85.
- Alpizar, F., F. Carlsson, and M. A. Naranjo (2011). The Effect of Ambiguous Risk, and Coordination on Farmers’ Adaptation to Climate Change—A Framed Field Experiment. *Ecological Economics* 70(12), 2317–2326.
- Andaleeb, S. S. (2001). Service Quality Perceptions and Patient Satisfaction: a

- Study of Hospitals in a Developing Country. *Social Science & Medicine* 52(9), 1359–70.
- Annan, J., C. Blattman, and R. Horton (2006). The State of Youth and Youth Protection in Northern Uganda: Findings from the Survey of War Affected Youth. Report for UNICEF Uganda.
- Aoki, M. (2001a). Community Norms and Embeddedness: A Game Theoretic Approach. In M. Aoki and Y. Hayami (Eds.), *Communities and Markets in Economic Development*, Chapter 4, pp. 97–129. Oxford: Oxford University Press.
- Aoki, M. (2001b). *Towards a Comparative Institutional Analysis*. Cambridge, MA: MIT Press.
- Aquinas, T. (1274). *Summa Theologica*.
- Arellano, M. (1993). On the testing of correlated effects with panel data. *Journal of Econometrics* 59(1-2), 87–97.
- Ariely, D., A. Bracha, and S. Meier (2009). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review* 99(1), 544–555.
- Axelrod, R. and W. D. Hamilton (1981). The Evolution of Cooperation. *Science* 211(4489), 1390–1396.
- Bähre, E. (2011). Liberation and Redistribution: Social Grants , Commercial Insurance, and Religious Riches in South Africa. *Comparative Studies in Society and History* 53(2), 1–22.
- Banerjee, A., E. Duflo, G. Postel-Vinay, and T. Watts (2010). Long-Run Health Impacts of Income Shocks: Wine and Phylloxera in Nineteenth-Century France. *Review of Economics and Statistics* 92(4), 714–728.
- Banerjee, A. V., R. Glennerster, and E. Duflo (2008). Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System. *Journal of the European Economic Association* 6, 487–500.
- Barker, D. J. P. (1990). The Fetal And Infant Origins of Adult Disease: The Womb

- May Be More Important than the Home. *British Medical Journal* 301(6761), 1111.
- Barr, A., T. Packard, and D. Serra (2014). Participatory Accountability and Collective Action: Experimental Evidence from Albania. *European Economic Review* 68, 250–269.
- Barreca, A. I. (2010). The Long-Term Economic Impact of In Utero and Postnatal Exposure to Malaria. *Journal of Human Resources* 45(4), 865–892.
- Basinga, P., P. J. Gertler, A. Binagwaho, A. L. B. Soucat, J. Sturdy, and C. M. J. Vermeersch (2011). Effect on Maternal and Child Health Services in Rwanda of Payment to Primary Health-Care Providers for Performance: An Impact Evaluation. *Lancet* 377(9775), 1421–1428.
- Battese, G. E. and T. J. Coelli (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics* 20, 325–332.
- Bauer, M., A. Cassar, J. Chytilová, and J. Henrich (2014). War’s Enduring Effects on the Development of Egalitarian Motivations and In-group Biases. *Psychological science* 25(1), 47–57.
- Becker, G. S. (1993). Nobel Lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy* 101(3), 385–409.
- Beckett, C., B. Maughan, M. Rutter, J. Castle, E. Colvert, C. Groothues, J. Kreppner, S. Stevens, T. G. O’Connor, and E. J. S. Sonuga-Barke (2006). Do the Effects of Early Severe Deprivation on Cognition Persist Into Early Adolescence? Findings From the English and Romanian Adoptees Study. *Child Development* 77(3), 696–711.
- Bellows, J. and E. Miguel (2009). War and Local Collective Action in Sierra Leone. *Journal of Public Economics* 93(11-12), 1144–1157.
- Bénabou, R. and J. Tirole (2006). Incentives and Prosocial Behavior. *American Economic Review* 96(5), 1652–1678.
- Bhutta, A. T., M. A. Cleves, P. H. Casey, M. M. Cradock, and K. J. S. Anand

- (2002). Cognitive and Behavioral Outcomes of School-Aged Children Who Were Born Preterm. *JAMA* 288(6), 728–737.
- Binswanger, H. P. (1980). Attitudes toward Risk: Experimental Measurement In Rural India. *Agricultural and Applied Economics Association* 62(3), 395–407.
- Bisin, A., G. Topa, and T. Verdier (2004). Cooperation as a Transmitted Cultural Trait. *Rationality and Society* 16(4), 477–507.
- Bisin, A. and T. Verdier (2001). The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory* 97, 298–319.
- Björkman, M. and J. Svensson (2009). Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda. *Quarterly Journal of Economics* 124(2), 735–769.
- Black, S. E., P. J. Devereux, and K. G. Salvanes (2007). From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes. *Quarterly Journal of Economics* 122(1), 409–439.
- Black, S. E., P. J. Devereux, and K. G. Salvanes (2016). Does Grief Transfer Across Generations? Bereavements during Pregnancy and Child Outcomes. *American Economic Journal: Applied Economics* 8(1), 193–223.
- Blanchard, E. B., J. Jones-Alexander, T. C. Buckley, and C. A. Forneris (1996). Psychometric Properties of the PTSD Checklist (PCL). *Behaviour Research and Therapy* 34(8), 669–673.
- Blattman, C. (2009). From Violence to Voting: War and Political Participation in Uganda. *American Political Science Review* 103(02), 231–247.
- Blattman, C. and J. Annan (2010). The Consequences of Child Soldiering. *Review of Economics and Statistics* 92(4), 882–898.
- Bloom, D. E. and D. Canning (2000). The Health and Wealth of Nations. *Science* 287(5456), 1207–1209.
- Bloomfield, F. H., M. H. Oliver, P. Hawkins, M. Campbell, D. J. Phillips, P. D. Gluckman, J. R. G. Challis, and J. E. Harding (2003). A Periconceptual Nutritional Origin for Noninfectious Preterm Birth. *Science* 300, 606.

- Bonfrer, I., E. Van de Poel, and E. Van Doorslaer (2014). The Effects of Performance Incentives on the Utilization and Quality of Maternal and Child Care in Burundi. *Social Science & Medicine* 123, 96–104.
- Bouma, J., E. Bulte, and D. van Soest (2008). Trust and Cooperation: Social Capital and Community Resource Management. *Journal of Environmental Economics and Management* 56(2), 155–166.
- Bowles, S. (1998). Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions. *Journal of Economic Literature* 36(1), 75–111.
- Bowles, S. (2008). Policies Designed for Self-Interested Citizens May Undermine “The Moral Sentiments”: Evidence from Economic Experiments. *Science* 320(5883), 1605–1609.
- Bowles, S., J. K. Choi, and A. Hopfensitz (2003). The Co-Evolution of Individual Behaviors and Social Institutions. *Journal of Theoretical Biology* 223, 135–147.
- Bowles, S. and H. Gintis (2011). *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton, NJ: Princeton University Press.
- Brañas-Garza, P., J. Kovářík, and L. Neyse (2013). Second-to-Fourth Digit Ratio Has a Non-Monotonic Impact on Altruism. *PLoS ONE* 8(4), 1–10.
- Brenzel, L. (2009). Taking Stock: World Bank Experience with Results-Based Financing for Health. World Bank Technical Brief.
- Buckholtz, J. W. and R. Marois (2012). The Roots of Modern Justice: Cognitive and Neural Foundations of Social Norms and Their Enforcement. *Nature Neuroscience* 15(5), 655–61.
- Buckles, K. S. and D. M. Hungerman (2013). Season of Birth and Later Outcomes: Old Questions, New Answers. *Review of Economics and Statistics* 95(3), 711–724.
- Bulte, E. H. and R. D. Horan (2011). Intergenerational Transmission of Preferences. *Economics Letters* 112, 85–87.
- Burke, M., S. M. Hsiang, and E. Miguel (2015). Global Non-Linear Effect of

