Bilingualism, Hypothetical Norms & Misperceptions

Joshua Miles Edward Hill

3546667

A thesis presented for the degree of

Doctor of Philosophy

School of Global Development

University of East Anglia

United Kingdom

27 February 2025

Contents

List of Figures

List of Tables

Acknowledgments

The completion of this thesis was only possible due to the great support from a number of people and organisations, and I am deeply grateful to each and every one who was involved in this process.

First of all, I would like to thank my supervisors, Paul Clist and Arjan Verschoor, for their excellent and indispensable support throughout this journey. I am truly grateful for their intellectual guidance and inspiration, their generous and valuable advice, and for their trust in my academic abilities. Moreover, this has been a long journey, so I also deeply appreciate their tenacity and patience.

I would also like to thank Pieter Serneels and Bereket Kebede for supporting my confirmation process, in addition to their kind support and words of encouragement. I would also like to enthusiastically thank Joshua Balungira, and all of the facilitators and enumerators at The Field Lab. This thesis genuinely would not have been possible without your support, guidance and hard work.

I am grateful to the Centre for Behavioural and Experimental Social Science ('CBESS') and the South and East Network for Social Sciences ('SeNSS') for experimental funding and my doctoral studentship. This thesis would not have been possible without their support. I would like to thank the University of East Anglia (UK) and Lira University (Uganda) Ethical clearance Boards for their diligent review and valuable feedback.

I want to show my appreciation to all of the early stage researchers and faculty at the School for Global Development, as well as anyone who provided me with feedback at international conferences. Your comments and insights have been most constructive.

Lastly, I am also grateful for my collective personal experiences, and to all of those who shaped them, for helping lead me to a more promising and meaningful path.

Chapter 1

Introduction

"A good decision is based on knowledge and not on numbers"

Plato

Understanding human choice behaviour has been the enduring pursuit of economics, psychology and many other disciplines. These questions are of central importance: the world is more connected and sophisticated than ever, and humanity has never been required to make so many complex decisions, at such a rapid pace. Our hunter-gatherer ancestral homo-sapien forebears did not have to contemplate the intricacies of international monetary flows, climate change justice or artificial intelligence. The complexity and scale of modern society often eclipses the ability of our mammalian brains to comprehend, process and rationalise the world around us. Behavioural economics has sought to better understand how biases and heuristics (evolved for the savannah) affect decision making in the 21st Century, and importantly, how these insights can help society solve the great challenges of our time.

Human society now operates at an unprecedented scale, as do the challenges it faces. The global population surpassed 8 billion in 2022, and despite the United Nations' optimistic rhetoric, the international development agenda has largely fallen short of its ambitions (UNFPA, 2022; Bhandari, 2024). The 2024 UN-hosted Summit of the Future aimed to reinvigorate progress toward the faltering Sustainable Development Goals (SDGs) through a proposed 'Pact for the Future' (UN, 2024). Held in the aftermath of the COVID-19 pandemic, the summit failed to deliver concrete actions, and critical issues such as climate change received minimal attention. While existential threats like climate change and pandemics ought to unify humanity, context, culture, and self-interest remain powerful drivers of behaviour.

To navigate these global challenges, we must better understand what influences and motivates human behaviour. Individuals are not solely self-interested; they are shaped by a complex array of external factors that extend beyond *Homo Economicus*. Preferences vary significantly across populations (Falk et al., 2018), as do cultures, social norms, and languages. Recognising how context and culture shape decision-making is essential for effective policy design at local, national, and international levels. This thesis explores these dynamics, with a particular focus on the role of social norms, language, and context. For much of the world, multiculturalism is a lived reality, with distinct cultural practices embedded in local norms. Language, as a key repository of culture, plays a central role—most people globally speak at least two languages daily (see Crystal, 2012). Understanding how these factors influence individual choices at the moment of decision-making offers valuable insights for policymakers.

While 80% of the global population lives in Non-Western, Educated, Industrialised, Rich and Democratic ('WEIRD') countries, a remarkably high proportion of economics research has been conducted in the United States and Western Europe (see Sharma and Siddique (2024) for a recent literature review). Indeed, this is also where the increasing majority of the world will continue to live (UNDESA, 2024). For this reason, all field work in this thesis has been conducted in Uganda. Uganda is a linguistically and culturally diverse country of about 47 million people and 44 languages, and at least 20 distinct indigenous ethnic groups (Eberhard et al., 2019). This cultural diversity is

representative of much of the developing world. This makes Uganda the perfect country in which to explore the central questions of this thesis, being how language, context (whether hypothetical or actual) and misperceptions may influence decision making, and how a better understanding of these factors can lead to more effective policies to support the positive changes our society needs.

This thesis combines behavioural and experimental economics in a developmental context. Social norms, values and preferences serve as the common thematic structure throughout this thesis and Chapter 2: *'Norms, Values & Preferences'* summarises the key theoretical and substantive literature in this area. Chapter 3: *'Uganda - Culture & Context'* presents a brief history of Uganda, including economic background, social context and deep-dives on the three ethnic groups of interest throughout this work: the Gisu, Ganda and Acholi. The proceeding three chapters comprise the empirical heart of this work.

Chapter 4: *'Bilinguals in the Lab: Why do Norms and Expectations not Predict Contributions?'* explores whether the language you speak affects the decisions that you make?[1] To answer this, this chapter randomly assigns the language of a public goods game for bilingual subjects to determine whether there is a difference, and if so, I explore why this is the case. This chapter initially replicates the largest previous result using 349 subjects in Uganda. I find that subjects playing a one shot public goods game in Luganda (the national language) contribute 28.9% more than those playing in Lugisu (their tribal language). This effect size is within 0.1 percentage points of the previous finding. I rule out several confounds, and test three mechanisms. First, I am able to reject the idea that language changes behaviour by activating different norms, as they do not differ by language. Second, language does not act as a coordination device, as expectations do not differ by language. Third, this leaves the explanation of different preferences in different languages. Rich anthropological evidence describes the Gisu as having a self-sufficient and non-cooperative culture. Speaking the associated language appears to activate this low-cooperation cultural frame, whereas speaking another leads to higher cooperation. These results show a large and robust language effect on behaviour, which in this case is best explained by language-dependent preferences for cooperation.

[1]This chapter is joint work with my primary PhD supervisor Paul Clist and was pre-registered in 2019 (Clist and Hill, 2019). I was jointly responsible for developing the 2019 research idea and study design, developing the pre-analysis plan, organising and overseeing the data collection in Uganda, conducting the statistical analyses, and writing the first draft of the paper. My co-author gave input on the research design, analyses and was involved in writing the draft of the paper being prepared for publication.

Chapter 5: *'Hypothetical Norms & Cultural Exposure'* explores whether individuals house multiple sets of social norms (as in Hoff and Stiglitz (2016)) and whether these can be elicited hypothetically across different contexts (by using a modified version of the Krupka and Weber (2013) methodology)[2]. The context applied is twofold: geographical 'place' (Kampala / Mbale, 2019 experiment) and 'frame' (Community / Wall Street Game, 2024 experiment). I find that norms differ significantly by both place and frame, but that expectations do not. In the 2019 experiment, because of the bilingual design, I also find a significant interaction between place and language for norms. This means that speaking either Lugisu or Luganda in Mbale affects norms, but norms are unaffected by speaking either language in Kampala. For the 2024 experiment, I find that actual and hypothetical contributions and norms are approximately equal, in that there is insufficient evidence to reject the null that they are equal. This was not found to be true for expectations. This chapter provides methodological enhancements which may be of use to policy makers; not only can hypothetical norms be measured out of context, but I have also shown that these are approximately accurate in signalling actual behaviours.

Chapter 6: *'Social Norms, Value & Misperceptions'* explores the nature of misperceptions and how these differ between in and out-groups[3]. Looking at political in-groups in the USA, several studies have found Republicans tend to think other Republicans are more like them than they are, and that Democrats are less like them then they are (Bursztyn and Yang, 2022). These studies also suggest that Republicans are more accurate in their perceptions of Republicans when compared to Democrats. Building on this, I explore similar ideas but instead of focusing on political identity in the USA, I use ethnic identity in Uganda. I replicate and expand these finding and extend the analysis in a number of unique ways. My design considers five domains (politics, religion, deception, violence and adultery), across three ethnic groups (the Gisu, Ganda and Acholi) in a non-WEIRD setting (Uganda). My design also allows the underlying mechanisms to be determined, for both the in and out-group. I find that what one believes about others (second order beliefs) is significantly influenced by own beliefs (first order beliefs), and this is more extreme for the in-group than the out-group. Considering the mechanisms, I find evidence that Motivated Reasoning and Cultural Identity are significant drivers of perceptions for both in and out-groups, though this varies by domain. For the in-group only, the Curse of Knowledge is also significant across all domains. Considering inaccuracy, participants are less accurate about their in-group despite feeling most confident about these responses. These results will be of value to policy makers and researchers alike. Misperception corrections have become an increasingly popular policy tools, and they have been shown to be effective, at least in the short term. However, this chapter highlights a number of potential situations where outcomes may not be as expected.

[2]This chapter is joint work with my primary PhD supervisor Paul Clist. Both the 2019 (Clist and Hill, 2019) and 2024 (Clist and Hill, 2024) experiments were pre-registered. I was jointly responsible for developing the 2019 research idea and study design, developing the pre-analysis plan, organising and overseeing the data collection in Uganda, conducting the statistical analyses, and write-up. My co-author gave input on the research design and analyses. For the 2024 experiment I led on all aspects of the experimental design, developing the pre-analysis plan and associated analysis, ethical clearance, data collection in Uganda (with the support of The Field Lab), analysis and write-up. My co-author contributed to aspects of the research design and feedback on the pre-analysis plan.

[3]This is the author's sole work and was pre-registered in 2024 (Hill, 2024).

Lastly, chapter 7 *'Conclusion'* concludes and summarises. This thesis makes several contributions to the literature. Firstly, Chapter 4 replicates the study with the largest existing effect size to within 0.1%, despite changing both the location (Mable to Kampala) and the sample demographic. Secondly, this chapter explores the underlying mechanisms for the first time, ruling out both norms and expectations and concluding that preferences, at least in this instance, differ by language. Chapter 5 contributes methodological improvements to the Krupka and Weber (2013) approach by introducing hypothetical elements, across both place (within-subject) and by frame (across-subject). This demonstrates that not only can different norms be elicited hypothetically, but also that hypothetical responses can also be used to approximately determine actual results. Additionally, this chapter explores for the first time whether there are interaction effects, by language, between hypothetical and actual norms and expectations. Chapter 6 makes numerous contributions, by expanding beyond the USA political domain, where groups are only differentiated by political beliefs (i.e. Democrats vs Republicans). The experimental design and analysis in this thesis considers religion, deception, violence and adultery (in addition to politics), in a non-WEIRD setting, across three distinct ethnic groups (one in-group, two increasingly distant out-groups). Secondly, I contribute by investigating four potential mechanisms driving the noted difference in misperceptions between in and out-groups, which have only been hypothesised in the literature to date. Overall, this thesis provides numerous insights which may be of use to local policy makers attempting to promote behaviours which help to address some of the larger challenges we face.

Chapter 2

Norms, Values & Preferences

"Man is as much a rule-following animal as he is a purpose-seeking one."

F.A. Hayek

2.1 Definition of Key Terms

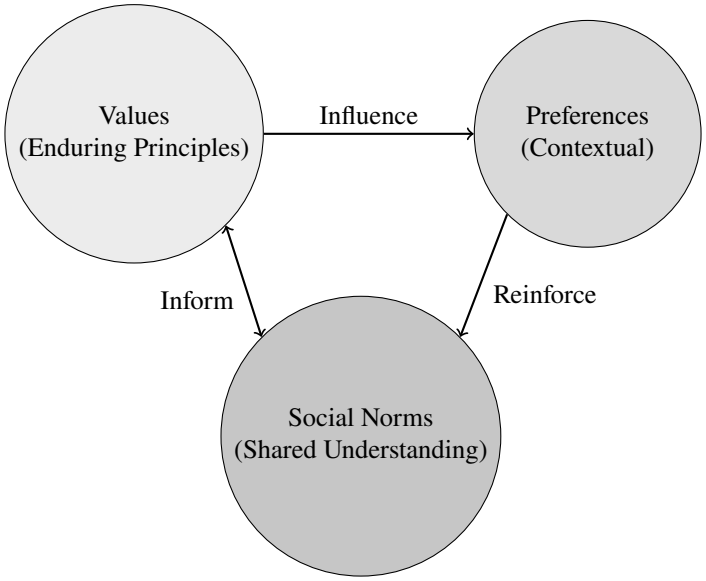
Social norms are critical to understanding human behaviour, influencing actions and interactions in diverse social contexts. These are unwritten rules or expectations about how individuals should behave in specific social situations. Norms are often context-dependent and can vary between cultures, communities, and even different social settings within the same culture (Cialdini et al., 2006). They are also closely related to the concept of values and preferences, though are distinct. The literature around each of these concepts is diverse and multi-disciplinary, thus definitions can vary. Though crudely, social norms set out a behavioural expectation as to how one *should act*, whereas values represent principles, worldviews, narratives and beliefs of what is desirable, or in other words what one *should believe* (Schwartz, 2012). Values and social norms provide the foundation, and reinforce or constrain preferences, being the optimal choice, given a specific context, an agent adopts (Samuelson, 1948).

From a psychological perspective, Warren et al. (2011) argue that preferences are constructed and shaped by immediate context, emotions, and social norms, while values are more abstract and enduring. In other words, preferences, or the choice which maximises an individual's utility, is dependent on factors such as the pull of social norms that exist in that context, the values an individual holds, and the emotional lens or frame being applied in the moment. Hoff and Stiglitz (2016) conceptualise this in a simple taxonomy where an individual's preferences and cognition are subject to two deep social influences: (a) 'social context' (exposure to environment); and (b) 'mental models' (cultural categories, identities and narratives) used to process information. An individual may have multiple mental models (or schema) which shape how they will interpret, remember and emotionally respond to the information they encounter (DiMaggio, 1997). These mental modes can be inconsistent (or contradictory) and thus lead to very different choice behaviour, depending on context (d'Andrade, 1995; Swidler, 2013). Importantly, social context creates the set of mental models available to the individual, and affects the scenarios that prime alternative mental models (Hoff and Stiglitz, 2016).

Schwartz (1992) defines values as trans-situational goals that vary in importance and serve as guiding principles in life. These can be considered the enduring underlying foundations, routed on principles such as self-direction, achievement, hedonism and power (see Schwartz (2012) for a comprehensive overview). Values provide the foundation for preferences, while preferences reflect values in specific contexts. Social norms, which are fundamentally a shared understanding of what is normative and expected behaviour (see Bicchieri (2016) for summary), can inform values by shaping what is perceived as acceptable or important within a society. Social norms reinforce or constrain

preferences by establishing boundaries for behaviour that align with group expectations (i.e. the pull of norms). These relationships are summarised in figure 2.1. While values inform social norms (and vice versa), considering population level heterogeneity, at the individual level it is possible that values and norms may disagree.

Figure 2.1: Fundamental Relationship Between Social Norms, Values & Preferences



As social norms are featured across a range of distinct disciplines, I provide a number of standard definitions here, consistently applied throughout this thesis. *Descriptive norms* indicate what behaviours are shared within a specific group or culture. They influence individual behaviour by providing information about what is 'typical' or 'normative' (Cialdini et al., 1991; Bicchieri, 2010). Descriptive norms can affect behaviour, for instance, when individuals learn that their peers engage in a specific behaviour, they may also feel pressured to adopt it in order to 'fit in'. Closely linked are *injunctive norms* (sometimes also called *prescriptive norms*), which set out what behaviours are perceived as socially acceptable or unacceptable within a group. They often elicit an emotional response, such as regret or fear of ostracism, which can motivate individuals to conform (Cialdini et al., 1991). Research shows that emphasising injunctive norms can be particularly effective in promoting pro-social behaviours (Cialdini et al., 2006; Schultz et al., 2007), and promoting desired behaviour change more generally (Kollmuss and Agyeman, 2002). *Personal norms* are internalised standards which individuals feel personally obligated to follow. These norms are shaped by moral beliefs, values, and the desire to act in accordance with one's self-concept (Schwartz, 1977; Chung and Rimal, 2016).

2.2 Measuring Social Norms

Across disciplines, a variety of methods have been employed to study social norms. These include both qualitative techniques, such as ethnographic studies (particularly prevalent in anthropology and sociology), and quantitative approaches, such as surveys and experiments. Experimental designs often use RCT (Randomised Controlled Trails) methodologies to assess the impact of norm-based interventions, and typically adopt lab-based methodologies from psychology (Camerer et al., 2004). I will focus specifically on efforts to measure social norms in the economics literature, as this is most relevant to this thesis, though it is important to note the cross disciplinary influences both theoretically and methodologically.

The inclusion of social norms as a determinant of behaviour and decision making marks a departure from classical individualistic thinking, towards a more social model of choice behaviour. Indeed a large number of experiments have produced evidence which infers that social norms can affect decision making across a large range of economic and social scenarios (Nosenzo and Görges, 2020). For example, norms have been shown to influence cooperative behaviour (Reuben and Riedl, 2013), honesty (Abeler et al., 2019), fair sharing (Krupka and Weber, 2013; Gächter et al., 2013), corruption (Gneezy et al., 2020) and discrimination (Barr et al., 2018). Fundamentally, this evidence suggests that individuals care about how others perceive them in reference to a given norm (Akerlof and Kranton, 2000; Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016; d'Adda et al., 2016). Individuals are motivated to adhere to social norms, even when this leads to choices which contradict with an individual's self-interest or intrinsic preference. This is because doing so spares them the disapproval (or gains the approval) of their peers. Thus, the utility lost/gained from social disapproval/approval offsets the opportunity costs of an increase/decrease in self interest (see Bicchieri (2006, 2016); Fehr and Schurtenberger (2018) for discussion).

Economists have devised three specific ways of measuring social norms: the Belief Survey Method, the Krupka-Weber Method and the Opinion Matching Method. The *Belief Survey Method* is the most simplistic approach and asks participants to provide second order beliefs to a survey question, specifically how would 'most other people' would rate the appropriateness of various actions available to an agent in a hypothetical decision situation (Nosenzo and Görges, 2020). This method is simple to administer and also allows the distributions of relative social acceptability to be captured (i.e. how the social acceptability of alternative actions may differ). The drawback of this methodology is that it is not incentivised. This introduces two issues. First, as these tasks are cognitively demanding, participants may provide only partially considered responses, introducing bias. Second, respondents may be tempted to respond in a socially desirable manner as their responses are often directly observed (Rustichini and Villeval, 2014).

Krupka and Weber (2013) (the *Krupka-Weber Method*) introduce a methodology which is similar to the Belief Method, but is unique in that respondents are paid monetary rewards if they rate the appropriateness of behaviour in the same way as most other respondents do. In other words, this conceptually removes both of the drawbacks of the

Belief Method, as participants are extrinsically incentivised to coordinate with their actual belief of others and what they perceive to be socially (un)acceptable. Indeed subsequent analysis has shown that the Krupka-Weber Method does not suffer from significant levels of bias (Erkut, 2020; d'Adda et al., 2016). However, the methodology has received criticism as other aspects of the design may act as coordination devices which may be cognitively more accessible or substitute for social norms in situations where an established norm does not exist. However this concern has been somewhat allayed (Fallucchi and Nosenzo, 2022). A second more challenging concern is whether the method elicits second order beliefs, or whether what is actually being measured is third (or more) order beliefs, or a combination thereof.