- Temperature on Economic Production. *Nature*.
- Buser, T. (2012). Digit Ratios, the Menstrual Cycle and Social Preferences. *Games and Economic Behavior* 76(2), 457–470.
- Callen, M., M. Isaqzadeh, J. D. Long, and C. Sprenger (2014). Violence and Risk Preference: Experimental Evidence from Afghanistan. *American Economic Review* 104(1), 123–148.
- Camacho, A. (2008). Stress and Birth Weight: Evidence from Terrorist Attacks. *American Economic Review* 98(2), 511–515.
- Camerer, C., G. Loewenstein, and D. Prelec (2005). Neuroeconomics: How Neuroscience Can Inform Economics. *Journal of Economic Literature* 43(1), 9–64.
- Cameron, A. C., J. Glebach, and D. L. Miller (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics* 90(3), 414–427.
- Cárdenas, J. C., A. Chong, and H. Ñopo (2009). To What Extent Do Latin Americans Trust, Reciprocate, and Cooperate? Evidence from Experiments in Six Latin American Countries. *Economía* 9(2), 45–94.
- Cassar, A., P. Grosjean, and S. Whitt (2013). Legacies of Violence: Trust and Market Development. *Journal of Economic Growth* 18(3), 285–318.
- Cassill, D. L., A. Casella, J. Clayborn, M. Perry, and M. Lagarde (2015). What Can Ants Tell Us about Collective Behavior During a Natural Catastrophe? *Journal of Bioeconomics* 17(3), 255–270.
- Cecchi, F. and J. Duchoslav (2017). The Effect of Prenatal Stress on Cooperation: Evidence from a Violent Conflict in Uganda. Working paper.
- Cecchi, F., J. Duchoslav, and E. Bulte (2016). Formal Insurance and the Dynamics of Social Capital: Experimental Evidence from Uganda. *Journal of African Economies* 25(3), 418–438.
- Cecchi, F., K. Leuvel, and M. J. Voors (2016). Conflict Exposure and Competitiveness: Experimental Evidence from the Football Field in Sierra

- Leone. *Economic Development and Cultural Change* 64(3), 405–435.
- Centers for Medicare and Medicaid Services (2011). CAHPS Hospital Survey (HCAPS): Quality Insurance Guidelines Version 6.0. <http://www.hcahpsonline.org> (Accessed: 2016-04-02).
- Cesarini, D., C. T. Dawes, J. H. Fowler, M. Johannesson, P. Lichtenstein, and B. Wallace (2008). Heritability of Cooperative Behavior in the Trust Game. *Proceedings of the National Academy of Sciences of the United States of America* 105(10), 3721–3726.
- Cesarini, D., C. T. Dawes, M. Johannesson, P. Lichtenstein, and B. Wallace (2009). Genetic Variation in Preferences for Giving and Risk Taking. *Quarterly Journal of Economics* 124(2), 809–842.
- Chamarbagwala, R. and H. E. Morán (2011). The Human Capital Consequences of Civil War: Evidence from Guatemala. *Journal of Development Economics* 94(1), 41–61.
- Chandrasekhar, A. G., C. Kinnan, and H. Larreguy (2016). Social Networks as Contract Enforcement: Evidence from a Lab Experiment in the Field. NBER Working Paper no. 20259.
- Choi, I. (2001). Unit Root Tests for Panel Data. *Journal of International Money and Finance* 20(2), 249–272.
- Coate, S. and M. Ravallion (1993). Reciprocity Without Commitment: Characterization and Performance of Informal Insurance Arrangements. *Journal of Development Economics* 40, 1–24.
- Coates, J. M., M. Gurnell, and A. Rustichini (2009). Second-to-Fourth Digit Ratio Predicts Success Among High-Frequency Financial Traders. *Proceedings of the National Academy of Sciences of the United States of America* 106(2), 623–628.
- Conybeare, D., E. Behar, A. Solomon, M. G. Newman, and T. D. Borkovec (2012). The PTSD Checklist-Civilian Version: Reliability, Validity, and Factor Structure in a Nonclinical Sample. *Journal of Clinical Psychology* 68(6), 699–713.
- Coumou, D. and S. Rahmstorf (2012). A Decade of Weather Extremes. *Nature*

- Climate Change* 2(7), 491–496.
- Craig, M., D. Le Sueur, and B. Snow (1999). A Climate-Based Distribution Model of Malaria Transmission in Sub-Saharan Africa. *Parasitology Today* 15(3), 105–111.
- Csathó, Á., A. Osváth, É. Bicsák, K. Karádi, J. Manning, and J. a. Kállai (2003). Sex Role Identity Related to the Ratio of Second to Fourth Digit Length in Women. *Biological Psychology* 62(2), 147–156.
- Currie, J. and T. Vogl (2013). Early-Life Health and Adult Circumstance in Developing Countries. *Annual Review of Economics* 5, 1–36.
- Daniels, L. (2011). Measuring Poverty Trends in Uganda with Non-monetary Indicators. Presented at the Fourth Global Conference on Agricultural and Rural Household Statistics in Rio de Janeiro, Brazil.
- Davies, A. R. and J. E. Ware (1988). Involving Consumers in Quality of Care Assessment. *Health Affairs* 7(1), 33–48.
- De Dreu, C. K. W., L. L. Greer, M. J. J. Handgraaf, S. Shalvi, G. A. V. Kleef, M. Baas, F. S. T. Velden, E. V. Dijk, and S. W. W. Feith (2010). The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict Among Humans. *Science* 328, 1408–1411.
- de Janvry, A., V. Dequiedt, and E. Sadoulet (2014). The Demand for Insurance against Common Shocks. *Journal of Development Economics* 106, 227–238.
- de Jong, J. (2002). *Trauma, War and Violence*. New York, NY: Springer.
- de Kloet, C. S., E. Vermetten, E. Geuze, E. G. W. M. Lentjes, C. J. Heijnen, G. K. Stalla, and H. G. M. Westenberg (2008). Elevated Plasma Corticotrophin-Releasing Hormone Levels in Veterans with Posttraumatic Stress Disorder. *Progress in Brain Research* 167(1), 287–291.
- de Onis, M., A. W. Onyango, E. Borghi, A. Siyam, C. Nishida, and J. Siekmann (2007). Development of a WHO Growth Reference for School-Aged Children and Adolescents. *Bulletin of the World Health Organization* 85(1), 660–7.
- De Weerd, J. and M. Fafchamps (2011). Social Identity and the Formation of

- Health Insurance Networks. *Journal of Development Studies* 47(8), 1152–1177.
- Dekker, M. and A. Wilms (2010). Health Insurance and Other Risk-Coping Strategies in Uganda: The Case of Microcare Insurance Ltd. *World Development* 38(3), 369–378.
- Delahanty, D. L., a. J. Raimonde, and E. Spoonster (2000). Initial Posttraumatic Urinary Cortisol Levels Predict Subsequent PTSD Symptoms in Motor Vehicle Accident Victims. *Biological Psychiatry* 48(9), 940–947.
- Dell, M., B. F. Jones, and B. A. Olken (2009). Temperature and Income: Reconciling new Cross-Sectional and Panel Estimates. *American Economic Review* 99(2), 198–204.
- Dercon, S. and P. Krishnan (2003). Risk Sharing and Public Transfers. *Economic Journal* 113(486), C86–C94.
- Dercon, S., R. Vargas Hill, D. Clarke, I. Outes-Leon, and A. Seyoum Taffesse (2014). Offering Rainfall Insurance to Informal Insurance Groups: Evidence from a field Experiment in Ethiopia. *Journal of Development Economics* 106, 132–143.
- Deschênes, O., M. Greenstone, and J. Guryan (2009). Climate Change and Birth Weight. *American Economic Review* 99(2), 211–217.
- Dictionary.com (2017). Environment. <http://www.dictionary.com/browse/environment?s=t> (Accessed: 2017-03-02).
- Dobie, D. J., D. R. Kivlahan, C. Maynard, K. R. Bush, M. Mcfall, A. J. Epler, and K. A. Bradley (2002). Screening for Post-Traumatic Stress Disorder in Female Veteran’s Affairs Patients: Validation of the PTSD Checklist. *General Hospital Psychiatry* 24(1), 367–374.
- Doepke, M. and F. Zilibotti (2017). Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission. *Econometrica* (forthcoming).
- Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2012). The Intergenerational Transmission of Risk and Trust Attitudes. *Review of Economic Studies* 79,

645–677.