Nosenzo and Görges (2020) introduce the *Opinion Matching Method*, following numerous concerns raised in the literature (for example, see Bicchieri et al. (2022)). This method consists of a two-stage procedure. In stage 1, a group of participants are asked to state their belief (or opinion) about the appropriateness of a range of actions (on a Likert or similar scale). For stage 2, participants are informed about stage 1 and are then asked to guess the most common response. In other words, they are asked to explicitly state their second-order beliefs of appropriateness, which are incentivised. This method has a few advantages relative to the two discussed above. As well as maintaining the Krupka and Weber advantages over the Belief Method, the sequential elicitation of first and second-order beliefs removes the strategic component that is present in the coordination game of Krupka and Weber and also has a clear focus on second-order beliefs. However, this methodology has two disadvantages. This first is that the method is more cumbersome to deploy as it requires two iterations of questions between subject, or two separate groups, which is both time consuming and costly. Secondly, the first-step responses are not incentivised as these are personal beliefs/opinions. This means that these responses could be vulnerable to responding biases, potentially even more so than the responses collected with the Belief Survey Method. This could undermine the measurement of both the first and second order beliefs.

Chapters 4 and 5 of this thesis make use of the Krupka and Weber (2013) methodology for three key reasons. The first is that, when the experiment was being design and conducted in 2019, this methodology was the most widely applied. As the 2024 experiment (relevant for only chapters 5 and 6) expanded elements of the original experiment, it was preferable to align the design (where appropriate) to the original 2019 experiment. Further, both experiments were time consuming and approximately the required size based on pre-registered power analysis. Given the limited funding available for this thesis, it would have been difficult to expand either the experimental duration or cost. Lastly, while it is not possible to tell whether second order beliefs were accurately measured, any bias introduced within the design would have been uniformly applied across different treatments. As across-subject analysis, between language or context-based treatment, is the primary focus of this work, it's questionable whether applying the Opinion Matching Method would have been of benefit.

2.3 Brief Theoretical and Historical Context

The concept of social norms has been discussed by numerous thinkers throughout history, but its systematic exploration as a field of study within the social sciences emerged in the late 19th and early 20th centuries. Émile Durkheim (1858-1917) is often regarded as the first to systematically discuss the concept of social norms within a sociological framework. In his works, particularly *The Division of Labour in Society* (Durkheim, 1893) and *Rules of Sociological Method* (Durkheim, 1895), he explored how norms maintain social order and cohesion and thus are essentially *'Functionalist'* in nature. Durkheim emphasised the role of collective consciousness in reinforcing norms, suggesting that norms provide stability in a society by aligning individual behaviour with group expectations. This was then notably expanded by Gabriel Tarde (1843-1904), who emphasised the role of imitation and social interaction in the development of norms. His work laid the foundation for understanding how norms spread within groups and societies (see Barry and Thrift (2007)).

Although Max Weber (1864-1920) did not explicitly focus on norms as a stand-alone concept, his analysis of social action in *Economy and Society* (Weber, 1921) touched on how norms influence behavior by defining what is considered legitimate or acceptable within a group. By the the mid-20th century, the understanding of norms expanded through the works of economists and sociologists, notably with the arrival of *Game Theory* (Nash, 1950) and *Structural Functionalism* (which extends Durkheim's work to include social structures and functions, jointly acting so that society behaves and evolves like an organism - see Parsons (1951)). The Game Theoretic framework illustrates how social norms can emerge as equilibria in repeated interactions. Elster (1991) discusses how rational actors develop norms to facilitate cooperation and reduce conflict, suggesting that norms are often the result of strategic interactions.

Bandura (1977) introduced Social Learning Theory, which suggests that individuals learn norms through the observation and imitation of others. This perspective highlights the importance of role models and reinforcement in the internalisation of social norms, and indeed experimentally understanding these themes represent some of the open questions in the field today (Nosenzo and Görges, 2020). Tangentially, Bicchieri (2006) was one of the first to argue that social norms are constructed through interactions and shared expectations within communities (i.e. that they are *'Constructivist'*). This dynamic perspective acknowledges that norms are not static; they adapt as societies evolve and new interactions occur.

2.4 Evolution, Formation and Transmission of Social Norms

The evolution of social norms reflects the shifting structures of society across historical periods. In pre-industrial societies, norms were primarily shaped by kinship ties, communal living and subsistence economies. These norms

emphasised collective well-being and were reinforced through oral traditions and rituals (Durkheim, 1893). As societies transitioned into agrarian systems, the domestication of plants and animals led to the development of hierarchical structures and property ownership, which introduced norms around inheritance, land use and labour division (Weber, 1921; Diamond, 1998; Harari, 2014). This period also saw the emergence of formalised legal systems to regulate increasingly complex social interactions. With industrialisation, social norms underwent significant transformation. Industrial societies prioritised individualism, efficiency, and economic productivity, driven by urbanisation and the proliferation of wage labour (Parsons, 1951). This shift marked a departure from traditional, community-based norms to those centred around bureaucratic institutions and capitalist values (Barry and Thrift, 2007). These historical transitions demonstrate the dynamic nature of social norms as they adapt to the changing needs and structures of society.

Social norms evolve through a dynamic interplay of cultural transmission, social learning, and environmental adaptation. They arise from the collective behaviours and expectations of a group, shaping and regulating individual actions to maintain social order (Bicchieri, 2006). Initially, norms often emerge as informal guidelines that address immediate communal needs, such as cooperation for resource sharing or conflict resolution. Over time, these informal rules become institutionalised through mechanisms like laws, traditions, and rituals, solidifying their role in society. Boyd and Richerson (1987) argue that the transmission of norms occurs through social learning pathways; vertical (parent to child), horizontal (peer to peer), and oblique (from leaders or elders to younger generations). This allows cultural traits to persist or adapt depending on their perceived utility. This evolutionary process ensures that norms responding effectively to environmental or societal pressures are retained, while maladaptive ones are discarded. However, some norms can become 'evolutionary kludges[1]', persisting even when they no longer serve their original purpose or when they hinder adaptive responses to new environmental conditions (Nunn, 2021). Outdated norms may continue to exist due to their historical entrenchment and the costs associated with changing them, resulting in a mismatch between current societal needs and the norms that govern behaviour, leading to maladaptive behaviours and social friction.

The evolution of social norms is also influenced by structural changes in society, such as technological advancements, economic shifts, and demographic changes. Rogers (1994) highlights how evolutionary pressures (such as resource scarcity and uncertainty) shape behaviours and preferences, which can in turn influence the development of norms prioritising short-term or long-term benefits. For example, in stable agrarian societies, norms emphasising familial loyalty and inheritance persisted, while industrialisation gave rise to norms prioritising individualism and efficiency. Giuliano and Nunn (2021) emphasise that environmental stability promotes cultural persistence, while volatile environments foster adaptive change, allowing norms to evolve in response to shifting social and economic

[1] An evolutionary kludge refers to a biological trait or system that has developed through evolution in a piecemeal, inefficient, or inelegant way. This is often because natural selection works with existing structures rather than designing optimal solutions from scratch. The concept and language is borrowed from engineering and computing (Granholm, 1962), where a kludge is a workaround or makeshift fix that's clumsy but functional.

conditions. Thus, the evolution of social norms reflects a continuous balancing act between tradition and innovation, shaped by both historical legacies and contemporary challenges.

2.5 Mechanisms for Norm Change & Adherence

Social norms embody intergenerational wisdom and have historically acted as social heuristics, particularly in pre-agrarian societies, where written language remained undeveloped. This legacy and role can mean that social norms can persist and are often resistant to change. Young (2015) explores the tension between stability and change in social norms, noting that while some norms may remain stable over time, others may be subject to rapid change, with change appearing to be stochastic. Despite potential changes, many norms exhibit resilience due to their deep-rooted nature in social structures and individual identities. Norm changes often require sustained efforts and the presence of alternative models of behaviour to encourage new norms. Cultural transmission across generations represents the maintenance of the status quo. However, social learning and feedback mechanisms introduce pressure for norms to evolve. Bikhchandani et al. (1992) propose that cultural phenomena like fads and fashions can be explained through informational cascades, where individuals base their decisions on the observed choices of others rather than their own private information. Thus, collective actions, shifts in public opinion, and the influence of social movements can facilitate changes in social norms, either permanently or temporarily. Conflicts between established and emerging norms can lead to significant changes. For instance, changing attitudes towards smoking and environmental conservation have prompted shifts in public norms due to increased awareness and advocacy (Berkman et al., 2000). Institutions are also important, including legal systems and educational bodies, which play a crucial role in shaping and enforcing norms. Laws can codify social norms, leading to their widespread acceptance and internalisation (Elster, 1989a).

Individuals often conform to social norms due to the perceived social pressure and influence, where the desire for acceptance leads to behaviour alignment (Asch, 2016). This phenomenon is further reinforced by informational social influence, where individuals look to others for guidance in ambiguous situations. Compliance is maintained through formal mechanisms, such as laws, and informal mechanisms, such as social sanctions (Elster, 1989a). Informal enforcement can include social ostracism or ridicule, which are powerful deterrents against non-conformity. Peer pressure is one of the most commonly cited mechanisms enforcing norm compliance. Norms create expectations that can lead to peer pressure, encouraging individuals to conform to group behaviours. When individuals perceive strong descriptive norms, they are more likely to adopt behaviours to align with group expectations (Cialdini et al., 1990). This is particularly true where social norms instil a sense of moral duty, driving individuals to participate in collective actions that align with group values.

There are numerous empirical studies which illustrate the power of peer effects to drive changes in norms

across numerous domains. For example, environmental activists may feel a moral imperative to engage in conservation efforts due to collective beliefs about environmental stewardship (Bennett et al., 2018). Social norms have also been shown to play a crucial role in promoting sustainable energy use, by providing consumption data relative to peers/neighbours (Schultz et al., 2007). Norms surrounding civic participation can impact voter turnout and community involvement. Lastly, Bennion and Nickerson (2011) demonstrate that normative messages encouraging civic engagement can significantly enhance participation rates, particularly among young voters.

2.6 Misperceptions

By nature a normative standard is commonly known by all group members, where group membership is finite (Fehr and Schurtenberger, 2018). This common understanding is fundamental to the operation of social norms, but differences in norm certainty (and/or accuracy) across group members can lead to a wide variation in how social norms actually drive behaviour (Bursztyn and Yang, 2022). Bicchieri (2010) is typical in distinguishing between normative beliefs (the shared view of what is socially acceptable) and empirical expectations (the extent to which others obey the norm). In order for a social norms to hold sway over actions they need to be established, with normative beliefs and empirical expectations being aligned (Bicchieri, 2006, 2016). In other words, unless norms are commonly and accurately known, and group members are generally expected to adhere to those norms, they do not function as common reference point to inform decision making.

Bursztyn et al. (2020) discuss *pluralistic ignorance* and the importance of second order beliefs in understanding how social norms may be driving behaviour. Saudi Arabia has low female labour force participation. When measuring first order beliefs of men in Saudi Arabia, privately the majority of men do not strongly believe that women should not participate in the workforce. Thus, it seems that norms are not participating to behaviour. However, when measuring second order beliefs (i.e. what men believe other men believe) the influence of norms becomes obvious - most believe that others believe it is not acceptable. This misalignment between first and second order beliefs has been termed 'misperceptions'. This illustrates the difference between private (first order) beliefs (which can also be conceptualised as a preference or taste), and normative (second order) beliefs. Here a norm is driving actual behaviour (i.e. normative beliefs and empirical expectations are aligned), but the normative belief is fundamentally mis-calibrated compared to the actual distribution of first order beliefs.

Misperceptions are particularly prevalent in domains such as politics, public health, and climate change, where cognitive biases and social dynamics strongly influence belief formation. Nyhan and Reifler (2010) identify confirmation bias and identity-protective cognition as central mechanisms reinforcing false beliefs. Individuals are more likely to accept information that aligns with their existing worldviews while rejecting contradictory evidence,

especially on polarising issues. These dynamics are further exacerbated by the Dunning-Kruger effect, which suggests that individuals with limited knowledge are often unaware of their own ignorance (Kruger and Dunning, 1999). Bursztyn and Yang (2022) summarise a large literature, focusing specifically on in vs out-group biases in both perceived likeness and accuracy. They suggest four potential mechanisms from the pyschological and economics literature: the Curse of Knowledge (Camerer et al., 1989), Projection Bias (Madarász, 2012), Motivated Reasoning (Kunda, 1990) and Cultural Identity (i.e stereotypes - Bonomi et al. (2021)). This reveals the diversity of both cognitive and cultural biases potentially driving misperceptions.

A significant body of work examines the persistence of misperceptions. Corrective efforts, such as fact-checking, often fail or even backfire, leading to strengthened false beliefs ('the backfire effect' - Nyhan and Reifler (2010)). This is particularly common when corrections challenge deeply held ideological commitments, specifically those steeped in cultural stereotypes. However, Wood and Porter (2019) suggest that backfire effects may be less common than initially thought and that corrections can reduce misperceptions under certain conditions. Further, techniques such as pre-bunking (providing prior warnings about mis-information) and re-framing corrective messages to align with individual's values can be effective (Lewandowsky et al., 2017). Fact-checking and trusted messengers have shown some success, but their impact is limited by ideological polarisation and declining trust in traditional media and experts (Flynn et al., 2017). However, on balance, Bursztyn and Yang (2022) find that attempts to correct misperceptions have been attributed to significant behavioural change, particularly in the short-term.

2.7 Conclusion

This chapter has explored the relationships between social norms, values, and preferences, highlighting their significant influence on human behaviour and decision-making. Through a review of key theoretical literature and empirical evidence, I have explored how social norms are not only shaped by, but also shape individual values and preferences. The dynamic interplay between these elements underscores the complexity of human decision making, where social context and mental models play pivotal roles in behaviour, crucially at the point at which a choice is made. This chapter has also highlighted the key methodological developments in how social norms can be measured, including the limitations and advantages of each method.

Further, I have examined the evolution and transmission of social norms, summarising how historical, cultural and environmental factors have contributed to their persistence and change. The mechanisms for norm adherence and change, including social learning, peer pressure, and institutional influences, have been discussed, emphasising both the resilience and adaptability of social norms. Lastly, this chapter has discussed misperceptions and their relationship to social norms. I have also highlighted their potential impact on behaviour, a number of the commonly cited mechanisms,

as well as efforts to correct mis-calibrations.

Chapter 3

Uganda - Culture & Context

3.1 Introduction

Uganda is a linguistically and culturally diverse country of about 47 million people and 44 languages (Eberhard and Fennig, 2022). Indeed Alesina et al. (2003) estimate it as the most ethnically and linguistically diverse country in their dataset of around 200 countries. This cultural diversity makes Uganda the perfect context in which to explore the central questions of this thesis. This chapter provides a summary of the history and cultural diversity of Uganda, with a specific focus on the three tribal/cultrual groups of interest: the Gisu, Ganda and Acholi.

This chapter provides a number of priority insights which underpin each of the empirical chapter of this thesis. Fundamentally, this chapter explores how Uganda's complex colonial and post-colonial history has profoundly shaped tribal identities, governance structures, and patterns of development. The Gisu, with their egalitarian social organisation and rich cultural heritage rooted in the eastern highlands, are typified by their community resilience and individual autonomy. In contrast, the Ganda's historically privileged position under British rule and their centralised kingdom structure have afforded them enduring political influence, while the Acholi have been deeply affected by conflict and marginalisation. These key insights are essential in understanding how behaviour driven by language, social norms or perception, may differ between groups.

Figure 3.1 presents a map of Uganda, highlighting (in blue) the three geographical regions in which data was collected or lab-in-the-field experiments were conducted. [A] highlights Mbale, a major urban city in eastern Uganda. Mbale is located in the Bugisu sub-region, which is the home to the Gisu, the primary cultural group of interest throughout this work. [B] highlights Kampala, the largest and capital city of Uganda, located in central Uganda. While Kampala is increasingly culturally diverse, it is located in the Buganda sub-region, home to the Ganda. [C] highlights Gulu, located in northern Uganda and the major urban centre in the Acholi sub-region of Uganda. This region is home to the Acholi. Mbale, Kampala and Gulu have population sizes of 290,000, 1,875,000 and 135,000 respectively and, as per the latest 2024 census, there are 2.2 million Gisu, 13.0 million Ganda and 2.0 million Acholi in Uganda (UBoS, 2024).

Figure 3.1: Map of Uganda Showing Regions of Interest

Adapted from https://www.nationsonline.org/oneworld/map/uganda-map.htm, accessed 14/01/2024

Figure 3.2 sets out a range of development statistics for Uganda. Despite significant progress over recent decades, the HDI index rates Uganda as having a low level of human development, with around 40% of Ugandans below the $1.90 per day poverty threshold, meaning Uganda ranks 159th of 189 countries and territories (World Bank, 2022). Falk et al. (2018) provide country-level averages for a range of social preferences in 76 countries, finding that Uganda is in the lowest quartile for altruism, positive reciprocity and negative reciprocity. It also has the lowest level of average trust in the sample, at around 0.7 standard deviations below the global average.

The population of Uganda (around 47.1 million in 2022) has been growing rapidly, with 5.0 live births per

women in 2022. This continues to place significant strain on infrastructure, health, and education systems (World Bank, 2023b). Despite economic growth at an estimated 6.7% in 2022, GDP per capita remains low, around $692 in 2022. Uganda's economy is largely based on agriculture, which employs a major part of its workforce, though there has been a gradual shift toward industry and services, particularly tourism and small-scale manufacturing (World Bank, 2024). Health and education indicators have shown gradual improvement over the last 50 years. Life expectancy, currently at around 65 years for females (60.4 years for men), has increased due to better access to healthcare and reduced infant mortality rates, which now stand at approximately 46.1 per 1,000 live births. While these are still significantly below the global average, Uganda has made steady progress on all fronts (World Bank, 2023a). Literacy rates are around 77%, and access to primary education is widespread, although secondary and tertiary educational still remains relatively limited (UNESCO, 2021).

Figure 3.2: Summary of Development Metrics, for Uganda

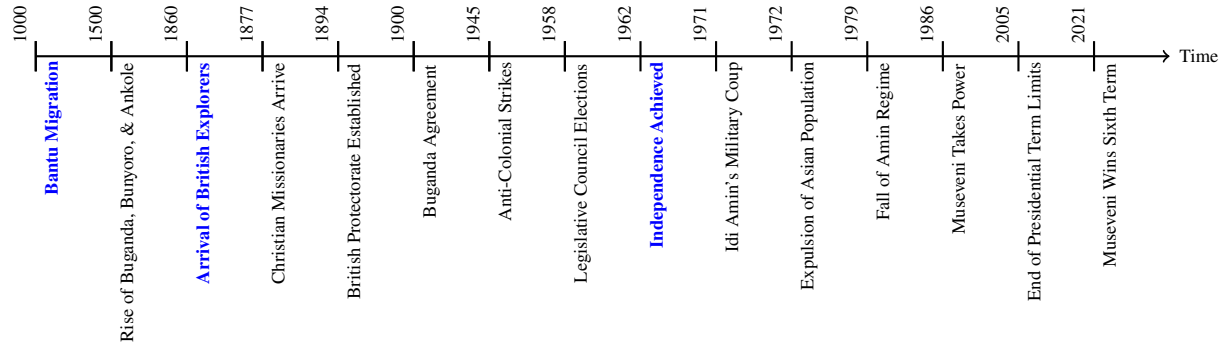
Development Metric	**Year**	**Value**
Population	2021*	47.1 million
GDP (Nominal)	2021*	$32.6 billion USD
GDP Growth Rate	2021*	6.7%
GDP per Capita	2021*	$692
Life Expectancy	2021*	65.0/60.4 years (females/males)
Infant Mortality Rate	2021*	46.1 deaths per 1,000 live births
Education: Government expenditure (% of GDP)	2021*	2.1%
Access to Electricity	2021*	28.5%
Fertility Rate (live births per women)	2021*	5.0
Unemployment Rate	2021*	1.9%
Access to Clean Water	2021*	79.2%
Human Development Index (HDI)	2022**	0.528
Gender Inequality Index (GII)	2022***	0.554

Note: *Taken from UN data 2021 `https://data.un.org/en/iso/ug.html`. **Taken from UNDP data 2022 `https://hdr.undp.org/data-center/human-development-index#/indicies/HDI`. ***Taken from UNDP data 2022 `https://hdr.undp.org/data-center/thematic-composite-indices/gender-inequality-index#/indicies/GII`. All data accessed 08/11/2024.

Uganda continues to face significant developmental challenges, especially in areas like electricity access, which is only available to about 28.5% of the population (World Bank, 2021). Additionally, gender inequality remains a significant issue, reflected in Uganda's Gender Inequality Index (GII) score of 0.554 (UNDP, 2021). Urbanisation continues to grow, with 23.6% of the population living in urban areas, adding pressure to urban infrastructure and services (UNDP, 2023).

3.2 Brief History of Uganda

Figure 3.3: A Brief History of Uganda



Uganda is located in East Africa and has a varied history summarised in figure 3.3. This history can be conceptualised in three key periods: i) The rise of indigenous kingdoms, notably the Bantu, from West Africa around 1000 AD; ii) European Colonisation c. 1860 AD; and iii) Post-colonial (1962+) governance, social and economic challenges.

3.2.1 The Rise of Indigenous Kingdoms c.1000 AD

This region's early history is characterised by the Bantu, and later, Nilotic migration around the first millennium AD. This brought agricultural and iron-working technologies from West Africa and the establishment of settlements in fertile regions around Lake Victoria and the western highlands. In addition to physical technology, these migrations also imported social structures and hierarchy, leading to the establishment of powerful kingdoms and clan-based communities (Twaddle, 1993). By the 15th century, centralised kingdoms like Buganda (hereafter referred to simply as the 'Ganda'), Bunyoro and Ankole had emerged, each with distinct governance structures, economic systems, and cultural practices (Reid, 2017). Of these, the Ganda became particularly influential, establishing a strong centralised monarchy and a structured, hierarchical society based on clan networks (Twaddle, 1993). Hansen and Twaddle (1988) argues that this shift from clan-based societies to organised states helped communities accumulate wealth and power, shaping early governance and cultural practices, ultimately leading to nation building.

The earliest notable large scale society was the Bunyoro-Kitara kingdom, which controlled large parts of the Great Lakes region (in western Uganda, near Lake Albert). The Bunyoro's influence extended widely, managed through an increasingly sophisticated social and administrative structure, which was centralised around a number of controlled tributary states (Low, 1971). However, by the 17th century, the Ganda had expanded and rivaled the Bunyoro

both organisationally and militarily, mainly due to their more advantageous positioning near Lake Victoria (Twaddle, 1993). The clan system played a central role in organising social life both for the Ganda and Bunyoro. The clan system provided cohesion and allowed for cultural integration of various groups through political assimilation (Low, 1971). For example, each clan worshipped its own unique deities and maintained specific rituals. These beliefs reinforced loyalty to local leaders and fostered unity within communities (Twaddle, 1993).

By the 18th century, long-distance trade was common, linking Uganda's kingdoms to regions beyond East Africa. The Ganda emerged as a key participant, engaging in the trade of iron, ivory and salt. These trade connections introduced foreign goods, cultures and ideas, gradually integrating Uganda's kingdoms into broader East African economic systems (Low, 1971). Contact with Arab traders, who were active in the region, and other African societies brought new crops, such as bananas and cassava, which became agricultural staples. This increased the Ganda's agricultural productivity and enabled a larger population to be supported (Hansen and Twaddle, 1988). Conflicts between kingdoms during this period were frequent, especially between the Bunyoro and Ganda. These conflicts were a significant driver of political change, leading to strategic alliances and territorial shifts (Twaddle, 1993; Low, 1971). By the mid-19th century, the Bunyoro's power had waned due to internal conflicts and the Ganda's expansion. The Ganda's success in consolidating territory and controlling resources meant it was the dominant force in the region by the time European explorers arrived (Low, 1971).

3.2.2 European Colonisation

By the mid-19th century, European explorers (such as John Hanning Speke) reached the Great Lakes region of Uganda. The mapping and exploration of the region paved the way for future colonial expansion by military force and through strategic partnerships (Hansen and Twaddle, 1988). In 1894, the British formally established the Ugandan 'Protectorate', employing an indirect rule system that empowered local leaders, specifically the Ganda elites, to manage administrative tasks. This drastically altered regional power dynamics, with the primary intention being to increase tax revenues for the British Empire. This system introduced western institutions and Christianity, as well as a number of cash crops, such as cotton and later coffee. These crops became the dominant part of the of the colonial economy, supported by forced labour, leading to increased food insecurity for the Ugandan population (Low, 1971; Heald, 1995). However, British reliance on Ganda elites to govern other regions fostered inter-ethnic tensions, notably with the Gisu, as these arrangements benefited the Ganda disproportionately (Mamdani, 1976). While the Ganda held a privileged position, British interference undermined traditional power structures at all levels. The British favoured some chiefs over others, often appointing compliant leaders and dismissing those who resisted colonial policies. This weakened traditional clan systems across all tribal and cultural groups, particularly in areas outside the Buganda sub-region, leading to social stratification, tension and conflict.

Christian missionaries were among the first Europeans in Uganda and were central to the establishment of schools (though education was largely limited to basic reading and writing). Consistent with the distribution of power during the colonial-era, the Ganda were the main beneficiaries of this education, leading to a relatively literate elite, while other regions and groups saw little benefit (Twaddle, 1993; Hansen and Twaddle, 1988). Missionary schools also introduced western religious beliefs and values, encouraging the spread of Christianity. Over time, Christianity gained a stronghold, particularly in the Buganda sub-region, altering Ugandan spiritual practices and cultural identity, leading to a shift in Ugandan cultural norms.

The colonial-era also saw the development of cities, such as Kampala and Jinja, increased urbanisation, and some modest investment in infrastructure (Mamdani, 1976). Cities became centres of trade and western influence, and with urbanisation came changes in clothing, diet, and social behaviour, marking a shift from rural and clan-based lifestyles. The British also introduced basic health care services. Malaria, sleeping sickness, and syphilis were major concerns, with colonial authorities carrying out mass immunisations and establishing basic health facilities in cities. However, the primary motivation was to ensure the continued healthy supply of cheap labour, and these services were largely limited to urban centres (Iliffe, 1998). Both the preferential treatment of the Ganda and limited development of rural regions had long lasting impacts on development outcomes and cultural identity.

3.2.3 Post-Colonial Uganda

Figure 3.4: Post-colonial Ugandan Leaders

Left: Milton Obote (1962 - 1971). Photo source: `https://cdn.britannica.com/68/91668-050-DAF8C1A0/Milton-Obote.jpg`. **Middle:** Idi Amin (1971 - 1978). Photo source: `https://apnews.com/article/uganda-politics-idi-amin-legacy-contested`. **Right:** Yoweri Museveni (1986 - present) Photo source: `https://edition.cnn.com/2024/07/22/africa/uganda-anti-corruption`. All accessed 11/11/2024.

Throughout the colonial period there were isolated instances of resistance, such as the Nyangire Rebellion in Bunyoro in 1907, a direct protest against the appointment of the the Ganda as overlords (Twaddle, 1993). However, it was not until the 1940s that more structured and well organised resistance appeared. Leaders such as Ignatius Musaazi, who had benefited from a Western education, united various ethnic groups in demanding greater representation and

independence, leading to widespread protests, strikes and the formation of nationalist political movements (Low, 1971). In the wake of World War II, shifts in geopolitical and economic power dynamics weakened the European colonial powers, making it increasingly difficult to maintain control over their colonies (Young, 2001). This was accompanied by the spread of ideas about self-determination, both domestically and within the colonies themselves. The formation of the United Nations also provided a visible platform for de-colonisation (Darwin, 1988).

After a succession of incremental concessions by the British throughout the 20th century, which included granting the Ganda autonomy, the London Conference of 1960 finally saw the British decide to permit an election on Uganda independence. On the 9th October 1962 Uganda gained its independence, with Milton Obote becoming the first Prime Minister, leading a mix of representatives from across Uganda, including Ganda separatists and religious leaders (Twaddle, 1993; Hansen and Twaddle, 1988). However, this period saw significant political instability, cultural division and accusations of corruption. In 1971, Major General Idi Amin staged a successful and self-interested coup, ousting Obote while he was attending the Commonwealth Summit in Signapore. Under Amin's dictatorship, the following eight years saw widespread human rights abuses, ethnic cleansing (with an estimated death toll between 100,000 and 500,000) and economic turmoil (Twaddle, 1993). In 1972, Amin expelled Uganda's Asian population leading to economic collapse and significant political isolation from the western powers. Amin's rule ended in 1979 after the Uganda-Tanzania War, following Amin's invasion of Tanzania a year earlier (Tindigarukayo, 1988).

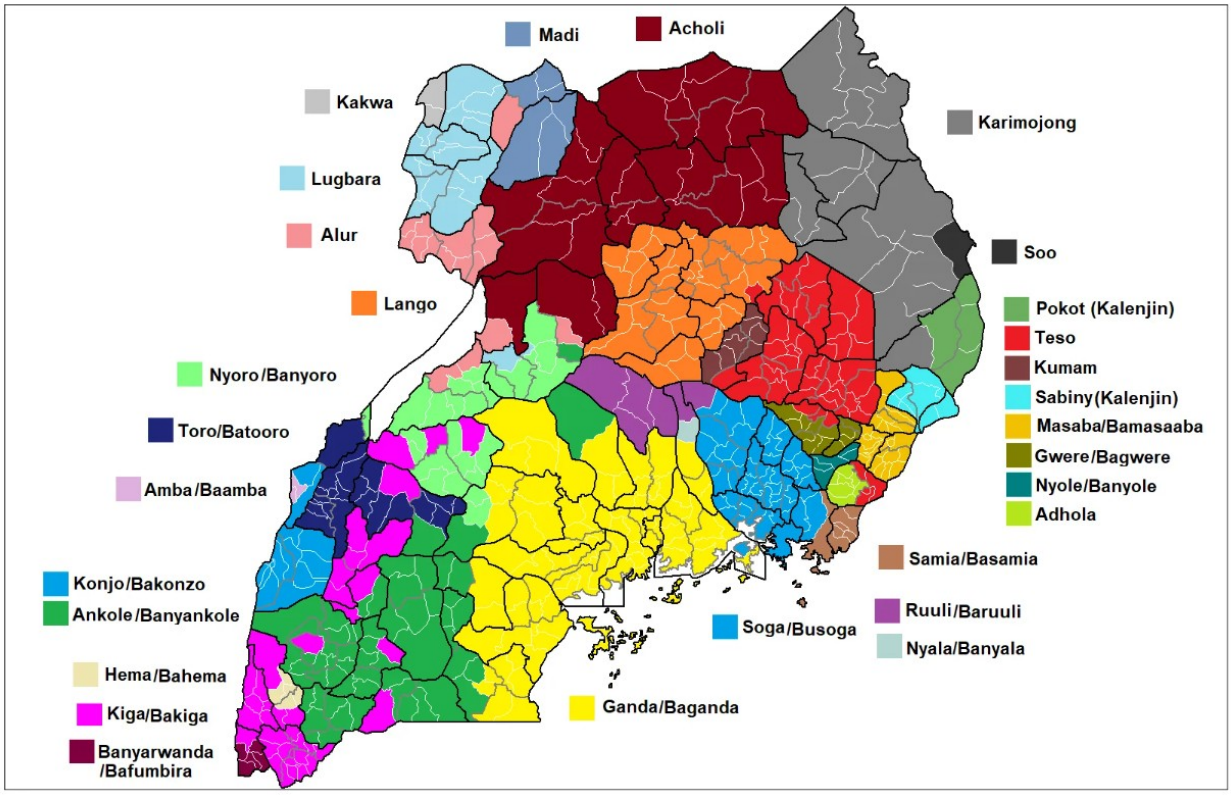
Following the invasion and capture of Kampala by the joint Tanzanian-Ugandan National Liberation Front (UNLF) force in 1979, Amin withdrew into exile. Uganda then faced years of further turmoil, with power struggles, including an attempted resurgence by the unpopular Obote. The Ugandan Bush War (1981-1986) saw Yoweri Museveni's National Resistance Army (NRA) eventually overthrow the government in 1986. Museveni's government introduced economic reforms, established a policy of decentralisation, and implemented policies to address public health crises like HIV/AIDS, which was a leading public health crisis by the late 1980s (Tripp, 2010). Under Museveni, Uganda has achieved relative stability and economic growth, with policies supporting infrastructure development, agriculture and education. However, Museveni's rule has not been without controversy. Accusations of electoral fraud and the suppression of political opposition have been consistent, as have concerns about human rights abuses (Golooba-Mutebi and Hickey, 2008).

3.3 Uganda's Cultural Diversity

There are 29 significant indigenous ethnic groups within Uganda (Afrobarometer, 2023; UBoS, 2024), which are summarised in figure 3.5. The Gisu (also known as the Masaba/Bamasaaba) are shown in orange-yellow in eastern Uganda. The Ganda are shown in yellow, in central-southern Uganda and the Acholi in burgundy in northern Uganda.

These groups are relevant, to varying degrees, to all three empirical chapters of this thesis, as summarised in table 3.2.

Figure 3.5: Ethnic Groups of Uganda 2024 Preliminary Census



Source: `https://www.ubos.org/wp-content/uploads/publications/National-Population-and-Housing-Census-2024-Preliminary-Report`, page 8, accessed 11/11/2024, based on 2024 preliminary census.

Table 3.2: Summary of Ethnic Groups in Empirical Chapters

	Ethnic Group	Languages Used
Chapter 4 - Language	Gisu	Lugisu / Luganda
Chapter 5 - Hypothetical Norms	Gisu	Lugisu / Luganda
Chapter 6 - Misperception	Gisu / Ganda / Acholi	Lugisu / Luganda / Luo

3.3.1 Languages of Uganda

Uganda's long past and geographical features predispose the area to relatively high linguistic diversity (Michalopoulos, 2012; Ashraf and Galor, 2013), and British colonial powers then cemented this diversity by designing Uganda's borders with relatively little attention to local cultural and geographic features (de la Cuesta and Wantchekon, 2016). As shown in figure 3.6, Uganda's 44 indigenous languages fall into four main families, of which Bantu and Nilotic are the largest (Eberhard and Fennig, 2022). Both Lugisu (Gisu) and Luganda (Ganda) are Bantoid languages, while Luo (Acholi) is

a Nilotic language. Figure 3.7 summarises the evolutionary history of each language, showing where each language branches off from any common ancestor. Luo is clearly distinct from both Lugisu and Luganda and does not share a common ancestor.

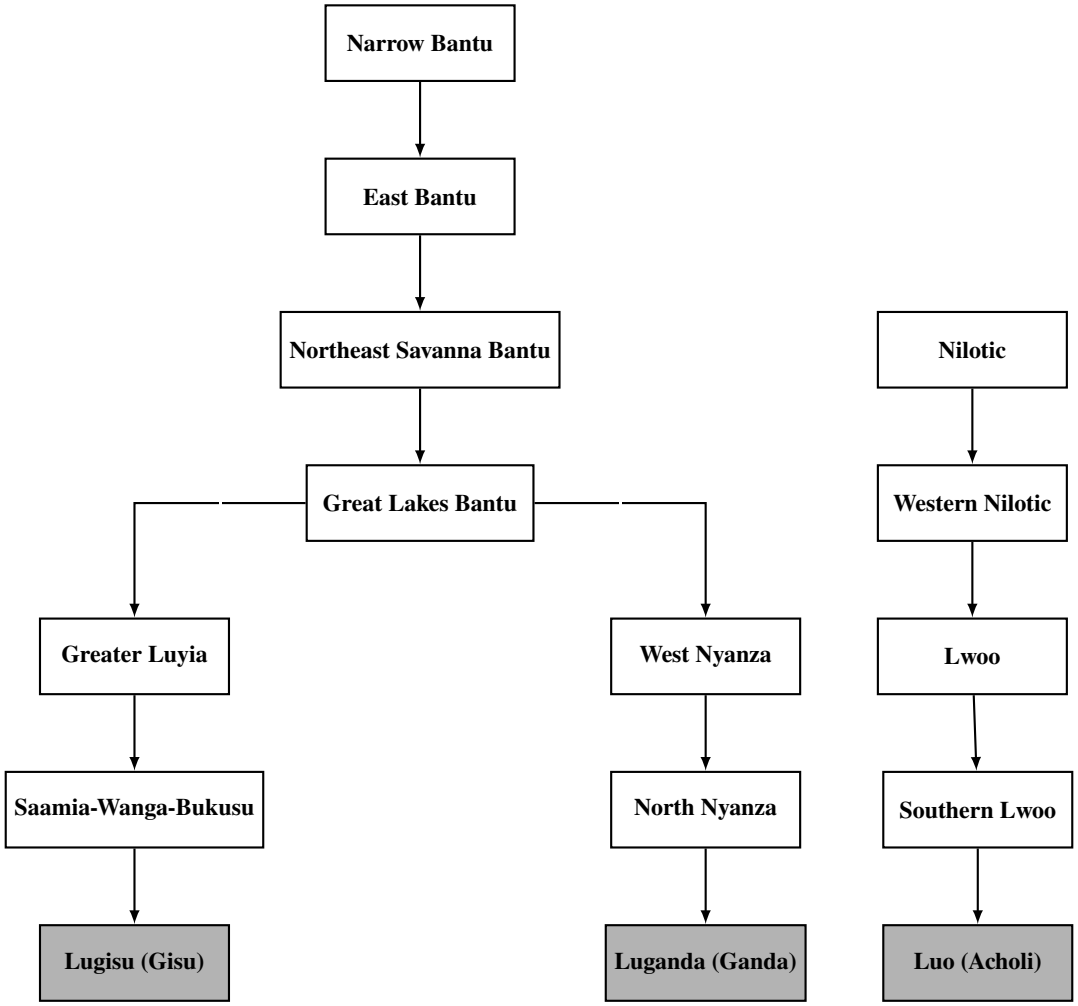
Figure 3.6: Language Families of Uganda

Source: https://en-academic.com/dic.nsf/enwiki/19555. Accessed 14/11/2024.

Ethno-linguistic diversity has been found to negatively correlate with income and a host of desirable outcomes, with one channel being lower levels of generalised trust (Alesina and La Ferrara, 2000, 2002; Alesina and Ferrara, 2005; Algan and Cahuc, 2013). Uganda has the kinds of linguistic diversity that has been found to be detrimental, on average. Desmet et al. (2012) find deeper distinctions (e.g. ELF 1, which focuses on differences at very high level of aggregations of language families) predicts civil war and the degree of redistribution, whereas shallow distinctions (e.g. ELF10-15, which includes distinctions between closely related languages) matters for literacy and infant mortality. Uganda is in the top quartile for both ends of the spectrum. Further, inequality between ethnic groups is abnormally high in Uganda. Hodler et al. (2020) point estimates rank Uganda as the second most ethnically stratified country

in their sample of 26 sub-Saharan African countries. They find this measure strongly correlates with stated trust in relatives, neighbours and other acquaintances.

Figure 3.7: Language Comparison: Lugisu, Luganda & Luo



Note: 'Narrow Bantu' stems from Atlantic-Congo, and all omitted preceding levels are shared. Source: Bantu: `https://glottolog.org/resource/languoid/id/masa1299`, Nilotic: `https://glottolog.org/resource/languoid/id/adho1241`, accessed 14/11/2024.