Dolan, C. (2009). *Social Torture: The Case of Northern Uganda, 1986–2006*. New York, NY: Berghahn Books.

Doom, R. and K. Vlassenroot (1999). Kony’s Message: A New Koine? The Lord’s Resistance Army in Northern Uganda. *African Affairs* 98(390), 5–36.

Dörner, G., F. Götz, W. Rohde, A. Plagemann, R. Lindner, H. Peters, and Z. Ghanaati (2001). Genetic and Epigenetic Effects on Sexual Brain Organization Mediated by Sex Hormones. *Neuroendocrinology Letters* 22(6), 403–409.

Duchoslav, J. (2017). Prenatal Temperature Shocks Reduce Cooperation: Evidence from Public Goods Games in Uganda. Working paper.

Duchoslav, J. and F. Cecchi (2017). Evaluating the Impact of Performance-Based Financing in Healthcare Provision: A Case Study in Uganda. Working paper.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review* 101(5), 1739–1774.

Duflo, E. and M. Kremer (2005). Use of Randomization in the Evaluation of Development Effectiveness. In G. Pitman, O. Feinstein, and G. Ingram (Eds.), *Evaluating Development Effectiveness*. New Brunswick, NJ: Transaction Publishers.

Dunkel Schetter, C. (2011). Psychological Science on Pregnancy: Stress Processes, Biopsychosocial Models, and Emerging Research Issues. *Annual Review of Psychology* 62, 531–558.

Eichler, R. (2006). Can “Pay for Performance” Increase Utilization by the Poor and Improve the Quality of Health Services. Discussion paper for the first meeting of the Working Group on Performance-Based Incentives at the Center for Global Development.

Eichner, A. S. (1983). Why Economics Is Not Yet a Science. *Journal of Economic Issues* 17(2), 507–520.

- Eijkenaar, F., M. Emmert, M. Scheppach, and O. Schöfski (2013). Effects of Pay for Performance in Health Care: A Systematic Review of Systematic Reviews. *Health Policy* 110(2-3), 115–130.
- Eisenberg, N., R. A. Fabes, and T. L. Spinrad (2006). Prosocial Development. In *Handbook of Child Psychology*, pp. 646–719. Hoboken, NJ: Wiley.
- Eisenegger, C., M. Naef, R. Snozzi, M. Heinrichs, and E. Fehr (2010). Prejudice and Truth about the Effect of Testosterone on Human Bargaining Behaviour. *Nature* 463, 356–359.
- Eldridge, C. and N. Palmer (2009). Performance-Based Payment: Some Reflections on the Discourse, Evidence and Unanswered Questions. *Health Policy and Planning* 24(3), 160–6.
- Ensminger, J. (2004). Market integration and fairness: evidence from ultimatum, dictator, and public goods experiments in East Africa. In J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis (Eds.), *Foundations of Human Sociality*. Oxford: Oxford University Press.
- Fafchamps, M. and F. Gubert (2007). The Formation of Risk Sharing Networks. *Journal of Development Economics* 83, 326–350.
- Fearon, B. J. D., M. Humphreys, and J. M. Weinstein (2009). Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post-Conflict Liberia. *American Economic Review* 99(2), 287–291.
- Fehr, E. and S. Gächter (2002). Altruistic Punishment in Humans. *Nature* 415(6868), 137–140.
- Fehr, E. and H. Gintis (2007). Human Motivation and Social Cooperation: Experimental and Analytical Foundations. *Annual Review of Sociology* 33(1), 43–64.
- Fehr, E. and K. Hoff (2011). Introduction: Tastes, Castes and Culture: The Influence of Society on Preferences. *Economic Journal* 121(556), 396–412.
- Figlio, D., J. Guryan, K. Karbownik, and J. Roth (2014). The Effects of Poor

- Neonatal Health on Children's Cognitive Development. *American Economic Review* 104(12), 3921–3955.
- Finnström, S. (2008). *Living with Bad Surroundings: War, History, and Everyday Moments in Northern Uganda*. Durham, NC: Duke University Press.
- Flood, M. M. (1958). Some Experimental Games. *Management Science* 5(1), 5–26.
- Flouris, A. D., Y. Spiropoulos, G. J. Sakellariou, and Y. Koutedakis (2009). Effect of Seasonal Programming on Fetal Development and Longevity: Links with Environmental Temperature. *American Journal of Human Biology* 21(2), 214–216.
- Fochmann, M. and J. Weimann (2013). The Effects of Tax Salience and Tax Experience on Individual Work efforts in a Framed Field Experiment. *FinanzArchiv* 69(4), 511–542.
- Fowden, A. L., D. A. Giussani, and A. J. Forhead (2006). Intrauterine Programming of Physiological Systems: Causes and Consequences. *Physiology* 21, 29–37.
- Freedly, J. R., M. M. Steenkamp, K. M. Magruder, D. E. Yeager, J. S. Zoller, W. J. Hueston, and P. J. Carek (2010). Post-Traumatic Stress Disorder Screening Test Performance in Civilian Primary Care. *Family Practice* 27(6), 615–624.
- Fretheim, A., S. Witter, A. K. Lindahl, and I. T. Olsen (2012). Performance-Based Financing in Low- and Middle-Income Countries: Still More Questions than Answers. *Bulletin of the World Health Organization* 90(8), 559–559A.
- Frey, B. S. and R. Jegen (2001). Motivation Crowding Theory. *Journal of Economic Surveys* 15(5), 589–611.
- Friedman, M. (1962). *Price Theory*. New Brunswick, NJ: Transaction Publishers.
- Galvan, D. (2004). *The State Must be Our Master of Fire: How Peasants Craft Culturally Sustainable Development in Senegal*. Berkley, CA: University of California Press.
- Garbarino, E., R. Slonim, and J. Sydnor (2010). Digit Ratios (2D:4D) as Predictors

- of Risky Decision Making for Both Sexes. *Journal of Risk and Uncertainty* 42(1), 1–26.
- Gates, S., H. Hegre, H. M. Nygård, and H. Strand (2012). Development Consequences of Armed Conflict. *World Development* 40(9), 1713–1722.
- Gerber, A. S. and D. P. Green (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W. W. Norton & Company.
- Ghobarah, H. A., P. Hurt, and B. Russett (2003). Civil Wars Kill and Maim People - Long after the Shooting Stops. *American Political Science Review* 97(2), 189–202.
- Gilligan, M. J., B. J. Pasquale, and C. Samii (2014). Civil War and Social Cohesion: Lab-in-the-Field Evidence from Nepal. *American Journal of Political Science* 58(3), 604–619.
- Gintis, H., S. Bowles, R. Boyd, and E. Fehr (2003). Explaining Altruistic Behavior in Humans. *Evolution and Human Behavior* 24(3), 153–172.
- Gneezy, U., S. Meier, and P. Rey-Biel (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives* 25(4), 191–210.
- Goldschmidt, W. R. (2006). *The Bridge to Humanity: How Affect Hunger Trumps the Selfish Gene*. Oxford: Oxford University Press.
- Governemnt of Uganda - Ministry of Health (2010). Health Sector Strategic Plan III: 2010/11-2014/15.
- Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology* 91(3), 481–510.
- Greif, A. (1993). Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition. *American Economic Review* 83(3), 525–548.
- Grimm, M., R. Hartwig, and J. Lay (2017). Does Forced Solidarity Hamper Investment in Small and Micro Enterprises? *Journal of Comparative Economics* (forthcoming).
- Grzymala-Busse, A. (2010). The Best Laid Plans: The Impact of Informal