However, de la Cuesta and Wantchekon (2016) argue policy makers are not powerless, pointing to the roles of local interactions and national languages. Desmet et al. (2020) find that higher levels of interaction between groups can avoid the potentially detrimental effects of ethnic diversity. Likewise, national languages can be an engine of greater national cohesion. English has been the only official national language of Uganda since independence. Though Swahili was added in 2005, there has not been a significant uptake on the ground. Luganda is the most widely spoken language in Uganda, though closely linked to the Ganda and their historical position of power, followed by English and

Swahili[1]. Similar processes have been seen in Tanzania's adoption of Kiswahili (Miguel, 2004), or Sierra Leone's use of the Krio dialect of English, which Glennerster et al. (2013) argue helps avoid low trust equilibria. Whilst adopting a new national language may affect trust in the long run, people must make decisions today about the language they will use for a given interaction. Only around 15% of Ugandans are monolingual, and there is no universal lingua franca (Afrobarometer, 2023).

3.3.2 Specific Ethnographic Details: The Gisu, Ganda & Acholi

In this subsection I summarise the relative anthropological context of the Gisu, Ganda and Acholi. The Ethnographical Atlas (Murdock, 1967; Kirby et al., 2016) has been used extensively to gain an insight into the anthropological features of different ethnic groups. Table 3.3 presents some of the 45 aspects available for all three groups, focusing on the differences or key common features. When comparing to the average response to these questions for Uganda as a whole (across all Ethnographic Atlas questions), the Gisu, Ganda and Acholi respond consistently 50.6%, 43.8% and 40.3% of the time respectively. To clarify, for each of the 95 Afrobarometer questions, such as *EA023 Counsin Marriage permitted*, if (on average) both cultures answered "Duolateral cross-cousin marriage permitted", then the pairwise value is 1, or nil in any other combination. Thus, 50.6% means that 48 of the 95 questions have the same response. While crude, this is a measure of how typical each group is compared to the average Ugandan. When considering a direct comparison between groups, the Gisu/Ganda, Gisu/Acholi and Ganda/Acholi agree 47.7%, 23.9% and 32.9% of the time[2]. Thus, the Gisu and Ganda are most likely to agree with each other and the Gisu and Acholi are the least.

[1]Using Afrobarometer (2023) Luganda is a common lingua franca for Bantu-speaking Ugandans, but English is more common for cross-tribe communication in the north.

[2]Pairwise analysis conducted on data taken from `https://github.com/D-PLACE/dplace-data/tree/master/datasets/EA` on 05/08/22. N=533 for African societies, filtered by Glottocode to isolate 29 Ugandan cultural groups.

Table 3.3: Relevant Historical, Cultural and Geographic Features

Feature	Ganda	Gisu	Acholi
Ethnographic Atlas, *Highlighting Differences Between the Tribes*			
Settlement patterns EA030	Villages/Towns	Dispersed Homesteads	Hamlets
Reliance on agriculture EA005	70.5%	60.5%	50.5%
J. hierarchy of local community EA032	Independent Families	Extended Families	Extended Families
J. hierarchy beyond local community EA033	Pre-Colonial State	Petty Chiefdom	Petty Chiefdom
Organisation of clan communities EA016	No exogamous clans	Clans	No exogamous clans
Largest patrilineal exogamous group EA018	Sibs	Lineages	Sibs
Descent: major type EA043	Patrilineal	Patrilineal	Patrilineal
Marital residence custom EA010	Neolocal	Patrilocal	Uxorilocal
Community marriage organisation EA015	Exogamous	Clans	Agamous
Domestic organisation EA008	Polygyny	Polygyny	Polygyny
Prohibits cousin marriage EA023	1st/2nd	1st	1st
Male genital mutilations EA037	Absent	Early Adulthood	Absent

Note: The Ethnographic Atlas data is from Murdock (1967) via Kirby et al. (2016), see also Bahrami-Rad et al. (2021). Relevant questions referenced in grey.

The following sections now provide some more detailed anthropological context of each of the three groups, with a specific focus on the Gisu, given that they are the focal group through this thesis.

The Gisu

"The Gisu, an agricultural people living on the slopes of Mount Elgon in eastern Uganda, have a long-standing reputation for quarrelsome and unruly behaviour. Resort to violence in personal disputes is common among men and women. Conjugal disputes, for example, frequently involve physical injury to either or both spouses and the destruction of property. Beer parties are often the scenes of brawls, some of which may result in death."

La Fontaine (1967), p.249

The Gisu people are primarily located in the eastern region of Uganda around Mount Elgon, a particularly rugged terrain. Prior to the Kenyan/Ugandan border being established in 1904, the Gisu interacted closely with other Kenyan tribes. Gisu society is organised into clans, which are critical for identity and social organisation. Each clan has specific roles and responsibilities, serving as a network of support and resource sharing. Kinship ties are central to Gisu life, with lineage typically traced through the male line. Heald (1995) notes that "clans are the building blocks of Gisu identity" (Heald 1995, p.45). Gender dynamics in Gisu culture are distinct and deeply rooted; men traditionally take on roles as providers and decision-makers, while women are primarily responsible for household management and agricultural production. "Women's roles, while often seen as subordinate, are critical in maintaining the social

fabric" (Heald 1995, p.72). More generally, the Gisu "visualise their society in terms of a series of progressively more inclusive patrilineal descent groups, from the smallest lineage segment to the category of all the descendants of Masaba" (Heald, 1995 p.140).

Lugisu is key component of Gisu identity and communication within the community. Lugisu is characterised by its rich phonetic and grammatical structure, including noun class systems that categorise words based on prefixes. This complexity reflects the nuances of Gisu cultural concepts and relationships. "Language is not merely a tool for communication; it embodies the cultural essence of the Gisu" (Heald, 1995, p.154). Lugisu is rich in idiomatic expressions, proverbs, and oral traditions. Storytelling, an important cultural practice, relies heavily on the Gisu language to convey moral lessons and historical narratives. Language plays a significant role in education within Gisu communities. Local schools often incorporate Lugisu (in addition to English and Swahili) fostering a sense of identity among the youth. Heald discusses how the Gisu maintain a strong sense of identity through their cultural practices, despite external influences and modernisation. She writes, "the preservation of language and customs is paramount for the Gisu, serving as a bulwark against cultural erosion" (Heald 1995, p.150). The vast majority of Gisu are multilingual, typically speaking Luganda and English fluently in addition to Lugisu.

The Gisu are predominantly agrarian, with farming forming the basis of the economy. Key crops include bananas (matoke), a staple food integral to Gisu diets, coffee and Maize, which are important cash crops that contribute to economic sustainability and trade. Heald emphasises the significance of communal labour in agricultural practices, stating, "the cooperative nature of farming fosters a sense of unity among community members" (Heald 1995, p.88). Heald also notes that in agriculture children are perceived to be of limited value, while women are seen as equals to men. Men typically control land ownership with land largely passing from father to the oldest son, either after 'Imbalu' or at death.

One of the most notable aspects of Gisu culture is the initiation ceremony known as 'Imbalu' which marks the transition to adulthood for boys. Heald describes this rite as "a crucial moment in the life of a Gisu male, embodying both physical and social transformation" (Heald 1995, p.112). This is considered a public act of bravery and occurs between the ages of 18 and 24. During this ceremony boys are expected to show "total fortitude under the knife, betraying no signs of fear, and even what might be regarded as involuntary twitches and tremblings, such as the blinking of the eyes, are evaluated negatively" (Heald 1995, p.60). The practice is designed to produce fierce men, where boys are able to freely manifest the powers to which their ancestry makes them heir. It's perhaps not surprising the the Gisu have a reputation for violence: "Ganda just steal but Gisu come with knives to kill you" (Heald 1995, p.30). During colonial times, where the Ganda acted as overlords, the Gisu formed vigilante groups to resist tax collectors and British rule more generally. However Gisu violence pre-dates colonialism; in Gisu culture killing is a matter of individual self-help and is seen as a male prerogative, specifically against thieves or witches. Following Imbalu, violence and conflict between fathers and sons is common, specifically around inheritance rights which are

conferred as part of the ceremony (Hargreaves Heap et al., 2012). Here, as elsewhere in Gisu life, the onus is upon the individual, and "kinship has little moderating effect on the attitude adopted" (Heald 1995, p.79). These combined factors lead to a striking lack of in-group preference amongst the Gisu, where an out-group preference is not uncommon (Hargreaves Heap et al., 2012), even with the Ganda. In Gisu culture, "there is a wide difference in Gisu attitudes to the killing of men and the killing of women and children" (Heald 1995, p.43). Indeed women are considered to be frail but furtive. A women can take her husband to court if he beats her too severely and 45% of all Gisu marriages end in divorce. Bridewealth (a form of dowry) is expected to be returned on divorce. Women are never committed to one household, and associate or marry Gisu men for their livelihoods and status. Thus, a Gisu man can only keep his wife by providing significant wealth or an estate, with the wealthiest Gisu men being polygamous.

Two other notable features of Gisu culture are 'Bukulo' and the believe in witchcraft. Bukulo, or the joking relationship, is common in Gisu culture, where reciprocal abuse and theft is seen a form of friendship which can form in response to a feud which has de-escalated peaceably. "This relationship is developed by the reciprocal 'snatching' - known by the special term, *xutubuta* - of each other's property. These thefts begin with small items, such as pots or chickens, and progress to goods of higher value, with the aim being finally to snatch cows from each other" (Heald 1995, p.164). While xutubuta is special form of theft, more widely in Gisu culture "there is no substantive difference between theft and witchcraft and how these two types of offences manifest themselves in Gisu life and ideology" (Heald 1995, p.6). Belief in witchcraft is widespread, and is still the primary reason for murder in the Bugisu sub-region. There is a belief that supernatural powers form with the wisdom of age and (because of this) there is a degree of fear between the young and the old.

The Ganda

The Ganda are the largest ethnic group in Uganda, predominantly residing in the central region around Lake Victoria. The Ganda have had a significant influence on Uganda's history, politics, and social structure, as summarised in section 3.2. The Ganda speak Luganda, a Bantu language that serves as a lingua franca in southern Uganda. Language plays a vital role in their identity and cultural expression. Traditional songs and stories in Luganda reflect their values and historical narratives (Twaddle, 1993). The Ganda are traditionally agriculturalists, with the main crop being matoke (a form of banana), supported by other crops, such as coffee, maize, and beans. The Buganda sub-region is typified by fertile soils and a favourable climate, partially explaning why the Ganda are the largest cultrual group in Uganda (Hansen and Twaddle, 1988). The Ganda are highly centralised and their 'Kabaka' (king or leader), has historically served as both the political and cultural leader. The kingdom's administration is organised hierarchically, with clan chiefs overseeing counties, sub-counties, and villages (Fallers, 1956). The Kabaka has considerable influence over political and social matters in Buganda, often working alongside a council of chiefs. While these traditional structures are still culturally important, increased centralisation has somewhat eroded their direct impact on Ganda life. However,

clan membership and identity is still fundamental to the Ganda.

Prior to the introduction of Christianity and Islam, the Ganda practised a form of ancestor worship and animism, believing in a pantheon of spirits ('Balubaale') that inhabit the natural world (Kiwanuka, 1971). Temples and shrines across the Buganda sub-region still continue to hold spiritual significance, and rituals are performed during key annual agricultural activities. With the arrival of missionaries in the 19th century, Christianity became widespread, and today, the majority of the Ganda identify as Christians (specifically Anglicanism or Catholicism). Despite this, traditional beliefs still influence cultural practices, especially in rural areas (Ray, 1991; UBoS, 2024).

Gender roles among the Ganda are rooted in the established patriarchal society, though these have seen changes in the face of modernisation. Men traditionally occupy positions of authority both in the household and in broader societal structures. Leadership roles are typically given to men who are responsible for all key decisions, land ownership and providing for the family (Fallers, 1956; Kiwanuka, 1971). Women, on the other hand, have historically been tasked with domestic duties, including child-rearing, food preparation, and maintaining the household. They play a crucial role in agriculture and, despite these traditional roles, women have significant influence over social and cultural life. Due to rapid urbanisation, increased female education and reduced heritable land following population growth, Ganda women are increasingly participating in professional careers and political life, increasing reshaping traditional gender dynamics (Wrigley, 1996).

The Acholi

The Acholi are a Luo-speaking ethnic group and originate from the northern region of Uganda. The major urban regions in this areas are Gulu and Kitgum. Traditionally the Acholi have been agro-pastoral, centered around millet and maize, alongside livestock rearing, which serves both economic and cultural purposes (Girling, 1960). Acholi society is structured around patrilineal clans, each led by a chief, who is responsible for resolving disputes and managing land allocation. The chief also performs ritual duties, such as mediating with ancestral spirits and maintaining harmony between the spiritual and physical worlds (Atkinson, 1994). Thus, in Acholi life, there is a unity between the social and spiritual. The pre-colonial Acholi belief system is centred around a supreme creator 'Rubanga' and ancestoral spirits. Rituals and sacrifices play a critical role in resolving misfortunes or illnesses, which are often attributed to spiritual disharmony (Girling, 1960). The colonial period and missionary efforts introduced Christianity, and today many Acholi identify as Christians while retaining elements of their traditional spirituality, particularly in rural areas (Atkinson, 1994; Behrend, 2000).

Historically the Acholi have endured significant upheavals, particularly during the Lord's Resistance Army (LRA) insurgency, led by Joseph Kony against Museveni's government, which began in the 1980s. This conflict

caused widespread murder and displacement of the Acholi, as well as disrupting traditional practices and leaving deep psychological scars on the community (Finnström, 2006; Branch, 2011). Many Acholi, up to 90%, were forced into internally displaced persons (IDP) camps, where their traditional livelihoods and social structures were undermined. Despite these challenges, efforts to rebuild Acholi society are on-going, highlighting the group's resilience (Allen, 2006). Women have taken on prominent roles in societal reconstruction, challenging traditional gender norms while continuing to uphold vital aspects of Acholi heritage. Acholi traditional gender roles have been firmly grounded in the patriarchy. Additionally, cultural festivals re-introducing traditional music, dance, and language are central to the Acholi's efforts to rebuild their identity (Branch, 2011). Acholi culture is renowned for its vibrant music and dance traditions, which are integral to social ceremonies and cultural identity. Notable among these is the 'bwola' dance, traditionally performed during royal occasions and other significant events. These cultural expressions not only serve as entertainment but also convey historical narratives and social values (Girling, 1960).

3.4 Conclusion

In conclusion, this chapter has summarised the cultural and historical intricacies of Uganda, focusing on the Gisu, Ganda, and Acholi ethnic groups. I have highlighted how Uganda's long history has contributed to a rich and substantial cultural endowment, social structures and linguistic diversity, all of which have shaped contemporary Ugandan society. I have presented further detail on three main ethnic groups: the Gisu, Ganda and Acholi, all of which are the focus of this broader work. Uganda is typical of many developing nations, specifically in terms of it's high degree of social fragmentation and complex history. Balancing nation-building and fostering social cohesion (while maintaining individual cultural identities) is challenging, and necessitates well-informed and effective policies.

Chapter 4

Chapter 4 - Bilinguals in the Lab: Why do Norms and Expectations not Predict Contributions?

4.1 Introduction

Economists have systematically measured cross-country differences in preferences for many years, with early attempts comprising just four nations (Roth et al., 1991). Amongst recent attempts, perhaps the most ambitious is Falk et al. (2018), who measure a range of economic preferences for 80,000 individuals in 76 countries. With such a variety of countries, trying to understand where preferences 'come from' has moved from finding individual-level correlates to country-level correlates. Language has recently emerged as an important factor: does the language you speak influence the decisions you make? This question has attracted attention from a wide variety of disciplines (De Saussure, 1916; Wittgenstein, 1922; Whorf, 1956; Rubinstein, 2000), with evidence that the structure of a language influences *perception* of colour, time and morality (Costa et al., 2014; Boroditsky, 2019). On *behaviour*, there is observational evidence that language affects savings, risky behaviour and female labour force participation (Chen, 2013; Jakiela and Ozier, 2018; Galor et al., 2018; Herz et al., 2021; Campo et al., 2024).

Some doubt this observational evidence. Linguists are sceptical of so-called Whorfian influences, where language structure affects thought (Pinker, 1994). In addition, there are potentially statistical problems with some common approaches (Roberts et al., 2015; Pepinsky, 2022). There are large numbers of grammatical features and outcome measures, and so a statistically significFant correlation could easily be found, even if there is no true underlying relationship. Further, languages are not entirely separate, which can effectively lead researchers to be too confident in their results (due to autocorrelation/the Galton problem). If researchers have an incentive to find significant results and/or these are more likely to be published, such findings could be misleading (Brodeur et al., 2020).

An alternative approach is to use experimental methods.[1] Bilingual subjects speak at least two languages fluently, and are thought to be in the majority globally (Grosjean, 2021). Experimenters can use bilingual subjects and randomly assign subjects to participate in an experiment in one of their languages. Previous research has found significant effects in Morocco (French/Arabic, Lambarraa and Riener, 2015), Hong Kong (Chinese/English, Li, 2017), Uganda (Lugisu/Luganda, Clist and Verschoor, 2017), Estonia (Estonian/Russian, Pérez and Tavits, 2017) and the UK (Chinese/English, Clist and Hong, 2023). Whilst the experimental approach does provide experimental control, it does not address the issue of publication bias. These papers could be unrepresentative precisely because they find interesting and significant results.

I replicate the largest experimental effect size previously found. Clist and Verschoor (2017) report randomising the language of a public goods game in rural Uganda leads to a third of a standard deviation increase in cooperation[2] (though importantly this study did not test the underlying mechanisms). I use university students in

[1] On occasion experimental methods are used to measure preferences between different groups, as per Sutter et al. (2018), in situations where the two groups differ in the mother tongue as well as, potentially, other factors.

[2] This increase occurs when participants speak Luganda (a semi-national language) instead of the baseline Lugisu (the tribal language). Participants on average were found to be less cooperative when speaking the tribal language compared to the semi-national language.

Kampala rather than a rural farming community near Mbale. This urban setting allows for a test of how robust and widespread language-level differences in cooperation are. I can rule out a Whorfian explanation, as the two languages used do not differ in any relevant grammatical feature.[3] I can also rule out more run-of-the-mill confounds such as differential language proficiency due to the setting.

I run a standard one-shot two-player public goods game in Uganda's capital city Kampala, randomising the language of sessions between Luganda (the national language) and Lugisu (the relevant tribal language). The 349 subjects are students from the Lugisu-speaking Gisu tribe's homelands, and speak both languages well. After measuring contributions, I also consider the underlying mechanisms. Bicchieri (2010) is typical in distinguishing between normative beliefs (the shared view of what is socially acceptable) and empirical expectations (the belief about how others actually behave). Ordinarily, empirical expectations are linked to a normative belief, but this does not necessarily have to be the case. In this terminology, I use *norms* to refer to normative beliefs and *expectations* to empirical expectations. Krupka and Weber (2013) introduced the four-point norms elicitation method, showing the importance of norms in simple dictator games. The method is a pure coordination game (Mehta et al., 1994), thus making it incentive compatible to reveal shared norms. This approach has since been widely adopted and/or adapted, and is central to this chapter (Kimbrough and Vostroknutov, 2016; Gächter et al., 2017; Barr et al., 2018; Chang et al., 2019)). Thus, after contribution decisions, I elicit norms and expectations.

My pre-registered design allows me to test four possible theories for any differences. First, language could activate different norms, which norm-following subjects then abide by (Benjamin et al., 2010; Binmore, 2010; Kimbrough and Vostroknutov, 2016). Second, if subjects are conditional cooperators, language could act as a coordination device; a high-cooperation language could lead subjects to choose a high-cooperation equilibrium (Fehr and Schurtenberger, 2018). Third, injunctive norms and empirical expectations could jointly underpin established norms, each being necessary conditions for differences in behaviour (Bicchieri, 2010). Fourth, it could be that rather than changing what subjects feel they should do, or what they think others will do, it could simply change what they want to do (see Gächter et al., 2013, for a relevant discussion). Language could do this by making one's identity more salient (Akerlof and Kranton, 2000), or by having a different resonance in one's associative memory (Bordalo et al., 2017).

The anthropological literature implies Lugisu and Luganda have different cultural associations. Various features in the Ethnographic Atlas differ between the Gisu (who traditionally speak Lugisu) and the Ganda (who traditionally speak Luganda), several of which have been found to correlate with modern-day development outcomes. In almost every case, there are reasons to think Gisu will be poorer, have lower generalised trust and have a higher propensity for conflict. Experimental measures find the Gisu have slight out-group bias and remarkably low levels of

[3] I am aware of two proposed Whorfian influences on trust and cooperation. Tabellini (2008) argues that the grammatical features of 'pronoun drop' and differentiating second person pronouns are related to levels of trust and respect. However, both Lugisu and Luganda allow pronoun drop, and neither have politeness distinctions. Therefore, any differences by language cannot be explained by these grammatical features.

trust with strangers (Hargreaves Heap et al., 2012), invest less when friends could cover their losses (D'Exelle and Verschoor, 2015), and have low cooperation with spouses (Iversen et al., 2011). The cultural value of independence and autonomy amongst the Gisu is perhaps their defining feature, as their independence "systematically downplays the dependence of people upon each other" (Heald, 1989, p.77). Luganda, by contrast, is associated with more of a national identity, as well as the historically powerful Ganda (see chapter 3 for further context).

I have five results. First, I show that subjects contribute 28.9% (1/3 SDs) more in the national language (Luganda) than in the local language (Lugisu). This effect size is within 0.1% of the original (Clist and Verschoor, 2017) study, who report a difference of 28.8%. My experiment changes several elements of the sample, location, year, and design, and so the virtually identical effect size is remarkable. Simple bayesian methods show the value of the replication (see appendix 2 for more, also Verhagen and Wagenmakers, 2014), and allow me to have greater confidence that there are genuine non-Whorfian language effects on behaviour. Second, injunctive norms do not appear to be a compelling mechanism, as they are almost identical in the two languages. Third, empirical expectations also do not appear to be a credible mechanism as they differ only slightly, and in the wrong direction. Subjects appear to exhibit wishful thinking in general, and mistakenly expect higher contributions in Lugisu. Fourth, I find that there is no established norm in either language, as empirical expectations and injunctive norms do not agree. Fifth, the average subject in my experiment is neither a conditional cooperator nor a norm follower. This may be because of the lack of established norms, or because of the subject pool is comprised of a culture that downplays cooperation. To summarise, I find language has a large and predictable effect on cooperative behaviour in Uganda. Having ruled out social norms as the mechanism, and more run-of-the-mill confounds like comprehension, I conclude language affects social preferences.

The chapter proceeds as follows. Section 2 presents the research design, section 3 the model and empirical considerations, section 4 the results and section 5 concludes.

4.2 Research Design

4.2.1 Context: The Gisu (Lugisu) & Ganda (Luganda)

This experiment uses two languages and Gisu participants. Lugisu and Luganda are, respectively, the languages traditionally spoken by the Gisu and Ganda tribes, who currently make up around 5% and 16.5% of the country's population (UBoS, 2016). As people move and marry, they learn multiple languages and even eschew 'their' tribal language, but cultural associations still persist. This study uses subjects from the Gisu tribe that have chosen to live and study about 120 miles away in Kampala, Uganda's capital city, as well as the historical base of the Ganda tribe's

powerful kings. This section provides a high level summary of the differences between the Gisu and Ganda as, from the perspective of the Gisu, each language will likely have a unique set of specific associations.

As set out in chapter 3 (specifically table 3.3), the Ganda have typically had a developmental advantage compared to the Gisu across a range of metrics, with the Gisu having lower levels of trust and tighter kinship scores (relying more on disgust, shame and revenge taking to maintain cooperation (Enke, 2019)). This is save for one noticeable advantage: the Gisu homelands are more rugged, and so more protected from the slave trade (Nunn and Puga, 2012). The higher ruggedness of the Gisu homelands imply slave exports would be lower by around one standard deviation (3.315), which in turn implies real GDP per person would be lower by $2,343.[4] Slavery did indeed affect the Ganda more than the Gisu, which implies one area in which the Gisu are relatively fortunate.

Differences between the two tribes were, if anything, amplified by their colonial experience. The task of subduing eastern Uganda, including the Gisu, was delegated to the Ganda by the British. In a sense this created *some* common feeling amongst the Gisu, as La Fontaine (1969, p.186) explains "...the fifty years of British administration in Uganda introduced into Gisu tribal life common interests and aims which had not existed formerly." Yet the sense of a cohesive broader tribe is elusive. Alongside a famous circumcision ritual La Fontaine (1969, p.188) argues the Gisu language is the "commonest [cultural] symbol, for it represents in its grammar and vocabulary a paradigm of a common way of life and enshrines common values." The Ganda, by contrast, remained a cohesive unit with some political power, and a strong cultural identity most visible in the Ganda kings (Kabaka), who still exist today albeit in a much curtailed fashion. The consequences of this history can be seen in later anthropologies. The Gisu had very high levels of homicide in the years after independence, with pressures on land and inheritance meaning fratricide was common. Across the border in Kenya they were referred to as cannibals, whereas in Uganda merely seen as 'personally aggressive' (Heald, 1989, p.30). Richards (1952, p.116) summarises the Ugandan view with the saying "Ganda just steal but [Gisu] come with knives to kill you." In more recent years, these pressures can be seen less in the crime rate, and more in cultural values. Individual autonomy is praised, and "...seen as basic, carrying with it the right to act in one's own interest and one's own defence" (Heald, 1989, p.8).

Available survey and experimental data paint similar pictures, but should be treated with some caution. I am interested in whether the language of an interaction changes one's decisions, and so the specific language used in an experiment or survey could matter. The language used is not always explicitly stated for experiments, and tribe-level sample sizes can be low in surveys. With these caveats in mind, survey data is summarised in table 4.1, broken down by stated ethnic identity. There is broad national context of very low generalised trust, for both tribes. Though interestingly, the Gisu have higher stated trust in all groups, variously significant at levels 1-10%. And yet, interviewers perceive Gisu respondents as significantly less cooperative and honest. It's not possible to determine

[4]This follows the calculation in Nunn and Puga (2012, p.30). The slavery calculation is $1.326 * (3.8 - 1.3)$. The GDP calculation is $e(ln(1784) - .206 * (-3.8 * 1.279)) - e(ln(1784) - .206 * (-1.3 * 1.279))$. Units are PPP-adjusted 2006 international dollars.

Table 4.1: Afrobarometer's Survey Data on Trust and Attitude, by Tribe

	Ganda		Gisu		T Test	
Question	Mean	N	Mean	N	b	se
How much do you trust each of the following types of people?						
Other Ugandans?	1.35	543	1.51	168	0.16**	0.0766
Your relatives?	1.82	545	2.10	168	0.28***	0.0793
Your neighbours?	1.46	545	1.78	168	0.32***	0.0730
Other people you know?	1.30	545	1.53	167	0.22***	0.0717
People from other religions?	1.36	545	1.48	168	0.13*	0.0718
People from other ethnic groups?	1.10	545	1.32	168	0.21***	0.0763
What was the respondent's attitude toward you during the interview?						
Friendly /Hostile	1.17	547	1.19	168	0.017	0.0338
Interested/ Bored	1.22	547	1.27	168	0.047	0.0381
Cooperative/ Uncooperative	1.21	547	1.35	168	0.13***	0.0398
Patient/ Impatient	1.24	547	1.24	168	0.0064	0.0441
At ease/ Suspicious	1.37	547	1.45	168	0.083	0.0532
Honest/ Misleading	1.27	547	1.49	168	0.22***	0.0453

Note: The trust questions range from 0 (Not at all) to 3 (A lot). The attitude question is asked to the interviewer, and ranges from 1 to 3, with higher numbers indicating more hostile participants, for example. The *b* is simply the difference from a t test, with *** denoting significance at 1%, ** at 5% and * at 10%. The Afrobarometer (2023) data is from round 9, which covers January 2022. Tribe is self-identified.

whether this is more to do with the interviewer's prejudice or the respondents actual behaviour. However, when Gisu respondents are interviewed by Gisu enumerators, they report significantly lower cooperativeness and honesty than non-Gisu enumerators. Gisu respondents are seen as significantly less cooperative and honest by Gisu enumerators, which drives the effect that the Gisu in general are seen as less cooperative and honest.

I can also summarise experimental findings from the area. With strangers, Hargreaves Heap et al. (2012) find Gisu men have some of the lowest giving & return rates in trust games found anywhere in the world (both around 25-35%). Further, they find evidence of slight out-group bias in trust, and no evidence of reciprocity. Amongst friends, D'Exelle and Verschoor (2015) find that the Gisu invest *less* when others can cover their losses. They appear to avoid incurring social debts or obligations, which closely echoes the characterisation that for the Gisu "...the pattern is of individual self-help, a pattern which is predicated upon the individual responsibility of each man for his own actions (and enmities) and which in turn leaves him standing alone" (Heald, 1989, p.151). Even amongst couples, there is evidence of low cooperation which is inefficient from a narrow economic standpoint. Iversen et al. (2011) ask mostly Gisu couples from the local area to play a two-player public good games with their spouse. Even with a multiplier of 1.5, most subjects do not contribute the full amount.

To summarise the relevant contextual material, Uganda has relatively high levels of ethnolinguistic diversity, low levels of trust/cooperation, and relatively low development outcomes. Such features are often found to correlate with one another, with language potentially a relevant factor in causing and overcoming the development challenge. There

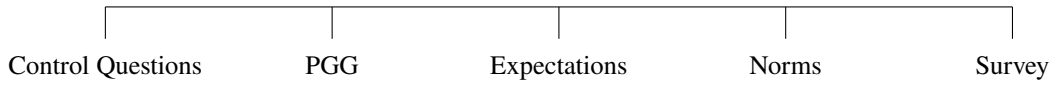
are reasons to think the two languages may hold different associations, values or cultural frames. From precolonial times, it seems the Ganda were richer, more politically organised, and had looser kinship networks than the Gisu. In almost every case that a feature has been found to correlate with later progress, the Gisu appear either equal with the Ganda, or to be hampered. Whilst these are general tendencies more recent data from anthropologies, surveys, and experiments all largely concur. And yet, the deep-rooted differences can apparently be overcome surprisingly easily. Clist and Verschoor (2017) find that people from the Gisu homelands are much more cooperative in Luganda than Lugisu. If language can circumvent a low-trust equilibrium, this offers a source of more immediate hope than longer term plans to instil a national language or increase contact between tribes.

4.2.2 Experimental Design

Location, sample and pre-screening

Figure 4.1 summarises the elements of the experiment, discussed in turn below:

Figure 4.1: Experimental Design



The experiment was conducted by the Field Lab[5] in two universities in Kampala. I restricted the sample to the Gisu, or people from the Bugisu region. As the Field Lab is based in Mbale, the main city in the Bugisu region, this interest should feel natural to participants. Local informants suggest that if a subject in Kampala speaks Lugisu they will also be fluent in Luganda, which was confirmed during the pilot, so all recruitment activities was undertaken in Lugisu. To recruit ethnically Gisu subjects (or subjects who have a strong link to the area) I used the semi-formal tribal social groups at each university. These groups maintain membership lists which represent the most accurate data of the local ethnic Gisu population. Willing subjects registered in advance, were pre-screened for their fluency in both languages and then randomly allocated to a session.

I randomised the language of the experiment at the session level, with 14 sessions in total, and 7 in each language. The allocated language was then used in *all* communication of that session, be it between facilitators, or with participants. Local informants suggest it is normal to hear both languages in informal and formal settings, so this too should feel natural. My sample of 360 comprises 200 over 8 sessions from Makerere University and 160 over 6 sessions from Kyambogo University. A constant set of facilitators delivered a script that was carefully translated and backtranslated to ensure consistency. All facilitator roles were maintained across both session languages to ensure

[5]A local research organisation with deep expertise in behavioural experiments: `https://thefieldlabuganda.com`.

comparability, though all are proficient in both languages. A full English script, answer sheet and survey questions are included in appendix 1.

In terms of how related Luganda and Lugisu are, language trees help objectively assess their linguistic distance; they are separated by 3 branches, sharing a common ancestor at level 7.[6] For intuition, this is the same distance as between English and German, or between French and Spanish.

Public goods game (PGG) and controls

I employ a standard linear one-shot two player PGG, which captures each participant's preference for cooperation. In this scenario, participants must choose between maximising their own payoff or the social benefit. Both individuals in the pair will be endowed with 12,000 Ugandan Shillings ('UGX'). At the time this purchased three hot meals on campus, and was equivalent to around £2.70 or $3.25[7]. The experiment lasted around 1 hour. Player 1's payoff (V_1) is calculated as:

$$V_1 = 12,000 - C_1 + 0.75(C_1 + C_2) \tag{4.1}$$

Where C_1 is their contribution to the joint fund and C_2 is the contribution from player 2. Each player can earn between 9,000 and 21,000 Ugandan Shillings, with the minimum and maximum pay out per pair totalling 24,000 and 36,000 Ugandan Shillings respectively.

In order to ensure participants fully understood instructions, the three corner solutions were used as the basis for control questions (see table 4.2). A basket represents the group fund ($C_1 + C_2$) and envelopes represent the amount of the initial endowment retained (12,000 — C_1). As per appendix 1, each scenario has a number of control questions, which also help subjects to think through the logic of the experiment. Based on piloting, understanding was enhanced where one control question was jointly solved by the room with open discussion. Eleven subjects (of 360, i.e. 3%) answered more than 10% of the individually answered control question steps incorrectly, and are excluded from any analysis.

[6]From `https://glottolog.org/resource/languoid/id/masa1299`. They share levels 1 to 9: Atlantic-Congo, Volta-Congo and Benue-Congo, Bantoid, Southern Bantoid, Narrow Bantu, East Bantu, Northeast Savanna Bantu and Great Lakes Bantu. They then diverge. Lugisu (also known as Masaaba) is under Greater Luyia, Luyia, Saamia-Wanga-Bukusu and Bukusuic. Luganda is under West Nyanza and North Nyanza.

[7]This level of compensation is high when compared to other recent Public Goods Games conducted in Uganda. Agarwal et al. (2021) paid 399 community workers UGX 31,240 per day. For experiments ranging between 2 - 4 hours, Gatiso et al. (2018) paid between UGX 5,000 and 20,000 across 450 participants in 34 villages, Sseruyange and Bulte (2018) between UGX 5,000 and 20,000 across a range of Ugandan workers and Clist and Verschoor (2017) between UGX 6,000 and 14,000 for a similar experiment in Eastern Uganda. Therefore, these incentives are sufficient to ensure serious and considered responses.

Table 4.2: Control Questions

Scenario	Player 1's Choice	Player 2's Choice	Final Pay-offs
1	Envelope: 12,000 Basket: nil	Envelope: 12,000 Basket: nil	Player 1: 12,000 Player 2: 12,000
2	Envelope: nil Basket: 12,000	Envelope: nil Basket: 12,000	Player 1: 18,000 Player 2: 18,000
3	Envelope: 12,000 Basket: nil	Envelope: nil Basket: 12,000	Player 1: 21,000 Player 2: 9,000

On completion of the control questions, subjects made their contribution decision: splitting 12,000 between their private envelope and a common basket (in an increment of 3,000). Payment was made to participants at the end of the experiment.

Expectations

The next section of each experimental session measured expectations of contributions. Participants were able to earn a bonus 4,000 Ugandan Shillings for whichever question in the norms and expectations section was randomly selected. For the expectation question, subjects are asked to guess the percentage of people in their session that gave the various amounts (UGX 0 to UGX 12,000 in UGX 3,000 increments) to the common basket. Subjects earnt the bonus if that question was selected and their guess was within 10% of the actual answer.

Norms

The next section of the session measured norms. Participants were asked to rank how socially acceptable each of the contributions levels were, using the now standard Krupka and Weber (2013) method. I employ a four-point scale; very socially unacceptable (- -), somewhat socially unacceptable (-), somewhat socially acceptable (+) and very socially acceptable (++). If a norm question was (randomly) chosen to determine payment, a subject was paid if they choose the most popular option. Participants were only eligible for this bonus if they only selected one option for each scenario. Note that while Krupka and Weber (2013) uses distinct samples, I follow d'Adda et al. (2016) and Erkut et al. (2015) in using the same sample for norms and contribution decisions.

Survey and payment

While participant payoffs are being calculated, participants were asked to complete an exit survey. Each question was read aloud, giving time for subjects to write their answers. Before answering the survey, a random question for the expectations and norms sections was selected to determine which question was used for payment. Further details on survey questions, including encoding, are included in appendix 1.

4.3 Theoretical and Empirical Approach

My experiment uses a standard one-shot public goods game. Individual i must decide upon some action a_c that best reflects their 'direct' utility $V_i(a_c)$. Benjamin et al. (2010) discuss this as norm-free utility, and it can more generally be understood as one's own social preference for greed, altruism or inequality aversion.

$$U_i(a_c) = V_i(a_c) \tag{4.2}$$

To enrich this analysis, early economic research focused on injunctive norms. For example, Elster (1989c) argues that where injunctive norms are shared, we should expect them to influence behaviour. These ideas find later echoes in more recent work in identity economics, and in explaining behaviour in experimental games (Akerlof and Kranton, 2000; Cappelen et al., 2007; Benjamin et al., 2010; Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016; Chang et al., 2019). A key debate regards whether behaviour is best explained by social preferences or injunctive social norms (i.e. $V()$ or $N()$ in the below; see Binmore, 2010 and Gächter et al., 2013 for opposing views). I include the pull of injunctive social norms below, where N is an increasing function, γ captures norm sensitivity and a_c^* is the prescribed action.

$$U_i(a_c) = V_i(a_c) - \gamma_i N(|a_c - a_c^*|) \tag{4.3}$$

A separate line of research focuses on empirical expectations. Ellingsen et al. (2012) find that cooperation is higher when a Prisoner's dilemma experiment is described as the Community Game than when it is described as the Stock Market Game. Their results indicate that the frame acts as a coordination device: "social frames enter people's beliefs rather than their preferences" (Ellingsen et al., 2012, p.117). A similar logic can be found elsewhere (e.g. Alekseev et al., 2017; Capraro et al., 2024), with certain treatments potentially changing one's expectations of others' behaviour rather than one's own preferences. I include this below using E to capture expectations as a strictly

increasing function, δ to capture sensitivity and a_c^e the expected action.

$$U_i(a_c) = V_i(a_c) - \delta_i E(|a_c - a_c^e|) \tag{4.4}$$

Others argue that injunctive norms and empirical expectations are both needed for norms to hold sway (Bicchieri, 2005, 2016; Bicchieri et al., 2022; Kölle and Quercia, 2021). Bicchieri (2005, p.11) argues that a "sufficiently large subset" of people need to agree on both empirical expectations *and* normative judgements for them to affect behaviour, as neither is a sufficient condition on its own. I include this idea below, with the expectation that both $\gamma > 0$ and $\delta > 0$ wherever there is an established norm. Unfortunately, the nature and conditions for a 'shared established norm' are neither shared nor established in the literature, but I return to this issue later.

$$U_i(a_c) = V_i(a_c) - \gamma_i N(|a_c - a_c^*|) - \delta_i E(|a_c - a_c^e|) \tag{4.5}$$

If language does affect contributions, it could do so according to any of the four reasons above. Language could change one's own valuation of different actions, activate injunctive norms, act as a coordination device by changing expectations, or make different established norms salient.