- Rules on Formal Institutions in Transitional Regimes. *Studies in Comparative International Development* 45(3), 311–333.
- Hack, M., H. G. Taylor, D. Drotar, M. Schluchter, L. Cartar, L. Andreias, D. Wilson-Costello, and N. Klein (2005). Chronic Conditions, Functional Limitations, and Special Health Care Needs of School-Aged Children Born with Extremely Low-Birth-Weight in the 1990s. *JAMA* 294(3), 318–325.
- Hansen, P. J. (2009). Effects of Heat Stress on Mammalian Reproduction. *Philosophical Transactions: Biological Sciences* 364(1534), 3341–50.
- Harada, M. (2013). Generalized Sensitivity Analysis and Its Application to Quasi-Experiments. Working paper.
- Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister (2014). Updated High Resolution Grids of Monthly Climatic Observations—the CRU TS3. 10 Dataset. *International Journal of Climatology* 34(3), 623–642.
- Harrison, G. W. and J. A. List (2004). Field Experiments. *Journal of Economic Literature* 42(4), 1009–1055.
- Hauert, C., A. Traulsen, H. Brandt, M. A. Nowak, and K. Sigmund (2007). Via Freedom to Coercion: The Emergence of Costly Punishment. *Science* 316(2007), 1905–1907.
- Hawking, S. (2016). This Is the Most Dangerous Time for Our Planet. *Guardian* (1 December).
- Hecht, R., A. Batson, and L. Brenzel (2004). Making Health Care Accountable. *Finance & Development* (March), 16–19.
- Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica* 47(1), 153–161.
- Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath (2001). In Search of Homo Economicus: Behavioral Experiments in 15 Small Scale Societies. *American Economic Review* 91(2), 73–78.
- Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe,

- D. Tracer, and J. Ziker (2006). Costly Punishment across Human Societies. *Science* 312(5781), 1767–1770.
- Hille, E. T. M., N. Weisglas-Kuperus, J. B. van Goudoever, G. W. Jacobusse, M. H. Ens-Dokkum, L. de Groot, J. M. Wit, W. B. Geven, J. H. Kok, M. J. K. de Kleine, L. a. a. Kollée, a. L. M. Mulder, H. L. M. van Straaten, L. S. de Vries, M. M. van Weissenbruch, and S. P. Verloove-Vanhorick (2007). Functional Outcomes and Participation in Young Adulthood for Very Preterm and Very Low Birth Weight Infants: The Dutch Project on Preterm and Small for Gestational Age Infants at 19 Years of Age. *Pediatrics* 120(3), e587–e595.
- Hobel, C. J., A. Goldstein, and E. S. Barrett (2008). Psychosocial Stress and Pregnancy Outcome. *Clinical Obstetrics and Gynecology* 51(2), 333–348.
- Hollingsworth, B. (2008). The Measurement of Efficiency and Productivity of Health Care Delivery. *Health Economics* 17(August), 1107–1128.
- Honda, A. (2013). 10 Best Resources on ... Pay for Performance in Low- and Middle-Income Countries. *Health Policy and Planning* 28(5), 454–7.
- Horrace, W. C. and R. L. Oaxaca (2006). Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model. *Economics Letters* 90(3), 321–327.
- Hsiang, S. M., M. Burke, and E. Miguel (2013). Quantifying the Influence of Climate on Human Conflict. *Science* 341(6151).
- Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge: Cambridge University Press.
- Hunt, J. (2010). Bribery in Health Care in Uganda. *Journal of Health Economics* 29(5), 699–707.
- Im, K. S., M. H. Pesaran, and Y. Shin (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* 115(1), 53–74.
- Imbens, G. W. (2003). Sensitivity to Exogeneity Assumptions in Program Evaluation. *American Economic Review* 93(2), 126–132.
- International Monetary Fund (2016). International Financial Statistics (Edition:

- January 2016). <http://dx.doi.org/10.5257/imf/ifs/2016-01> (Accessed: 2016-02-15).
- Ireland, M., E. Paul, and B. Dujardin (2011). Can Performance-Based Financing be Used to Reform Health Systems in Developing Countries? *Bulletin of the World Health Organization* 89(9), 695–8.
- Jackson, P. (2002). The March of the Lord's Resistance Army: Greed or Grievance in Northern Uganda? *Small Wars and Insurgencies* 13(3), 29–52.
- Jirtle, R. L. and M. K. Skinner (2007). Environmental Epigenomics and Disease Susceptibility. *Nature Reviews Genetics* 8, 253–262.
- Johansson, P., M. Oléni, and B. Fridlund (2002). Patient Satisfaction with Nursing Care in the Context of Health Care: a Literature Study. *Scandinavian Journal of Caring Sciences* 16(4), 337–44.
- Kaplan, R. and D. Saccuzzo (2012). *Psychological Testing: Principles, Applications, and Issues*. Boston, MA: Cengage Learning.
- Keen, S. M., C. J. Kutter, B. L. Niles, and K. E. Krinsley (2008). Psychometric Properties of PTSD Checklist in Sample of Male Veterans. *Journal of Rehabilitation Research and Development* 45(3), 465–474.
- Keverne, E. B. and J. P. Curley (2008). Epigenetics, Brain Evolution and Behaviour. *Frontiers in Neuroendocrinology* 29(3), 398–412.
- Kimball, M. S. (1988). Farmers' Cooperatives as Behavior toward Risk. *American Economic Review* 78(1), 224–232.
- Klohn, F. and C. Strupat (2013). Crowding out of Solidarity? – Public Health insurance versus informal Transfer Networks in Ghana. RUHR Economic Paper No. 432.
- Koppensteiner, M. F. and M. Manacorda (2016). The Effect of Violence on Birth Outcomes: Evidence from Homicides in Rural Brazil. *Journal of Development Economics* 119, 16–33.
- Kosfeld, M., M. Heinrichs, P. J. Zak, U. Fischbacher, and E. Fehr (2005). Oxytocin Increases Trust in Humans. *Nature* 435, 673–676.