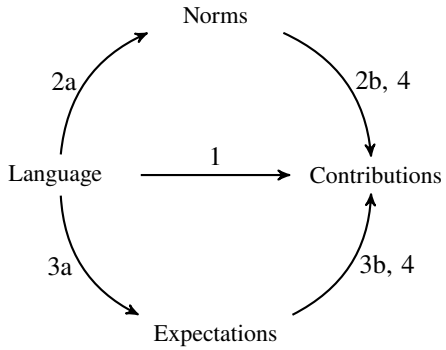
4.3.1 Family-Wise Error Rate (FWER) Strategy

Within each family of hypotheses, I will control for the number of tests conducted, controlling the FWER using the Hochberg procedure implemented in Stata by Newson (2010). I will use the 5% significance level for all decisions, using the appropriate q values. The families of hypotheses here are:

1. Does the language of the experiment affect contributions?
2. Does the language of the experiment affect contributions through norms?
3. Does the language of the experiment affect contributions through expectations?
4. What is the (additional) effect of expectations and norms on contributions, at individual and treatment level?

The families are also described in figure 4.2, showing when families are broken into two parts.

Figure 4.2: FWER and Research Questions



4.3.2 Empirical Strategy

All analysis was pre-registered (Clist and Hill, 2019). The first family of tests are simply whether language affects contributions in the public goods game, i.e. comparing $U_i(a_c)$ by language. I start with a parsimonious regression, with standard errors clustered at the session level. Given the anticipated small number of clusters, I use wild cluster bootstraps (implemented by Roodman et al., 2019). The first regression will use the basic set up:

$$Y_i = \alpha + \beta Language_i + \epsilon_i$$

where Y_i is the contribution to the common pot. Clist and Verschoor (2017) found lower contributions in Lugisu than Luganda, but I keep the more conservative two-sided tests as the urban setting may mean different behaviour by language. The second regression will include controls. These will comprise age (capped at 30, given the student population and the desire to not be overly influenced by outliers), gender, the number of people in the room a subject knows (which may increase cooperation) and the number of years a subject has spent in Kampala (coded as 0 for less than 2.5 years, 1 for 2.5-5.5 years and 2 for over 5.5 years). This variable is transformed into a categorical variable to avoid the pull of outliers (as pre-registered).

Moving to family 2, I ask whether language affects contributions through norms. This is a two-part hypothesis: first, that language affects norms, and in turn that norms affect contributions. I apply two FWER strategies: controlling for one family (2) or for each sub set (2a and 2b) separately. 2a questions whether norms differ by language and 2b investigates whether norms affect contributions. The rationale for this is that in order to conclude that evidence is supportive of 2a, it is best to control for all tests that are relevant to 2a. Similar logic can be applied to 2b. However, 2 rests on both 2a and 2b, and so all of the relevant tests need to be controlled for. In the language of equation 4.5, the

tests are that a) $N(|a_c - a_c^*|)$ differs by language, and b) that $\gamma > 0$. If there is support that contributions are different by language (family 1), but not for the mechanism through norms (family 2), it is possible that language changes the inherent preferences of individuals, i.e. that $V_i(a_c)$ differs by language (see Benjamin et al., 2010). Alternatively, the mechanism may be expectations rather than norms: that mechanism is explored in family 3.

For family 2a (whether norms differ by language) the standard approach (Krupka and Weber, 2013; Chang et al., 2019) has been to conduct a series of rank sum tests, to see whether the norm ratings differ between two groups. A concern here is that the series of tests mean that false positive results are made more likely. As such I follow the standard approach, but control for five tests. In translating a norm rating to a numerical scale I follow Krupka and Weber (2013) in coding the norm ratings as -1, -1/3, 1/3 and 1. In the language of equation 4.5, this is a test of $N(|a_c - a_c^*|)$ by language for $c = [0, 12]$.

Family 2b (whether norms affect contributions) is effectively testing the γ parameter in equation 4.5. I follow Gächter et al. (2017) in modelling unobserved heterogeneity in the response to the treatment-level norm by using a mixed logit (in other words, rather than averaging at the contribution level, this analysis will consider the full heterogeneity within the data). The result is an estimated mean and standard deviation for γ, the distribution of which can be shown graphically to reveal the degree to which participants conform to or flaunt norms. I include the level of the contribution as a control. If people were purely selfish and only constrained by norms, this would have a negative coefficient. However, I have no prior expectations over this parameter. It is worth noting that the identification of the effect of norms will be in relation to this linear effect of contributions.

Moving to family 3, I ask whether language affects contributions through expectations. As in family 2, I provide 3 FWER strategies: a, b and together, and do not control for the other mechanism. The 3a mechanism is analogous to family 2a, and so I keep all tests the same, conducting a series of rank sum tests, with expectations averaged as at the contribution level. There are differences in the nature of the data: norms is a four-point scale, whereas expectations could be any percentage that is a multiple of 10 between 0 and 100. However, pilot experiments lead me to expect a similar level of variation, with most guesses likely to be in the 10-40% range. To keep comparability with 2b, I will approach expectations in the same way as norms: as averaged over the entire treatment for a given contribution level (in family 4 I will return to the issue of individual level expectations/norms). This means I am not at this stage asking whether a given individual is a conditional cooperator. Rather, I ask whether the session-level differences in contributions can be explained by the session-level differences in expectations, as would be expected if conditional cooperation plays a part in determining behaviour.

Moving to family 4, I now consider the effects on norms and contributions at the same time. In testing family 4, I present two strategies. First, I follow the standard approach of the norms literature (e.g. Krupka and Weber, 2013) by including treatment-level ratings, in the case for norms and expectations. Second, I depart by running the same tests

but using individual-level ratings. This second test is able to show whether individuals are conditional cooperators, and whether they feel bound by the norm as they perceive it.

4.4 Sample Statistics

Table 4.3 reports summary statistics. The median participant is a 22 year-old male, who was born in Mbale and learnt Lugisu first. They have since spent 4 years in Kampala, speak Lugisu at their home, and know 2 other people in the session. I see some variation in where people are from, though only 3% are from non-Bagisu homelands. I note less variation in language with 85% having learnt Lugisu first, and 84% speaking Lugisu in their current home. The sample has more men than women, though this is broadly in line with the latest data on gender inequality in tertiary education in Uganda.[8] The random allocation mostly worked well, but there is one sign of a lack of balance. Those allocated to Lugisu sessions previously spent 1.74 more years in Kampala than those allocated to Luganda sessions. This variable is transformed into a categorical variable to avoid the pull of outliers (as pre-registered), and there is also a lack of balance in the transformed variable. Essentially, the randomisation allocated more morning sessions to Luganda (5/7). Students that spent less time in Kampala were more likely to be able to attend in the morning (presumably as they lived on campus), and so more recent arrivals ended up being statistically more likely to be allocated to a morning session. I was not aware of this, and had no control of this. The variable is only used once in one of the pre-registered tests, shown in table 4.4, where it is insignificant. When running additional specifications in a range of permutations across contributions, norms and expectations, by language, time in Kampala is never significant (either raw or transformed, at the individual or session level). Separately, as set out in chapter 5 section 5.5 the aggregated analysis on exposure across all specifications is insignificant

[8] World Bank data, available at `https://data.worldbank.org/indicator/SE.ENR.TERT.FM.ZS?locations=UG`, provides a gender parity index. The last observation is from 2014, and implies 42% of university students are female.

Table 4.3: Summary Statistics, by Treatment

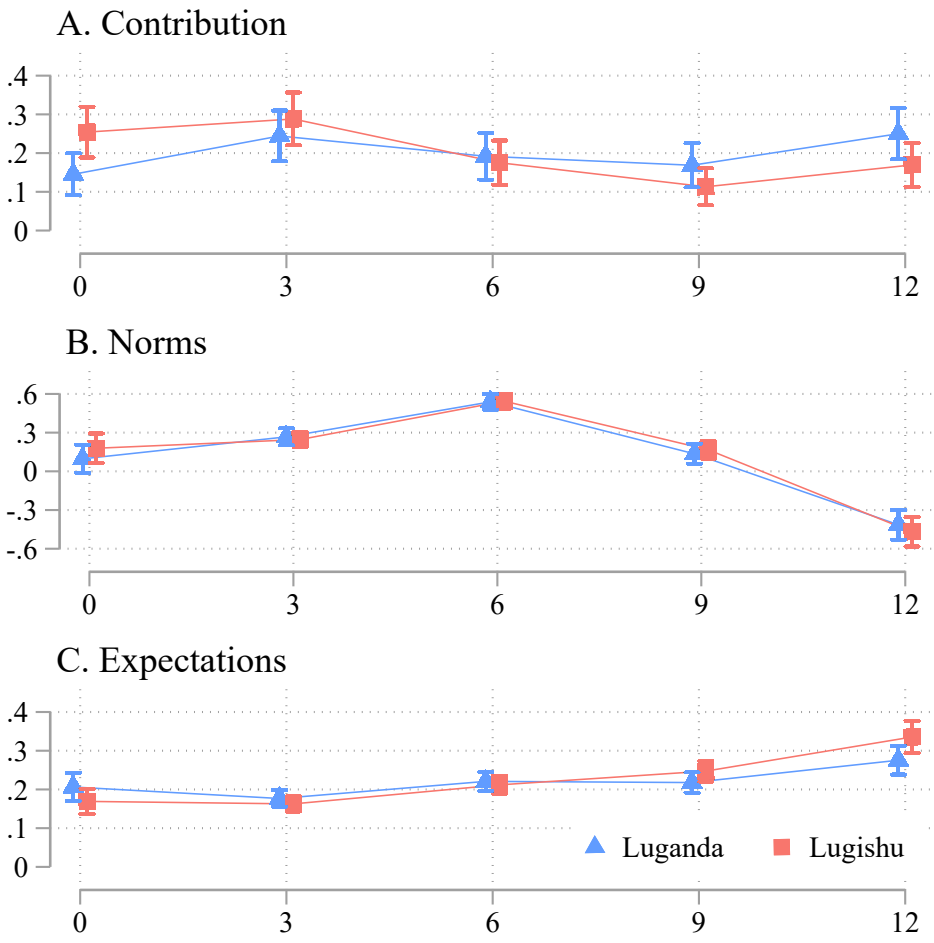
	Session Language			Balance Tests
Variable	Luganda	Lugisu	Total	*p*
Female	0.378 (0.486)	0.367 (0.483)	0.372 (0.484)	0.837
Age (capped at 30)	22.279 (2.084)	22.271 (2.322)	22.275 (2.205)	0.973
# people known in session	3.483 (3.841)	3.113 (3.021)	3.295 (3.449)	0.318
Years in Kampala, Raw	4.267 (3.384)	6.006 (5.178)	5.149 (4.466)	<0.001***
Years in Kampala, Transformed				0.017**
< 2.5 years	59 (34.3%)	41 (23.2%)	100 (28.7%)	
2.5-5.5 years	72 (41.9%)	72 (40.7%)	144 (41.3%)	
> 5.5 years	41 (23.8%)	64 (36.2%)	105 (30.1%)	
Main Language Used At Home				0.971
Lugisu	143 (83.1%)	149 (84.2%)	292 (83.7%)	
Luganda	11 (6.4%)	12 (6.8%)	23 (6.6%)	
English	15 (8.7%)	13 (7.3%)	28 (8.0%)	
Other	3 (1.7%)	3 (1.7%)	6 (1.7%)	
First Language Learnt				0.406
Lugisu	151 (87.8%)	145 (81.9%)	296 (84.8%)	
Luganda	7 (4.1%)	13 (7.3%)	20 (5.7%)	
English	8 (4.7%)	9 (5.1%)	17 (4.9%)	
Other	6 (3.5%)	10 (5.6%)	16 (4.6%)	
Birth Location				0.418
North Bugisu Region	53 (30.8%)	51 (28.8%)	104 (29.8%)	
South Bugisu Region	33 (19.2%)	48 (27.1%)	81 (23.2%)	
Central Bugisu Region	23 (13.4%)	18 (10.2%)	41 (11.7%)	
Mbale	58 (33.7%)	53 (29.9%)	111 (31.8%)	
Other	5 (2.9%)	7 (4.0%)	12 (3.4%)	
N	172 (49.3%)	177 (50.7%)	349 (100.0%)	

Note: The first four p values are for binary or continuous variables, and come from regressions. The second four come from Pearson tests. In each case *** denotes significance at 1%, ** at 5% and * at 10%. Standard deviations, or percentages, are provided in parenthesis. As pre-registered, I capped age at 30, which affects three people, aged 32, 33 and 47. The other languages used at home comprise Swahili, Lugere and Lusoga. The other languages first learnt include Swahili, Lugere, Lusoga, Luyankole, Ruqiso and Rutoro.

4.5 Experimental Results

The core results are shown in figure 4.3. The three panels respectively show the level of contributions as well as the norm ratings and expectation at each contribution level by language.

Figure 4.3: Average Contributions, Norms and Expectations, by Session Language



Note: Error bars denoting plus or minus two standard errors are shown in each panel.

In the next three subsections, I present statistical tests of whether these differ by language. Each subsection concerns one family of hypotheses (e.g. do contributions differ by language) for which we control for the family wise error rate. I present q values, calculated using the Hochberg procedure implemented in Stata by Newson (2010).

4.5.1 Does Language Affect Contributions?

Table 4.4 shows the treatment effect. The average contribution is around 5,000 in Lugisu and 6,400 in Luganda. The effect size is around a third of a standard deviation, and is similar with and without controls. It is significant at the 5% level after adjusting for the small number of clusters (Roodman et al., 2019) and correcting for multiple hypotheses tests. The top panel of Figure 4.3 shows the effect is clear across the distribution.

Table 4.4: Does Session Language Affect Contribution Choice?

	(1)	(2)
Lugisu Session	-1.435**	-1.469**
	(-2.74)	(-2.73)
Female		0.0814
		(0.17)
Age		0.117
		(0.90)
Years in Kampala		0.127
		(0.43)
# people known in session		-0.0173
		(-0.31)
Constant	6.401***	3.701
	(19.09)	(1.28)
Wild-Cluster Bootstrap p	0.023**	0.032**
Hochberg q	0.032**	0.032**
N	349	349

Note: The dependent variable is the contribution to the common pot in thousands of Ugandan Shillings, and spans 0-12 in increments of 3. There were 14 sessions, with an average of 24.9 subjects per session, and a range of 11 to 30. Standard errors are clustered at the session level, using Roodman et al. (2019) `boottest` command. t statistics are provided in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. Hochberg q values show adjusted p using the Hochberg procedure implemented in Stata by Newson (2010) to control for the FWER.

The effect size, at 28.9%, is remarkably close to Clist and Verschoor (2017) finding of a 28.8% difference. This is despite differences in subject pool (students, rather than the general population), year (the experiments were conducted four years apart), experimental design (the previous experiment included a controlled cheap talk message from one player to the other) and payments (the previous experiment used initial endowments of 8,000 Ugandan Shillings). This experiment can be partially viewed as a replication, considering four approaches in appendix 2. One might expect a smaller effect size to be found here, given that I replicate the largest language effect previously found, and replications typically find an effect of around two-thirds the original finding (Camerer et al., 2016). In this appendix, I show the effect sizes are similar, and that the evidence is strongly against a null effect.

Result 1: Contributions are about a third higher in Luganda than in Lugisu. The effect size is within 0.1 percentage point of previous work.

Having established an effect of language on contributions level, I will now investigate any treatment effects from language on norms and expectations.

4.5.2 Do Norms Differ by Language?

The middle panel of figure 4.3 shows there is very little difference in norms by language. Table 4.5 breaks this data down, reporting the norm ratings for each possible action by language. Following Krupka and Weber (2013), the scale runs from -1 to 1, denoting 'very socially unacceptable' and 'very socially acceptable', with mid points ($\pm 1/3$) referring to 'somewhat'. Table 4.5 shows that, once multiple testing corrections have been applied, the resulting q values are all above 0.995: injunctive norms do not differ by language.

The shape of norms, shown in figure 4.3, are not standard in Western laboratory experiments. Amongst the participants the most socially acceptable action is to give half of the endowment to the common pot, even though the midpoint was not mentioned in the script. More surprisingly, people correctly state that the norm is *not* to give everything; contributing everything is the least socially acceptable action, with an average rating of -0.44. 'Hump' players are well documented, but are not so prominent in Western experiments (Fischbacher and Gachter, 2010; Kimbrough and Vostroknutov, 2016; Weber et al., 2023). The hump in norms is not due to misunderstanding as 52% of subjects correctly guess the most common appropriateness level. This is comparable to previous work, with 55% guessing correctly in Krupka and Weber (2013). Further, the instructions were correctly interpreted, with only 11 of 360 subjects (3%) failing the control questions. Considering that the sample is comprised almost entirely of university students, the result is perhaps not surprising. Excessive generosity may be considered impolite in the absence of close social ties, where recipients feel disproportionately burdened to reciprocate. Equally, a lack of generosity may similarly be perceived as impolite (see Jones (2022) for a recent meta-analysis).

Table 4.5: Norm Ratings, by Contribution Level and Session Language

	Norm Rating								Test Stats.	
	Luganda				Lugisu				Rank Sum	
Contribution	-1	-1/3	1/3	1	-1	-1/3	1/3	1	p	q
0	32	43	51	46	33	42	35	67	0.25	0.9956
3	3	36	108	25	3	36	119	19	0.65	0.9956
6	1	9	98	64	0	9	103	65	1.00	0.9956
9	6	65	75	26	1	67	84	25	0.60	0.9956
12	94	32	19	27	108	25	16	28	0.35	0.9956

Note: N=172 for Luganda, and N=177 for Lugisu. The left panels show how many participants under each treatment (Lugisu and Luganda) selected each level of social acceptability for each contribution level. The ratings are consistent with the original Krupka and Weber (2013) design, where -1 / 1 means very socially unacceptable / acceptable, and -1/3 / 1/3 means somewhat socially unacceptable / acceptable. *** p<.01, ** p<.05, * p<.1. Hochberg q values show adjusted p using the Hochberg procedure implemented in Stata by Newson (2010) to control for the FWER.

Result 2: Norms do not differ by language, so cannot explain the difference in contributions by language. Average norms are hump (or triangle) shaped in both languages, unlike in experiments with Western subjects.

4.5.3 Do Expectations Differ by Language?

The bottom panel of figure 4.3 shows a small difference in expectations by language, with figure 4.4 presenting this as a cumulative distribution. Table 4.6 reports the mean and standard errors of expectations by language for each contribution level. In Lugisu, more people are expected to contribute the full amount (33.6% versus 27.5%), and fewer people are expected to contribute nothing (16.9% versus 20.6%). Whilst three of the five effects are weakly significant, the q values in Table 4.6 show language differences are not significant once controlling for the multiple tests.

It should be noted that the expectations do not total to 100% in either treatment. As set out in the experimental script in appendix 1, participants were not asked to ensure that these responses total 100%, but rather to state how many people in the room contributed the amounts shown in 10% intervals. As I am interested in the relative difference by treatment at the contribution level, this is not believed to affect the analysis. Indeed, forcing expectations to total 100% would; even if the proportions between the contribution levels were maintained, this would change the relative values by treatment at each contribution level. For robustness, when doing this, the results are consistent with the analysis presented once the FWER procedures have been implemented, in that expectations do not significantly vary by language.

Figure 4.4: Cumulative Expectations, by Language

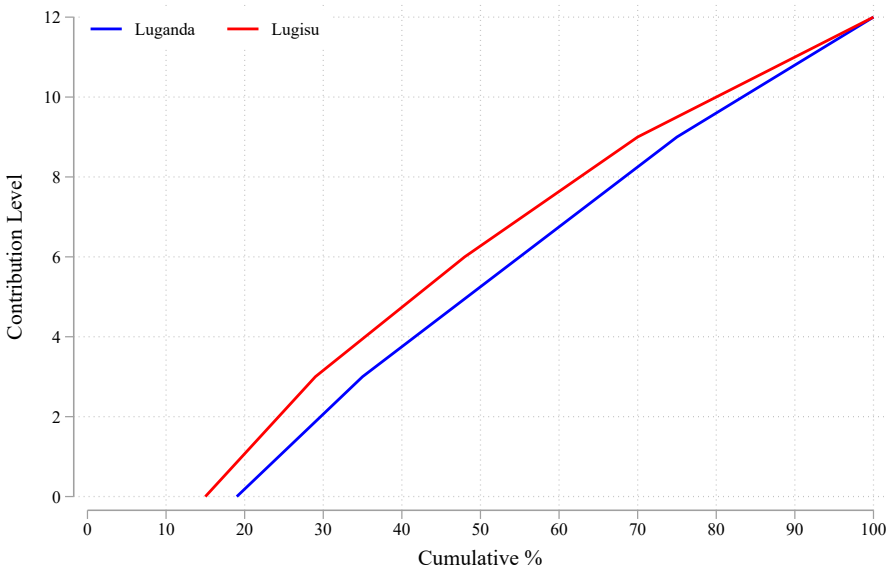


Table 4.6: Expectations, by Session Language

	Luganda		Lugisu		Test Stat.	
Contribution	Mean	s.e.	Mean	s.e.	p value	q value
0	20.6%	1.8	16.9%	1.7	0.08*	0.24
3	17.7%	1.1	16.3%	1.0	0.38	0.61
6	22.1%	1.2	21.2%	1.2	0.61	0.61
9	21.8%	1.3	24.7%	1.4	0.05*	0.22
12	27.5%	1.8	33.6%	2.1	0.05*	0.22

Note: The p values come from a rank sum test. Hochberg q values show adjusted p using the Hochberg procedure implemented in Stata by Newson (2010) to control for the FWER. Standard errors (s.e.) are provided.

Subject's empirical expectations are inaccurate on average, as people expect others to be more generous than they are. Pooling across languages, people contribute the maximum 21% of the time, but this is expected between a quarter and a third of the time (Table 4.6). Wishful thinking is common in other experiments, but typically smaller.[9] I find contributions are lower than expectations in both languages, but the difference is more pronounced in Lugisu. Whilst people expect slightly more generosity in Lugisu (an insignificant difference), they are actually more generous in Luganda.

Result 3: Expectations do not differ significantly by language, so language does not act as a coordination device. There is wishful thinking in both languages, as people expect others to be more generous than themselves. The gap is larger in Lugisu, with people mistakenly expecting higher contributions.

4.5.4 Do Norms And Expectations Agree?

Results 1-3 show that while contributions do differ by language, injunctive norms and empirical expectations do not. This seems to rule out norms and expectations as the potential mechanism for differing levels of cooperation by language. However, Bicchieri (2005, 2016) argues norms only hold sway over actions when they are established i.e. when injunctive norms and empirical expectations align. In this subsection, I ask whether that is the case in this setting - do people expect others to behave in accordance with an injunctive norm? If they do not align in this setting, I cannot test Bicchieri's view, as there is no established norm. I use three methods to answer this question.

First, I consider the whole range of actions, and ask whether there is agreement between norm ratings and expectations ($cov(N, E) > 0$) at the population level. This is clearly not the case from figure 4.3; contributing everything is both the most expected action and the least socially acceptable. Taking the average for each action

[9] Weber et al. (2023) run experiments in the USA, UK, Morocco and Turkey using a 0-20 scale. They find average unconditional cooperation of 6.4, a standard deviation of 6, and average expectations of 8.1. Contributions are then lower than expectations by around a quarter of a standard deviation. I find a mean expected contribution of 7.79, an average contribution of 5.67 and a standard deviation of 4.28. This gives a difference of half a standard deviation. Wishful thinking was present for each of their countries; it was largest in Turkey and smallest in the UK.

($cov(\bar{N}_c, \bar{E}_c)$ where c denotes a possible action), I can see the correlation is strongly negative, Pearson's $\rho = -0.73$ (though with only 5 data points, $p = 0.17$).

Next, I see whether individuals expect actions which they think are more social acceptable, i.e. $cov(N_i, E_i)$ at each level of contribution. Calculating a correlation statistic for each of the five possible actions, where identification comes from individual-level variation, people's norm ratings and expectations are positively associated. Pearson's ρ varies between 0.18 and 0.27, with the largest p value 0.0005. To give a sense of the effect size, I can report the coefficient from regressions of norms on contributions at all 5 levels. The beta regression coefficients range from 0.18 to 0.27, i.e. a one standard deviation higher norm rating is associated with around a quarter higher expectation. The association, at the individual level, is significant and positive: individuals that give high norm ratings to a given action also tend to expect it more.

Lastly, I ask whether there is a clear implication from norms and expectations ($maxN_c = maxE_c$), i.e. whether a single action is both the modal expectation and most socially acceptable. As long as there is a clear course of action implied by norms and expectations, the specific ratings regarding other actions are not materially important. Table 4.7 classifies subjects, showing how often they have a clear normative action, defined by a single peak in their injunctive norms or empirical expectations. The third line (Both, Strong) captures people that have the same single peak in norms and expectations - if people are norm following, there is a single clear action to follow here. The next line includes only people who have one peak in their norm or expectation which corresponds to one of two peaks in the other. This group seems to have a clear norm in Bicchieri (2010) definition, in that one action is clearly recommended, but not as clearly as the 'strong' group.

Table 4.7: Expectations and Norms, by Peak & Session Language

	Single Hump at					By Session Language		Total
Contribution:	0	3	6	9	12	Luganda	Lugisu	
Injunctive Norms	88	22	88	33	42	136	137	273
Empirical Expectations	61	23	49	56	122	154	157	311
Both, Strong	30	8	24	13	30	55	50	105
Both, Weak	17	4	10	11	9	26	25	51
N						172	177	349

Note: One subject has flat ratings, answering the same for all contribution levels for norms and expectations.

Table 4.7 shows that there is no single established norm in either language. Less than a third of subjects (105/349) think the most likely action is the most socially acceptable. Less than half have a Pareto-efficient norm to follow. Amongst the subset that do have clear courses of action, there is great disagreement about what that action should be, with the most popular options being to give everything and to give nothing. Table 4.7 also shows that there is no difference in the likelihood of a norm existing by language, with very similar numbers in all cases.

Result 4: Whilst there is an association between norms and expectations at the individual level, there is no established norm. Less than a third of subjects think the socially most acceptable action is the most likely action. More than half of subjects would need to choose between following injunctive norms and conditional cooperation. Amongst those that do have a clear course of action, there is disagreement about which action is the established norm.

4.5.5 Do Norms and Expectations Affect Contributions?

In this setting there is no established norm, but I can still test if either norms or expectations affect behaviour. To do this, I use a mixed logit model, which models unobserved heterogeneity in the sensitivity to norms and expectations: γ and δ in the language of (4.5). This is not possible using the original approach applied by Krupka and Weber (2013), which averages norms at the contribution level. Further, as set out in Gächter et al. (2017), the mixed logit model has a number of advantages over a standard logit model in terms of capturing heterogeneity in preferences among individuals and in relaxing the unrealistic Independence of Irrelevant Alternatives (IIA) assumption. In each case I include a linear effect of contributions, which would be negative if people are selfish. A conditional cooperator would have a positive coefficient on expectations, as they would be drawn towards what they expect others to do. Likewise, norm-followers would have a positive coefficient on norm ratings.

I measure both norms and expectations at two levels (see table 4.8). In column 1 I use 'treatment level' averages, i.e. the average norm or expectation rating for that contribution level in a given language. This is how Krupka and Weber (2013) identifies the effects of norms, and relies on differences by treatment and contribution level to identify effects. In column 2 I use 'individual level' ratings, to capture how a subject responds *to their own* norm ratings and expectations. Further, I can build on the idea of the last subsection, and see whether having a clearly implied norm affects behaviour.

Whilst there is no established norm (as people do not agree which behaviour is normative), some people expect the most socially acceptable behaviour to be the most popular. In table 4.8, column (3) includes a dummy for those with the same single peak in both norms and expectations. Column (4) includes a dummy for a clear signal (i.e. one hump in norms or expectations which coincides with one of two humps in the other). Whilst norms are not established and shared, this would be an intermediate step towards establishing a norm: people responding to their own judgements.

Table 4.8: Mixed Logit for Sensitivity to Norms and Expectations

	Treatment Level	Individual Level		
Definition of Signal :	-	-	Strong	Any
	(1)	(2)	(3)	(4)
Mean				
Contribution	0.0281	-0.0371**	-0.0359**	-0.0347**
	(1.01)	(-2.26)	(-2.24)	(-2.18)
Standardised Norm Rating	-0.170	-0.184**	-0.209***	-0.229***
	(-1.41)	(-2.51)	(-2.76)	(-3.04)
Standardised Expectation	-0.652***	-0.0446	-0.0770	-0.103
	(-3.82)	(-0.37)	(-0.78)	(-1.02)
Normative Signal			0.354	0.513*
			(1.00)	(1.66)
SD				
Standardised Norm Rating	0.881***	0.575***	0.521***	0.484**
	(3.78)	(3.12)	(2.62)	(2.39)
Standardised Expectation	0.439	-0.159	0.170	0.156
	(1.27)	(-0.28)	(0.52)	(0.47)

Note: This makes use of the `mixlogit` command in Stata (Hole, 2007). The dependent variable is *choice*, which is 1 if that contribution level was selected by the participant and 0 otherwise. t statistics are included in parentheses. *** $p < 0.01$, ** $p < .05$, * $p < .1$. N=1,660.

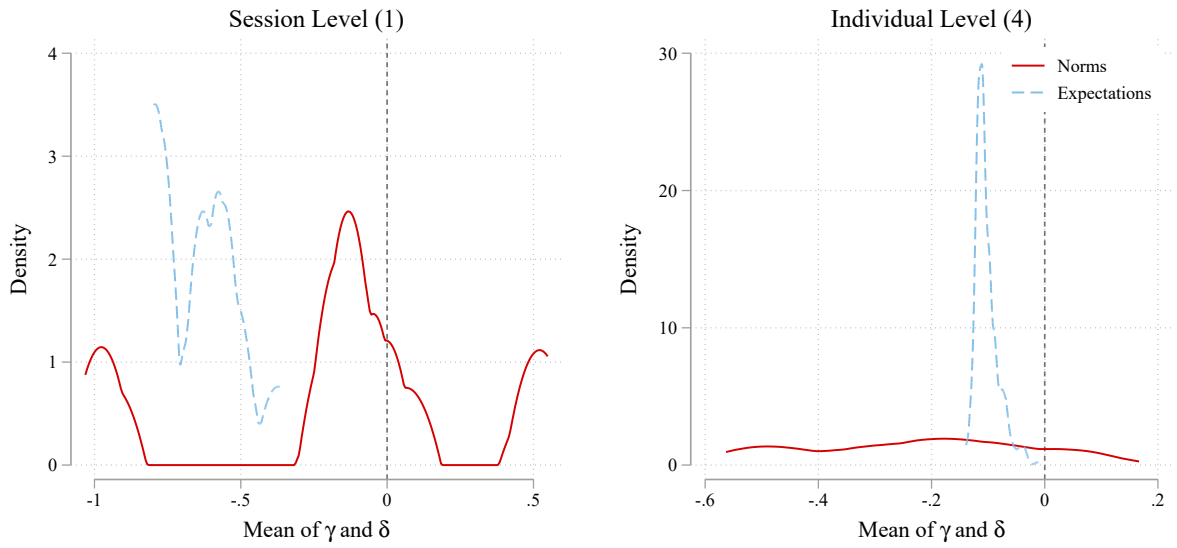
Across specifications, the average person is always estimated to have a negative coefficient on norms and expectations. This means subjects are more likely to flout norms and expectations than conform to them. Whilst true at both levels, this is particularly surprising at the individual level, as the average person acts counter to their own judgements of injunctive norms and empirical expectations. For expectations this is significant at the treatment level. This is identified using differences by treatment and contribution level. In essence this restates result 3, which shows wishful thinking in expectations, with subjects also expecting more generosity in the 'wrong' language. People tend to be less likely to do things if they expect others to do them, which runs contrary to conditional cooperation. The effect is clear in figure 4.3's simple data summary: the least expected behaviour (contribute 3,000 shillings) is also the most popular behaviour. For norms, this is significant at the individual level, meaning people are *less* likely to do actions which they themselves think are normative.

These results are easy to reconcile with Bicchieri (2010) account of norms holding sway, as there are no established norms in this setting. Yet, when there is a signal, it appears people do respond to it. This is true when an action is unequivocally the best action in both norms and expectations, or when it is the Pareto optimum. The effect size is reasonable, and weakly significant in (4). The marginal results imply that for actions with a signal, they are 4.6-7 percentage points more likely to be chosen; the low range comes from column (3), the higher from (4).

There is substantial evidence of heterogeneity in sensitivity to injunctive norms, at both session and individual levels. Figure 4.5 shows the distribution of parameters, using columns (1) and (4) from table 4.8. There is only limited

evidence of such heterogeneity in sensitivity to expectations. At the treatment-level there are some norm-followers, but the majority are either relatively indifferent or more likely to diverge than conform.

Figure 4.5: Heterogeneity in Sensitivity to Norms and Expectations



Note: I use 1,000 replications to plot the distribution of parameters. This makes use of Stata's `mixlogit` command (Hole, 2007).

The results show that the majority of the subject pool are neither norm-followers nor conditional cooperators. This can be rationalised in three ways. First, the lack of an established norm means normative and empirical judgements are often in conflict, which frees subjects to ignore normative concerns. Second, for the subset of people that do feel some nascent sense that there is no conflict between following the normative judgements and their empirical expectations, they are more likely to choose the appropriate action. Third, the subject pool differs from the majority of literature (Fehr and Schurtenberger, 2018), which mainly uses 'WEIRD' subjects (i.e. those from Western, Educated, Industrialised, Rich and Democratic countries; see Henrich et al., 2010). As such, a higher proportion of hump-shaped players, and few conditional cooperators, is potentially more representative.

Result 5: These non-WEIRD subjects mostly flout both norms and expectations. They are neither conditional cooperators, nor norm followers. Yet if an action is clearly implied by one's own empirical expectations and injunctive norms, there is suggestive evidence the norm is weakly influential.

4.5.6 Measuring Norms and Expectations

While the design adopted in this chapter has been widely applied across the literature (see Nosenzo and Görges (2020) for a review), it is possible that there is a disconnect between the contribution in the PGG and the more artificial assessment of norms and expectations. As discussed by Kahneman (2003), reasoning is done deliberately and effortfully (System 2), but intuitive thoughts seem to come spontaneously to mind, without conscious search or computation (System 1). A key attribute of System 1 thinking is accessibility, being the ease to which a thought comes to mind. Within my design, deciding how much to contribute in the PGG is likely to be achieved intuitively, through System 1 thinking. This decision is substantially more accessible than the more labour-intensive assessment of norms and expectations. The latter involves a judgement of both an artificial set of norms and expectations, as well as an assessment of how best to coordinate with other participants in order to maximise the probability of receiving the bonus payout. Thus, the assessment of norms and expectations likely requires System 2 thinking. This disparity may contribute to the insignificant results noted in sections 4.5.2 and 4.5.3.

Further, the noted results could also be explained by a norm of *ignoring norms*. While norms are successfully measured, as the majority of participants correctly guessed the most popular acceptability level for norms, or being within 10% for of the selected question for expectations, it may be that the actual norm is to ignore this. For example, Bicchieri (2010) discusses scenarios where individuals might choose to relax their commitment to social norms, especially when defying normative expectations carries no significant consequences. She suggests that in such situations, it may be reasonable for individuals to adjust their adherence to social norms. In the PGG, given the anonymous set up, there is little or no cost if participants choose not to adhere to the norms elicited. Thus, norms may still be a mechanism, where the *norm* is not to adhere to the elicited social norms. Additionally, a remarkably high proportion of economics research has been conducted in the United States and western Europe, despite 80% of the global population living in Non-Western, Educated, Industrialised, Rich and Democratic ('WEIRD') countries (see Sharma and Siddique (2024) for a recent literature review). Thus, it may be that findings from the existing literature do not directly translate where context and culture vary so substantially.

4.6 Discussion & Conclusion

I find that in a public goods experiment, randomising language has a predictable and large effect on contributions. I can rule out several compelling explanations; comprehension does not differ, and there are not important inherent differences in the languages themselves. They do not activate different norms, and are not used as a coordination device by conditional cooperators. Nor can this effect be dismissed as a one-off, as it closely replicates a previous experiment. The effect size is practically identical (28.9% versus a previous result of 28.8%), despite various changes including the location (the capital city rather than a rural area), and subject pool (students rather than rural Ugandans).

Having ruled out several explanations, I am left with the idea that language affects people's own preferences for cooperation. In other settings I think it is plausible that language could act as a coordination device, in the same way the name of a game has been found to do so (Ellingsen et al., 2012). Likewise, languages could activate different norms (Elster, 1989c), in other settings. I find language appears to affect one's preferences directly, or at least in a way that is not captured by the elicitation of norms and expectations (as per Gächter et al., 2013). The direction of the effect matches the available evidence of the subject pools' history; the Gisu are less cooperative when speaking their tribal language, which is associated with a range of development hurdles.

A major reason why language does not appear to affect actions through norms may be that they are not established in this context. Neither norms or expectations appears sufficient to affect behaviour on their own. This may also explain why most subjects in this setting are neither conditional cooperators nor norm followers, though that could also be due to the non-WEIRD subject pool. There is some evidence that for a small number of participants, for whom their normative expectations meet their empirical expectations, this affects behaviour.

These results may have implications for how we understand the importance of language in everyday life. The Ugandan context is typical of many developing countries, with most people communicating in multiple languages each day. I find the language in which an interaction takes place (keeping expectations over one's partner's identity constant) is important: around a 30% difference in cooperation. If this is representative of daily life, multiple languages may mean multiple possible outcomes. In other situations this could conceivably operate through differences in norms or expectations. However, the Gisu prefer to be unconditional cooperators in the national language and unconditional non-cooperators in their own. Investigating these relationship is an opportunity for further research, specifically investigating the impact of using different languages, across different groups and contexts.

Chapter 5

Chapter 5 - Hypothetical Norms & Cultural Exposure

5.1 Introduction

Norms are ubiquitous at all levels of human social interaction, from the mundane to the profound. When obeyed, norms permit predictable behaviour and social order which are essential for large scale contemporary societies (Elster, 1989b). The importance of norms in decision making illustrates the need to include social determinants of decision making within theories of choice, moving beyond the standard model of economics. Research in *behavioural economics* has shown that the context of the moment of decision can affect choice (Cialdini et al., 2006; Kremer and Levy, 2008; Herbst and Mas, 2015; Thaler and Sunstein, 2008). Hoff and Stiglitz (2016) conceptualise this in a simple taxonomy where an individual's preferences and cognition are subject to two deep social influences: (a) 'social context' (exposure to environment); and (b) 'mental models' (cultural categories, identities and narratives) used to process information. An individual may have multiple mental models (or schema) which shape how they will interpret, remember and emotionally respond to the information they encounter (DiMaggio, 1997). These mental modes can be internally inconsistent and thus lead to very different choice behaviour, depending on context (d'Andrade, 1995; Swidler, 2013). Importantly, social context *creates* the set of mental models available to the individual, and affects the scenarios that prime alternative mental models (Hoff and Stiglitz, 2016).