- Kosfeld, M., A. Okada, and A. Riedl (2009). Institution Formation in Public Goods Games Formation in Public Goods Games. *American Economic Review* 99(4), 1335–1355.
- Lacetera, N. and M. Macis (2010a). Do All Material Incentives for Pro-Social Activities Backfire? The Response to Cash and Non-Cash Incentives for Blood Donations. *Journal of Economic Psychology* 31(4), 738–748.
- Lacetera, N. and M. Macis (2010b). Social Image Concerns and Prosocial Behavior: Field Evidence from a Nonlinear Incentive Scheme. *Journal of Economic Behavior and Organization* 76(2), 225–237.
- Lajinian, S., S. Hudson, L. Applewhite, J. Feldman, and H. L. Minkoff (1997). An Association between the Heat-Humidity Index and Preterm Labor and Delivery: A Preliminary Analysis. *American Journal of Public Health* 87(7), 1205–1207.
- Landmann, A., B. Vollan, and M. Frölich (2012). Insurance versus Savings for the Poor: Why One Should Offer Either Both or None. IZA Discussion Paper No. 6298.
- Lauderdale, D. S. (2006). Birth Outcomes for Arabic-Named Women in California Before and After September 11. *Demography* 43(1), 185–201.
- Lawlor, D. A., D. A. Leon, and G. Davey Smith (2005). The Association of Ambient Outdoor Temperature throughout Pregnancy and Offspring Birthweight: Findings from the Aberdeen Children of the 1950s Cohort. *BJOG: An International Journal of Obstetrics and Gynaecology* 112(5), 647–57.
- Leonard, K. L. (2008). Is Patient Satisfaction Sensitive to Changes in the Quality of Care? An Exploitation of the Hawthorne Effect. *Journal of Health Economics* 27(2), 444–59.
- Levitt, S. and J. List (2009). Field Experiments in Economics: The Past, the Present, and the Future. *European Economic Review* 53(1), 1–18.
- Levitt, S. D. and J. A. List (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World? *Journal of Economic Perspectives* 21(2), 153–174.

- Ligon, E., J. P. Thomas, and T. Worrall (2002). Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies. *Review of Economic Studies* 69(1), 209–244.
- Lilley, T., T. Laaksonen, O. Huitu, and S. Helle (2010). Maternal Corticosterone but not Testosterone Level is Associated with the Ratio of Second-to-Fourth Digit Length (2D:4D) in Field Vole Offspring (*Microtus agrestis*). *Physiology & Behavior* 99(4), 433–437.
- Lin, W., Y. Liu, and J. Meng (2014). The Crowding-Out Effect of Formal Insurance on Informal Risk Sharing: An Experimental Study. *Games and Economic Behavior* 86, 184–211.
- List, J. A. (2006). Field Experiments : A Bridge between Lab and Naturally Occurring Data. *Advances in Economic Analysis and Policy* 6(2), Article 8.
- Lopes Cardozo, B., O. O. Bilukha, C. a. G. Crawford, I. Shaikh, M. I. Wolfe, M. L. Gerber, and M. Anderson (2004). Mental Health, Social Functioning, and Disability in Postwar Afghanistan. *Journal of the American Medical Association* 292(5), 575–584.
- Lu, M. (1999). Separating the True Effect from Gaming in Incentive-Based Contracts in Health Care. *Journal of Economics & Management Strategy* 8(3), 383–431.
- Lutchmaya, S., S. Baron-Cohen, P. Raggatt, R. C. Knickmeyer, and J. T. Manning (2004). 2nd to 4th Digit Ratios, Fetal Testosterone and Estradiol. *Early Human Development* 77(1-2), 23–28.
- Macdonald, D., S. Creel, and M. Mills (2004). Society: Canid Society. In D. Macdonald and C. Sillero-Zubiri (Eds.), *Biology and Conservation of Wild Carnivores*, pp. 85–106. Oxford: Oxford University Press.
- Macksoud, M. (1992). Assessing War Trauma in Children: A Case Study of Lebanese Children. *Journal of Refugee Studies* 5(1), 1–15.
- Mancuso, R. A., C. Dunkel Schetter, C. M. Rini, S. C. Roesch, and C. J. Hobel (2004). Maternal Prenatal Anxiety and Corticotropin-Releasing Hormone

- Associated with Timing of Delivery. *Psychosomatic Medicine* 66(5), 762–769.
- Manning, J. T., P. E. Bundred, D. J. Newton, and B. F. Flanagan (2003). The Second to Fourth Digit Ratio and Variation in the Androgen Receptor Gene. *Evolution and Human Behavior* 24(6), 399–405.
- Marley, K. A., D. A. Collier, and S. Meyer Goldstein (2004). The Role of Clinical and Process Quality in Achieving Patient Satisfaction in Hospitals. *Decision Sciences* 35(3), 349–369.
- Maynard, A. (2012). The Powers and Pitfalls of Payment for Performance. *Health Economics* 12, 3–12.
- McDonald, C., C. Schiller, and K. Ueda (1999). Income Distribution, Informal Safety Nets, and Social Expenditures in Uganda. IMF Working paper No. 99/163.
- Meessen, B., J.-P. I. Kashala, and L. Musango (2007). Output-Based Payment to Boost Staff Productivity in Public Health Centres: Contracting in Kabutare District, Rwanda. *Bulletin of the World Health Organization* 85(2), 108–115.
- Meessen, B., L. Musango, J.-P. I. Kashala, and J. Lemlin (2006). Reviewing Institutions of Rural Health Centres: the Performance Initiative in Butare, Rwanda. *Tropical Medicine & International Health* 11(8), 1303–1317.
- Mehta, P. H. and R. A. Josephs (2011). Social Endocrinology: Hormones and Social Motivation. In D. Dunning (Ed.), *Social Motivation*, pp. 171–189. New York, NY: Psychology Press.
- Miguel, E., S. M. Saiegh, and S. Satyanath (2011). Civil War Exposure and Violence. *Economics & Politics* 23(1), 59–73.
- Miller, K. E., P. Omidian, A. Rasmussen, A. Yaqubi, and H. Daudzai (2008). Daily Stressors, War Experiences, and Mental Health in Afghanistan. *Transcultural Psychiatry* 45(4), 611–38.
- Montesquieu (1748). *The Spirit of Laws*. La Brède: Montequieu.
- Moore, C., J. Barresi, and C. Thompson (1998). The Cognitive Basis of Future-Oriented Prosocial Behavior. *Social Development* 7(2), 198–218.

- Morgan, L. (2010). Some Days Are Better than Others: Lessons Learned from Uganda's First Results-Based Financing Pilot. World Bank Report No. 53985.
- Morten, M. (2016). Temporary Migration and Endogenous Risk Sharing in Village India. NBER Working Paper No. 22159.
- Newton, I. (1675). Letter from Sir Isaac Newton to Robert Hooke. http://digitallibrary.hsp.org/index.php/Detail/Object/Show/object_id/9285 (Accessed: 2017-03-05).
- Nijman, T. and M. Verbeek (1992). Nonresponse in Panel Data: The Impact on Estimates of a Life Cycle Consumption Function. *Journal of Applied Econometrics* 7(3), 243–257.
- North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- Nowak, M. A. (2006). Five Rules for the Evolution of Cooperation. *Science* 314(5805), 1560–1563.
- Nowak, M. A. and R. Highfield (2011). *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*. New York, NY: Simon & Schuster.
- O'Loughlin, J., F. D. W. Witmer, A. M. Linke, A. Laing, A. Gettelman, and J. Dudhia (2012). Climate Variability and Conflict Risk in East Africa. *Proceedings of the National Academy of Sciences of the United States of America* 109(45), 18344–9.
- Oreskes, N. (2005). The Scientific Consensus on Climate Change. *Science* 306(5702), 1686.
- Ostrom, E., J. Walker, and R. Gardner (1992). Covenants With and Without a Sword: Self-Governance is Possible. *American Political Science Review* 86(2), 404–417.
- Oxman, A. D. and A. Fretheim (2008). An Overview of Research on the Effects of Results-Based Financing. Norwegian Knowledge Centre for the Health Services Report No. 16-2008.
- Papageorgiou, V., A. Frangou-Garunovic, R. Iordanidou, W. Yule, P. Smith, and

- P. Vostanis (2000). War Trauma and Psychopathology in Bosnian Refugee Children. *European Child & Adolescent Psychiatry* 9(2), 84–90.
- Peabody, J. W., R. Shimkhada, S. Quimbo, O. Solon, X. Javier, and C. McCulloch (2013). The Impact of Performance Incentives on Child Health Outcomes: Results from a Cluster randomized Controlled Trial in the Philippines. *Health Policy and Planning*.
- Peabody, J. W., F. Tozija, J. a. Muñoz, R. J. Nordyke, and J. Luck (2004). Using Vignettes to Compare the Quality of Clinical Care Variation in Economically Divergent Countries. *Health Services Research* 39(6), 1951–70.
- Pejovich, S. (1999). The Effects of the Interaction of Formal and Informal Institutions on Social Stability and Economic Development. *Journal of Markets & Morality* 2(2), 164–181.
- Perrot, J., E. de Roodenbeke, L. Musango, and G. Fritsche (2010). Performance Incentives for Health Care Providers. WHO/HSF/HSS Discussion Paper No. 1-2010.
- Pfeiffer, A. and T. Elbert (2011). PTSD, Depression and Anxiety among Former Abductees in Northern Uganda. *Conflict and Health* 5(14), 1–7.
- Pham, P., P. Vinck, and E. Stover (2007). Abducted: The Lord 's Resistance Army in Northern Uganda. Report for the Berkley-Tulane Initiative on Vulnerable Populations.
- Polanyi, K. (1944). *The Great Transformation*. New York, NY: Farrar & Rinehart.
- Quintana-Domeque, C. and P. Ródenas-Serrano (2017). The Hidden Costs of Terrorism: The Effects on Human Capital at Birth. Working paper.
- Rand, D. G., A. Dreber, T. Ellingsen, D. Fudenberg, and M. A. Nowak (2009). Positive Interactions Promote Public Cooperation. *Science* 325(September), 1272–1276.
- Reichhuber, A., E. Camacho, and T. Requate (2009). A Framed Field Experiment on Collective Enforcement Mechanisms with Ethiopian Farmers. *Environment and Development Economics* 14(5), 641–663.

- Richards, P. (1996). *Fighting for the Rain Forest: War, Youth and Resources in Sierra Leone*. Oxford: James Currey.
- Roberts, B., K. F. Ocaña, J. Browne, T. Oyok, and E. Sondorp (2008). Factors Associated with Post-Traumatic Stress Disorder and Depression amongst Internally Displaced Persons in Northern Uganda. *BMC Psychiatry* 8(38), 1–9.
- Rodrik, D., A. Subramanian, and F. Trebbi (2004). Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. *Journal of Economic Growth* 9, 131–165.
- Rohner, D., M. Thoenig, and F. Zilibotti (2013). Seeds of Distrust: Conflict in Uganda. *Journal of Economic Growth* 18(3), 217–252.
- Roland, G. (2004). Understanding Institutional Change: Fast-Moving and Slow-Moving Institutions. *Studies in Comparative International Development* 38(4), 109–131.
- Roodman, D. (2009). How to Do xtabond2: An Introduction to Difference and System GMM in Stata. *Stata Journal* 9(1), 86–136.
- Ruggiero, K. J., A. A. Rheingold, H. S. Resnick, D. G. Kilpatrick, and S. Galea (2006). Comparison of Two Widely Used PTSD-Screening Instruments: Implications for Public Mental Health Planning. *Journal of Traumatic Stress* 19(5), 699–707.
- Rusa, L., M. Schneidman, G. Fritzsche, and L. Musango (2009). Rwanda: Performance-Based Financing in the Public Sector. In R. Eichler and R. Levine (Eds.), *Performance Incentives for Global Health: Potential and Pitfalls*, pp. 190–214. Washington, D.C.: Center for Global Development.
- Rushton, J. P. (1991). Is Altruism Innate? *Psychological Inquiry* 2(2), 141–143.
- Rutter, M. L. (1997). Nature-Nurture Integration: The Example of Antisocial Behavior. *American Psychologist* 52(4), 390–398.
- Ryan, A. (2009). Hospital-Based Pay-for-Performance in the United States. *Health Economics* 18, 1109–1113.
- Sahn, D. and D. Stifel (2003). Exploring Alternative Measures of Welfare in the

- Absence of Expenditure Data. *Review of Income and Wealth* 49(4), 463–489.
- Sapienza, P., L. Zingales, and D. Maestripieri (2009). Gender Differences in Financial Risk Aversion and Career Choices are Affected by Testosterone. *Proceedings of the National Academy of Sciences of the United States of America* 106(36), 15268–15273.
- Scheffler, R. M. (2010). Pay For Performance (P4P) Programs in Health Services: What is the Evidence? WHO World Health Report (2010) Background Paper No. 31.
- Scott, A. and W. Coote (2010). Do Regional Primary-Care Organisations Influence Primary-Care Performance? A Dynamic Panel Estimation. *Health Economics* 19, 716–729.
- Scott, J. C. (1985). *Weapons of the Weak: Everyday Forms of Peasant Resistance*. New Haven, CT: Yale University Press.
- Sekabaraga, C., F. Diop, and A. Soucat (2011). Can Innovative Health Financing Policies Increase Access to MDG-Related Services? Evidence from Rwanda. *Health Policy and Planning* 26 Suppl 2, ii52–ii62.
- Selten, R. and A. Ockenfels (1998). An Experimental Solidarity Game. *Journal of Economic Behavior & Organization* 34, 517–539.
- Shorter, J. R. and O. Rueppell (2012). A Review on Self-Destructive Defense Behaviors in Social Insects. *Insectes Sociaux* 59(1), 1–10.
- SHU (2013). Community Health Financing in SHU. Save for Health Uganda Internal Report.
- Simar, L. and P. W. Wilson (2007). Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes. *Journal of Econometrics* 136(1), 31–64.
- Skinner, B. F. (1938). *The Behavior of Organisms*. New York, NY: Appleton-Century.
- Skinner, B. F. (1953). *Science and Human Behavior*. New York, NY: Simon & Schuster.

- Skutch, A. F. (1935). Helpers at the Nest. *Auk* 52(3), 257–273.
- Smith, A. (1761). *The Theory of Moral Sentiments*. London: A. Millar.
- Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan and T. Cadell.
- Smith, R. and R. C. Nicholson (2007). Corticotrophin Releasing Hormone and the Timing of Birth. *Frontiers in Bioscience* 12, 912–918.
- Smith, V. L. (1976). Experimental Economics: Induced Value Theory. *American Economic Review* 66(2), 274–279.
- Smith, V. L. (1982). Microeconomic Systems as an Experimental Science. *American Economic Review* 72(5), 923–955.
- Smith, V. L. (2010). Theory and Experiment: What Are the Questions? *Journal of Economic Behavior and Organization* 73(1), 3–15.
- Soeters, R., C. Habineza, and P. Bob (2006). Performance-Based Financing and Changing the District Health System : Experience from Rwanda. *Bulletin of the World Health Organization* 84(11), 884–889.
- Soeters, R., P. B. Peerenboom, P. Mushagalusa, and C. Kimanuka (2011). Performance-Based Financing Experiment Improved Health Care in the Democratic Republic of Congo. *Health Affairs* 30(8), 1518–1527.
- Song, Y., D. Zhou, and X. Wang (2008). Increased Serum Cortisol and Growth Hormone Levels in Earthquake Survivors with PTSD or Subclinical PTSD. *Psychoneuroendocrinology* 33(8), 1155–1159.
- Spolaore, E. and R. Wacziarg (2013). How Deep Are the Roots of Economic Development? *Journal of Economic Literature* 51(2), 325–369.
- Steudte, S., I.-T. Kolassa, T. Stalder, A. Pfeiffer, C. Kirschbaum, and T. Elbert (2011). Increased Cortisol Concentrations in Hair of Severely Traumatized Ugandan Individuals with PTSD. *Psychoneuroendocrinology* 36(8), 1193–1200.
- Stigler, G. J. and G. S. Becker (1977). De Gustibus Non Est Disputandum. *American Economic Review* 67(2), 76–90.

- Stott, D. H. (1973). Follow-up Study from Birth of the Effects of Prenatal Stresses. *Developmental Medicine & Child Neurology* 15(6), 770–787.
- Sutter, M., S. Haigner, and M. G. Kocher (2010). Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations. *Review of Economic Studies* 77(4), 1540–1566.
- Taylor, L., J. Seeley, and E. Kajura (1996). Informal Care for Illness in Rural Southwest Uganda: The Central Role that Women Play. *Health Transition Review* 6, 49–56.
- Titmuss, R. M. (1970). *The Gift Relationship*. London: Allen and Unwin.
- Townsend, R. M. (1994). Risk and Insurance in Village India. *Econometrica* 62(3), 539–591.
- Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *Quarterly Review of Biology* 46(1), 35–57.
- Udry, C. (1994). Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria. *Review of Economic Studies* 61(3), 495–526.
- UN DESA (2012). Demographic Profiles. In *World Population Prospects: The 2012 Revision*, Volume 2. New York, NY: United Nations Department of Economic and Social Affairs.
- van Os, J. and J.-P. Selten (1998). Prenatal Exposure to Maternal Stress and Subsequent Schizophrenia: The May 1940 invasion of The Netherlands. *British Journal of Psychiatry* 172(4), 324–326.
- Vinci, A. (2005). The Strategic Use of Fear by the Lord's Resistance Army. *Small Wars and Insurgencies* 16(3), 360–381.
- vom Saal, F. S., D. M. Quadagno, M. D. Even, D. H. Keisler, L. W. Keisler, and S. Khan (1990). Paradoxical Effects of Maternal Stress on Fetal Steroids and Postnatal Reproductive Traits in Female Mice from Different Intrauterine Positions. *Biology of Reproduction* 43(5), 751–761.
- Voors, M., E. Bulte, A. Kontoleon, J. a. List, and T. Turley (2011). Using Artefactual Field Experiments to Learn about the Incentives for Sustainable

- Forest Use in Developing Economies. *American Economic Review* 101(3), 329–333.
- Voors, M. J., E. E. Nillesen, P. Verwimp, E. H. Bulte, R. Lensink, and D. P. Van Soest (2012). Violent Conflict and Behavior: A Field Experiment in Burundi. *American Economic Review* 102(2), 941–964.
- Voracek, M., J. T. Manning, and S. G. Dressler (2007). Short Report Repeatability and Interobserver Error of Digit Ratio (2D:4D) Measurements Made by Experts. *American Journal of Human Biology* 19(1), 142–146.
- Walder, D. J., D. P. Laplante, A. Sousa-Pires, F. Veru, A. Brunet, and S. King (2014). Prenatal Maternal Stress Predicts Autism Traits in 61/2 Year-Old Children: Project Ice Storm. *Psychiatry Research* 219(2), 353–360.
- Walker, S. P., P. S. Gaskin, C. A. Powell, and F. I. Bennett (2002). The Effects of Birth Weight and Postnatal Linear Growth Retardation on Body Mass Index, Fatness and Fat Distribution in Mid and Late Childhood. *Public Health Nutrition* 5(3), 391–396.
- Ward, I. L. and J. Weisz (1980). Maternal Stress Alters Plasma Testosterone in Fetal Males. *Science* 207(4428), 328–329.
- Ware, J. E. J., M. Kosinski, and S. D. Keller (1996). A 12-Item Short-Form Health Survey of Scales and Preliminary Construction Tests of Reliability and Validity. *Medical Care* 34(3), 220–233.
- Weathers, F. W., B. T. Litz, D. S. Herman, J. A. Huska, and T. M. Keane (1993). The PTSD Checklist (PCL): Reliability, Validity, and Diagnostic Utility. Presented at the Annual Meeting of International Society for Traumatic Stress Studies.
- Weil, D. N. (2007). Accounting for the Effect of Health on Economic Growth. *Quarterly Journal of Economics* (August), 1265–1306.
- Weinstock, M. (2008). The Long-Term Behavioural Consequences of Prenatal Stress. *Neuroscience and Biobehavioral Reviews* 32(6), 1073–86.
- Wells, J. C. K. and T. J. Cole (2002). Birth Weight and Environmental

- Heat Load: A Between-Population Analysis. *American Journal of Physical Anthropology* 119(3), 276–282.
- Wesson, R. (1991). *Beyond Natural Selection*. Cambridge, MA: The MIT Press.
- Wharam, J. F., M. K. Paasche-Orlow, N. J. Farber, C. Sinsky, L. Rucker, K. J. Rask, M. K. Figaro, C. Braddock, M. J. Barry, and D. P. Sulmasy (2009). High Quality Care and Ethical Pay-for-Performance: a Society of General Internal Medicine Policy Analysis. *Journal of General Internal Medicine* 24(7), 854–9.
- Wilde, J., B. H. Apouey, and T. Jung (2017). The Effect of Ambient Temperature Shocks During Conception and Early Pregnancy on Later Life Outcomes. *European Economic Review* (forthcoming).
- Wilde, L. L. (1981). On the Use of Laboratory Experiments in Economics. In J. C. Pitt (Ed.), *Philosophy in Economics*, pp. 137–148. Dordrecht: Springer Netherlands.
- Williamson, C. R. (2009). Informal Institutions Rule: Institutional Arrangements and Economic Performance. *Public Choice* 139, 371–387.
- Willmott, C. J. and K. Matsuura (2015). Terrestrial Air Temperature: 1900–2014 Gridded Monthly Time Series. http://climate.geog.udel.edu/~climate/html_pages/Global2014/README.GlobalTsT2014.html (Accessed: 2016-02-10).
- Wolszczak-Derlacz, J. and A. Parteka (2011). Efficiency of European Public Higher Education Institutions: A Two-Stage Multicountry Approach. *Scientometrics* 89(3), 887–917.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.
- World Bank (2011). *World Development Report 2011: Conflict, Security, and Development*. Washington, D.C: World Bank.
- World Bank (2016). World Development Indicators. <http://data.worldbank.org/products/wdi> (Accessed: 2016-02-06).
- Wynia, M. K. (2009). The Risks of Rewards in Health Care: How Pay-For-Performance Could Threaten, or Bolster, Medical Professionalism.

- Journal of General Internal Medicine* 24(7), 884–7.
- Yackerson, N., B. Piura, and E. Sheiner (2008). The Influence of Meteorological Factors on the Emergence of Preterm Delivery and Preterm Premature Rupture of Membrane. *Journal of Perinatology* 28(10), 707–711.
- Yajnik, C. S. (2002). The Lifecycle Effects of Nutrition and Body Size on Adult Adiposity, Diabetes and Cardiovascular Disease. *Obesity Reviews* 3, 217–224.
- Yamagishi, T. (1986). The Provision of a Sanctioning System as a Public Good. *Journal of Personality and Social Psychology* 51(1), 110–116.
- Zaff, J. F., K. A. Moore, A. Romano Papillo, and S. Williams (2003). Implications of Extracurricular Activity Participation During Adolescence on Positive Outcomes. *Journal of Adolescent Research* 18, 599–630.
- Zhang, S., W. Gao, and B. Fan (2015). Cognitive Ability and Cooperation: Evidence from the Public Goods Experiments. *Annals of Economics and Finance* 16(1), 43–68.
- Zheng, Z. and M. J. Cohn (2011). Developmental Basis of Sexually Dimorphic Digit Ratios. *Proceedings of the National Academy of Sciences of the United States of America* 108(39), 16289–16294.

Summary

All economic activity requires some degree of cooperation, and the process of economic development involves many social dilemmas. It is therefore crucial to understand how the preferences which guide our behavior vis-à-vis these situations are shaped. The ability and willingness to work for the benefit of the group rather than just one's own has evolved over many generations, and is – to some extent – innate to any healthy human being. At the same time, individual prosocial preferences are – also to a certain extent – endogenous to the physical and social environment within which we operate. This thesis identifies several ways in which environmental changes affect intrinsic prosocial preferences, and outlines a possible direction for fixing any such negative effects.

In Chapter 1, I introduce the topic of prosocial preferences. I briefly describe how prosociality has been viewed over the course of scientific history, and summarize the current state of knowledge about the formation of social preferences. I further outline how extrinsic incentives can influence prosocial behavior without affecting the preferences which underpin it. Finally, the chapter contains an overview of the methodologies used throughout this thesis.

In Chapter 2, I focus on an early formative factor of prosocial preferences—their fetal origins. I study how temperature shocks faced by pregnant women affect their children's later-life prosocial preferences. I find that exposure to higher than usual ambient temperatures during gestation reduces a child's probability of contribution to the public good, with the negative effect lasting into adulthood.

Chapter 3 continues in the same vein as Chapter 2, looking at the fetal origins of

prosocial preferences. In this chapter, I investigate how prenatal stress induced by random violence affects the preferences for cooperation among children born during an armed conflict. To do so, I exploit variations in the ratio of the lengths of the index and ring fingers—a marker of in utero hormone exposure negatively associated with high maternal distress during early fetal development. I show that prenatal stress reduces the probability that children contribute to the public good.

In Chapter 4, I move away from the physical aspects of human environment, focusing instead on the social ones. I study the effects of a sudden introduction of a formal institution on individual cooperative behavior within informal arrangements. In particular, I look at how an NGO intervention which helped create a mutual health insurance affected cooperative behavior in a public goods game. I find that the introduction of formal insurance reduces contributions to the public good. This reduction in cooperation levels is, however, not due to the adopters of the formal insurance who may now have less need for informal reciprocal networks, and who therefore (partially) withdraw from them. It is instead the non-adopters who become less cooperative towards the adopters.

To outline a possible direction for remedying the negative environmental effects on prosocial behavior described in the previous three chapters, I illustrate one of the ways in which prosocial behavior can be incentivized with a relatively simple and easily implementable policy. In Chapter 5, I evaluate the impact of introducing performance-based financial incentives on staff effort and, consequently, on allocative efficiency and output in healthcare provision. I show that in the case under investigation, financial incentives conditioned on output and worth roughly 5% of total expenditures increased staff effort to the extent that output rose by over 25%, without any detectable drop in the quality of the provided services. This not only shows the potential of incentive-compatible financing to improve the performance of underfunded healthcare systems in developing countries, but also that extrinsic motivation can be used to foster behavior which benefits the society rather than just the individual.

Finally, I combine the main findings from the core chapters of the thesis in Chapter

6. I discuss their policy implications, and point out the some of the outstanding questions, outlining the directions for future research.

Acknowledgements

This book is part of a personal journey which started many years ago, and which, surely, is nowhere near over. It began with Eyad and Ravi, who showed me the beauty of economics. It is thanks to them that this book is not about chemistry. Later on, David, Martha, and Charlotte helped me find the focal point of my work by introducing me to the field of development. Once in Wageningen, Marrit saw some potential in me, and Lonneke and Maarten gave me the opportunity to prove it. Erwin then gave me not only the independence of running my own research, but also the freedom to do it as I saw fit, trusting that I will deliver in the end. The fact that I did is largely due to his moral and financial support, as well as his pragmatic approach to problems and the willingness to solve them. James wonderfully complemented Erwin's flexibility by helping me to focus on the details when it was most needed. Finally, throughout the years in Wageningen, Francesco was everything from an office mate and a drinking buddy to a co-author and a co-promoter, and – above all – a good friend. The journey would have taken me in a very different direction had it not been for you all. Thank you for your guidance.

Many other people helped me along the way. I am humbled by the generosity with which hundreds of men, women, and children shared their time and opinions for the sake of my research, the purpose of which – despite my best efforts – must have often seemed silly. I am equally humbled by the enthusiasm and dedication of my research assistants, who worked long and hard to transform the respondents' feelings and life stories into hard numbers. My thanks also go to Francesco, Elena, Vuk, Aisha, Pippo, Esther, Vojta, Edward, and John, who

not only aptly supervised the data collection teams, but who were also excellent travel companions. Wherever we did travel, Muna always got us there safely and on time, providing incredible insights and wisdom along the way. He has been the older brother I never had. Working behind the scenes, Eria helped me navigate the bureaucracy of Ugandan institutions, while Dineke, Betty, Karen, and especially Marian saw me through the bureaucratic maze of Wageningen University. Thank you for making my work so much easier.

Of course, the journey has not been only about work. In the office, Ire, Koen, Maria, Landry, and the rest of my fellow PhD students and colleagues kept me sane by reminding me that there is more to life than deadlines, imperfect identification strategies, and desk rejections. Similar reminders came from Samba, Marta, Loïc, Stefano, Myro, and João who gracefully put up with my periodic grumpiness at home. My thanks go to you all, as well as to all the friends who made my time in Wageningen so memorable.

Over the years, this journey took me to many places, both physically and intellectually. However, it would have never been possible without the love of my family at home, in the Czech Republic. Never doubting and always supportive, they encouraged me to take on new challenges, but also made sure that I have a place to come back to. They were there whenever I needed someone to talk to, but stepped back when I got too busy to call. Thank you for giving me the freedom to pursue my dreams.

Finally, my warmest thanks belong to my best friend and life partner, who has been there for me in good times and bad. Tuma, you are the reason why I look into the future with joy and excitement. Thank you!

Jan Duchoslav
Wageningen School of Social Sciences (WASS)
Completed Training and Supervision Plan



Wageningen School
of Social Sciences

Name of the learning activity	Department/Institute	Year	ECTS*
A) Project related competences			
Lab-in-field Workshop	University of East Anglia	2013	1.5
GIS for Social Sciences	GeoDesk	2013	1.5
Advanced Behavioural Economics	Utrecht University	2014	5
Theory and Practice of Efficiency & Productivity Measurement	WASS	2015	6
B) General research related competences			
Research Proposal	WASS	2014	4
Data Management	WUR Library	2014	0.4
<i>'Performance-Based Financing: The Ultimate Catholicon?'</i>	CSAE Annual Conference, University of Oxford	2014	1
Presenting with Impact	WGS	2015	1
Summer School in Development Economics	IDEAS / Università di Verona	2015	2
<i>'Institutional Change and the Dynamics of Social Capital'</i>	Symposium on Economic Experiments in Developing Countries, CERGE-EI, Prague	2015	1
<i>'Prenatal Trauma and 'Cooperation'</i>	CSAE Annual Conference, University of Oxford	2015	1
	Economics PhD Conference, University of Warwick	2016	
<i>'Prenatal Temperature Shocks Reduce Cooperation'</i>	Gorman Seminar, University of Oxford	2016	1
Summer School in Development Economics	IDEAS / Università di Verona / Monash University Prato Centre	2016	2
C) Career related competences/personal development			
WASS Introduction Course	WASS	2014	1
Regular PhD Meeting	DEC	2012-16	4
Intern Supervision	DEC	2012-13	1
Teaching Assistant	DEC	2013	1
Teaching Assistant	DEC	2014	1
Teaching Assistant	DEC	2015	1
Visiting Student	University of Oxford	2015-16	2
Total			38.4

* One credit according to ECTS is on average equivalent to 28 hours of study load

This research received funding from The Netherlands Organization for Scientific Research (NWO) as part of project number W 07.72.108, and under grant number 453.10.001.

Cover design by Vojtěch Duchoslav