Social norms govern behaviour in interactions that have multiple equilibria and thus are by definition highly susceptible to context. This means that social norms can cue or prime a specific mental mode in the moment, or directly inform the mental modes with increased exposure. By nature a normative standard is commonly known by all group members, where group membership is finite (Fehr and Schurtenberger, 2018). Thus, an individual can have membership to multiple 'sets' of social norms, some of which will overlap and be contrary. Social norms can be further broken down into constituent components. I adapt the terminology applied in Bicchieri (2010) and use *expectations* to refer to empirical expectations and *norms* to refer to normative beliefs. This distinction permits an investigation into the impact of norm adherence; an individual may be aware that different norms exist across contexts, but that doesn't mean that they adhere to different norms equally.

Norms are typically self-enforcing at a group level and evolve over time through a course of experimentation and adaptation (Young, 2015). Several types of psychological and social mechanisms support the emergence and maintenance of norms. Social pressure, reference points and symbolism all contribute to supporting behavioural norms. Conversely while norms are typically persistent, occasional bursts of change reform old norms into new (the "punctuated equilibrium" effect – see Young (1998)). An important feature here is local conformity and the resulting global diversity, particularly when communities interact only very occasionally, norms may follow very different trajectories of evolution. Thus, individuals are likely to be aware of and encounter scenarios where norms are contrary.

While the inclusion of social determinants broadens the explanatory potential of economics, in reality predicting real-world behaviour is challenging. The interaction between context, culture and identity is often difficult

to disentangle, particularly when individuals have membership to multiple overlapping groups. For example, context may only interact with identity in one geography, but not another. Measuring these interactions would have significant benefits to policy makers and provide real-world applications for theory. Like norms, cooperation is often present in almost every form of human interchange. Social dilemmas pervade a myriad of social scenarios, both for our ancestors and modern humans. Understanding why individuals cooperate (as oppose to free-ride) is a central question in within the social sciences, particularly in addressing market failure. There is a large literature linking cooperation to social norms (Fehr and Schurtenberger, 2018; Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016; Bicchieri, 2005), but more limited experimental research exploring whether individuals house multiple sets of social norms which are context dependent. Hypothetical norm elicitation permits a within-subject exploration of the impact of context (or 'place') or frame on social norms. This approach, introduced here, asks subjects to first elicit in-session social norms for a scenario (such as a public goods game), and then to imagine they are in the same scenario, but in another place (Kampala vs Mbale 2019: 'Experiment 1') or under another frame (Community vs Wall Street Game 2024: 'Experiment 2'). In all situations, subjects are attempting to coordinate with other subjects in an incentive compatible way. This extends the Krupka and Weber (2013) methodology and is the first contribution of this chapter.

A further contribution of this chapter is with regards to the interaction between place and language (and the effects on norms and expectations). Language is an important behavioural cue with strong cultural and ethnic associations (Habyarimana et al., 2009). This is externally relevant as multilingualism is a global norm and daily necessity for the majority of the world, with 7,105 known active languages across 200 nations (Crystal, 2012; Lewis, 2013). Both language (Jakiela and Ozier, 2018; Clist and Verschoor, 2017; Chen, 2013; Li, 2017; Lambarraa and Riener, 2015) and place (Berge et al., 2018; Cameron et al., 2015) have individually been shown to affect decision making. To date the interaction between the variables has not yet been explored but could prove to be of significant benefit in policy making.

I investigate this by conducting two separate pre-registered lab-in-the-field experiments. The first 2019 experiment ('Experiment 1') was conducted in Kampala (Uganda) with ethnic Gisus. This experiment applies the Krupka and Weber (2013) norm elicitation methodology to measure within subject norms and expectations across hypothetical geographical contexts ('place'). I also randomise the language of the sessions amongst bilinguals using the extensively adopted one-shot version of the Public Goods Game to measure cooperation (Zelmer, 2003; Cardenas and Carpenter, 2008; Levitt and List, 2007). First, I measure whether there are significant differences by norms. Second, I test whether expectations differ across each context. Third, I test the interaction between context and language, a proxy for cultural and identity factors. Fourth, I test whether differences in norms/expectations by place can be explained by cultural exposure, being the amount of time subjects, who all originate from Bugisu region, have spent within Kampala.

The second 2024 experiment ('Experiment 2') is conducted in Mbale (Uganda) and also features Gisu

participants. This experiment applies two treatment frames to the standard one-shot Public Goods Game, the *Community Game* and the *Wall Street Game*, as used by Ellingsen et al. (2012). All participants play both games, though the first game in the treatment is 'actual', where participants are asked to play the game normally. The second is 'hypothetical', where participants are asked to imagine what their responses would have been under the different frame in an incentive compatible design. Others have examined whether data is affected by eliciting norms for oneself or a different group (d'Adda et al., 2016), but not whether multiple norms can sensibly be elicited in one sitting. For this experiment contributions, expectations and norms are all measured across the two treatments. First I test whether contributions within treatment differ, and whether hypothetical contributions can predict actual contributions for each frame. I then repeat these tests for expectations and norms. The main contribution of this chapter is methodological.

Theories of cultural integration have an established heritage within the social sciences, broadly summarised as assimilation theory, structuralism and multiculturalism (for a review see Algan et al. (2012)). Fundamentally, these theories attempt to explain how cultural integration reflects learning about (and adherence to) a new set of social norms, or a change in one's own identity and norms. While typically applied to international migration, these theories are arguably equally pertinent to national rural-to-urban migration. There are significant pressures on all migrants to assimilate as failing to coordinate is inherently costly, as is ostracism from a social group (for summary see Edgerton (2010)). This is consistent with psychological theories which emphasise the importance of peer effects in socialisation, which may mean that migration (or a new set of peers) could result in an individual adapting their preferences substantially (Bisin and Verdier, 2001). However, exposure and time are needed to learn and adhere to new norms. Previous studies have shown that one year is sufficient for behaviours to start to change (Cameron et al., 2015; Berge et al., 2018). This experiment measures actual and hypothetical norms and expectations across two geographies, and uses subjects which have migrated between them. This allows an expansion of the existing literature by investigating whether national migration is equivalent to international migration, or whether 'linkeages' substantially alter this relationship (Stites et al., 2014; Gibson et al., 2020; Clist and Hong, 2023).

From Experiment 1 I have four main results. First, I show that norms are statistically different by place, with it being more socially acceptable to contribute less to the common pot when hypothetically in Mbale, as opposed to (actually being in) Kampala. This is consistent with the anthropological literature. Second, there is little evidence to support that expectations vary by place. Third, norms in Kampala are similar in both languages, but there is a language-treatment-effect in Mbale of 5.5%. However, the interaction for expectations is small and insignificant. Fourth, while norms do differ by place, cultural exposure does not appear to explain this; either Kampala norms are learnt within a few months or they were known before migration.

From Experiment 2 I have three main results. First, for contributions, there is insufficient evidence to reject the null that actual and hypothetical responses are equal for each game (i.e. within game but across treatments). These findings imply that participants are both able to identify different levels of contributions by game, and that hypothetical

responses are a statistically accurate proxy for actual contributions. Second, for norms, there is insufficient evidence to reject the null that the populations are equal when comparing actual and hypothetical norms within game frames. Third, this is not true for expectations which do not significantly, suggesting that distinct distributions of expectations do not exist by game/frame.

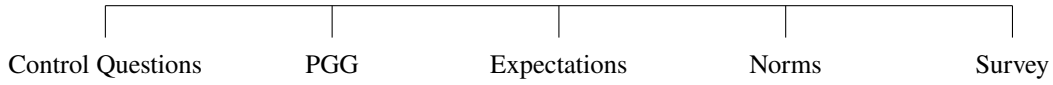
The paper proceeds as follows. Section 5.2 sets out the empirical overview, section 5.3 presents the research design, sections 5.4 and 5.5 present the results for experiment 1, and section 5.6 the results for Experiment 2. Section 5.7 concludes.

5.2 Empirical Overview

5.2.1 Overview of Experiments

This chapter contains two separate experiments, Experiment 1 (conducted in Kampala in 2019) and Experiment 2 (conducted in Mbale in 2024). The phases of Experiment 1 are summarised in Figure 5.1. Experiment 1 was conducted at universities in Kampala, Uganda's capital city, with students who had moved from the Bugisu region to study. Mbale is the largest city within the Bugisu region. During recruitment (and again prior to the experiment) all participants were screened to ensure that they were fluent in both Luganda and Lugisu, where Luganda is a semi-official national language, and Lugisu the tribal language of the Gisu. It was explained to participants that they would be playing a simple one shot Public Goods Game in anonymous pairs, with an initial endowment of UGX 12,000 and a 1.5 multiplier. All contributions to the common pot would be divided equally. The language of the sessions was randomised and conducted in either Luganda or Lugisu.

Figure 5.1: Experiment 1 (2019 Kampala) Design



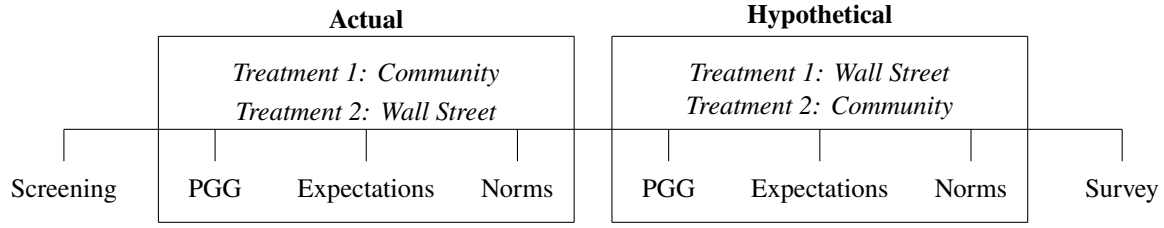
Control questions were asked to ensure that participants understood the rules of the Public Goods Game, after which participants played the game. Contributions to the common pot were set in intervals of UGX 3,000, between UGX 0 and 12,000. After the game, I measured expectations by asking participants "in intervals of 10%, how many people in this room do you think contributed the amounts shown". Next, norms were measured using the Krupka and Weber (2013) methodology. Each of the five contribution levels (UGX 0, 3,000, 6,000, 9,000 and 12,000) were ranked in terms of social acceptability on a four point scale: '++' very socially acceptable, '+' socially acceptable, '-' socially unacceptable and '- -' very socially unacceptable. These steps measure actual expectations and norms in Kampala

respectively, and both are incentivised with a bonus payment of UGX 4,000.

At this point participants were then told "we previously played this game with over one hundred, randomly selected people in Nakaloke sub-county, near Mbale". They were then asked "in intervals of 10%, how many people in Mbale do you think contributed the amounts shown". This measures the hypothetical expectations in Mbale. Next, participants were askd to "imagine that instead of playing in Kampala, we are playing in Mbale. Everyone in this room is also playing the game, and that everything else is the same. The only difference is that we are imagining playing the game in Mbale. Please rate the same 5 choices as either very socially unacceptable (–), somewhat socially unacceptable (-), somewhat socially acceptable (+) or very socially acceptable (++)". This measured the hypothetical norms in Mbale. 'Place' refers to either Kampala or Mbale, where norms and expectations in Kampala are termed 'actual' and those elicited for Mbale 'hypothetical'.

The Experiment 2 design is set out in Figure 5.2 and again is centred around an anonymous one-shot two player Public Goods Game, with a 1.5 multiplier of the common pot which is then split equally (though with a smaller initial endowment of UGX 8,000). The experiment is instead conducted in Mbale, in the Bugisu region, though again recruits Gisu participants. Rather than varying 'place' hypothetically, this experiment varies the frame applied. There are two treatments, the *Community Game* and the *Wall Street Game*. Similar to Experiment 1, this experiment pre-screens participants to ensure that the Public Game is fully understood. This design does not include a language element and thus all experiments are conducted in Lugisu.

Figure 5.2: Experiment 2 (2024 Mbale) Design



In Treatment 1, the Community Game is played first, with possible contributions to the common pot increasing in multiples of UGX 2,000 (i.e. UGX 0, 2,000, 4,000, 6,000 or 8,000). Participants are asked to select their contribution level. Next expectations and norms are measured in the same way as Experiment 1, with responses incentivised with a UGX 3,000 potential bonus. These responses are termed 'actual'. At this point, participants are told the following: "now imagine that you have just walked into this room. Instead of being told that you are going to be playing the Community Game we told you that you will be playing the Wall Street Game. All of the conditions and rules are exactly as before. Take a moment to visualise this. Please mark what you would have contributed to the common pot in the Wall Street Game". Participants then provide a 'hypothetical' response to the Wall Street Game frame. In same

way, participants are then asked to re-answer the expectations and norms questions by imagining that they are playing under the Wall Street Frame. These expectations and norms are 'hypothetical'.

In Treatment 2, everything is identical to Treatment 1, except that the Wall Street Game is played first (i.e. is the 'actual') and the Community Game is played second (being the 'hypothetical'). 'Actual' is always the game frame played first, and 'hypothetical' the second frame applied. Full details of the experimental designs are included in section 5.3.

5.2.2 Theoretical Framework and Model

I do not propose a new framework for this chapter, as the supporting theory has been well established within the existing literature. Instead, I propose a simple vignette where individuals have a vector of social acceptability ratings (with accompanying expectations) for a given set of possible actions in a set of scenarios. For this framework, social norm sets are discrete distributions of norms and expectations activated by a specific environment, culture, frame, language, identity, narrative or combination thereof. To illustrate this, an individual will have a different set of norms and expectations when deciding how much to tip in hamburger restaurant in Tokyo compared to New York. In the former, tipping is considered a social faux pas, where in the latter the opposite is true. Thus, the same individual will have a different distribution of social acceptabilities for norms (i.e. a different Krupka and Weber (2013) distribution of social acceptabilities), as well as a different distribution of expectations for each context.

An individual will have a default social norm set for all scenarios, being the norms and expectations learnt as a child in their community of origin. An individual will only develop multiple sets where exposure (time spent or the number of interchanges) to influencing factors increases. For example, an American child may have been taught to tip a waiter 20% in their community of origin (let's assume this is New York) when growing up. This will form the child's worldview until they are exposed to alternative social norms directly through social interactions, or are taught about them hypothetically (i.e. without the direct experience). Let's assume, when grown, the individual decides to learn Japanese, becoming fluent and thus bilingual. The individual likes to practice their Japanese at local Japanese restaurants, but still believes it is most socially acceptable to, and expects the majority of people will, tip 20% as these restaurants are in New York. Subsequently, the individual travels to Tokyo without an in-depth knowledge of the cultural differences regarding tipping. It is not known until the individual is exposed to the new norm in context, makes errors and subsequently adjusts their behaviour. At this point, a new set of social norms becomes distinct, as the individual has had sufficient exposure to identify that both norms and expectations differ in the new context. Physical exposure isn't necessarily required, moreover exposure can be conceptualised as any experience which increases knowledge of differing cultures (for example a travel guide could have informed the individual, at least in part, of the cultural difference).

This example is obviously highly simplified; language embodies culture and, in reality, it would be almost impossible for the individual to learn Japanese without also gaining insights into Japanese culture. Additionally, the establishment of new social norm sets is likely gradual, with accuracy improving with increased exposure. This increases the likelihood that there may be an interaction between language and context. For example, when the individual returns to New York, they may feel it is less appropriate to tip 20% when speaking Japanese in a Japanese restaurant, now that the social norms of Japanese culture are better understood. However, if the individual speaks English, the original social norms may become 'activated', resulting in the 20% tip becoming the most socially appropriate. Further, it may be that this distinction is not relevant in New York (i.e. irrespective of language, the individual feels the most socially acceptable action is to tip 20%), but that it is in Toyko. In other words, the individual may believe that it is socially unacceptable to tip if speaking Japanese in Tokyo, but socially acceptable to tip if speaking English in Tokyo.

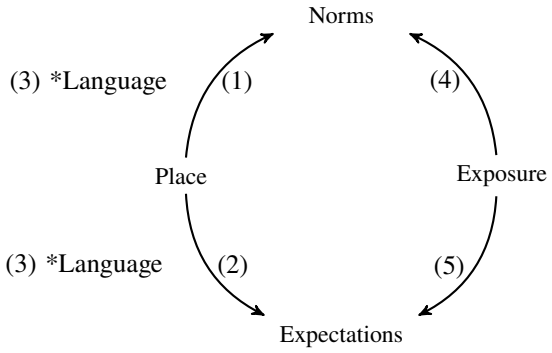
This framework is informed by two key strands within the literature. The first strand is evolutionary game theory, which provides models which help to explain norm dynamics by reference to how individuals update their expectations (see Young (2015) for review). These models reveal that norm dynamics are inherently stochastic, even when individuals share socioeconomic characteristics. Social pressure, symbolism and bench-marking are key mechanisms in supporting norms, but there are inevitably 'tipping points' which lead to a global diversity of norms, despite local conformity. The second strand is summarised by Hoff and Stiglitz (2016), who note the importance of exposure to environment in shaping the behaviour of encultured actors. Specifically, social context creates the set of mental models available to the individual, and affects the scenarios that prime alternative mental models. The key link here is that individuals are expected to have multiple sets of social norms by context (which can be triggered by a frame), as social norms evolve differently in each context, and individuals are exposed and have an awareness of these different contexts, which remain consistent, irrespective of whether or not they are actualised or hypothesised.

The primary contribution of this study is to the empirical experimental literature, namely through the use of hypothetical conditions to elicit social norms, both by reference to a physical geography (place) or by use of a frame (Community / Wall Street Game). This involves asking participants to rate actions 'as if' they and the others in their session were operating in a different context, or under a different frame. This will be achieved by introducing different places (Kampala/Mbale - Experiment 1) and frames (Community/Wall Street Game - Experiment 2) in the context of a single shot Public Goods Game (Liberman et al., 2004; Zelmer, 2003; Cardenas and Carpenter, 2008; Levitt and List, 2007).

5.2.3 Experiment 1 (2019 Kampala) - Empirical Approach

All analysis was pre-registered (Clist and Hill, 2019). Within each family of hypotheses, I control for the number of tests conducted, controlling the FWER using the Hochberg procedure implemented in Stata by Newson (2010). I adopt the 5% significance level for all decisions, using the appropriate q values. Figure 5.3 summarises the testing families for my FWER approach and empirical tests.

Figure 5.3: FWER and Research Questions



Family 1 simply tests whether norms differ by place in a public goods game where norms are elicited using a modified version of the Krupka and Weber (2013) methodology. The innovation involves asking subjects to elicit in-session and hypothetical norms of a separate geographical location ('place'). In terms of analysis, the standard approach (see Chang et al. (2019)) has been to conduct a series of rank sum tests, to see whether the norm ratings differ between two groups. A concern here is that the series of tests mean that false positive results are made more likely. As such I follow the standard approach, but control for five tests. In translating a norm rating to a numerical scale I follow Krupka and Weber (2013) in coding the norm ratings as -1, -1/3, 1/3 and 1.

The approach for family 2 mirrors that of family 1. Instead of norms, family 2 asks whether expectations differ by place, where expectations are recorded in multiples of 10%. As highlighted, I follow Bicchieri (2010)'s approach in distinguishing the components of social norms into empirical expectations ('expectations' - family 2) and normative beliefs ('norms' - family 1). As these two components are distinct they form two families, each with one corrected p value for the five individual Wilcoxon rank-sum tests.

The separate effects of language on norms and expectations has been explored in chapter 4. Family 3 is concerned with the interaction *between* place and language and the effects on norms or expectations. To explore this I use the following regressions:

$$\lambda_{ic} = \alpha_{ic} + \beta_1 Language_{ic} + \beta_2 Place_{ic} + \beta_3 Language_{ic} * Place_{ic} + \epsilon_{ic} \tag{5.1}$$

$$\theta_{ic} = \alpha_{ic} + \beta_1 Language_{ic} + \beta_2 Place_{ic} + \beta_3 Language_{ic} * Place_{ic} + \epsilon_{ic} \tag{5.2}$$

where λ_{ic} represents norms on a four-point scale between -1 and 1 and θ_{ic} represents expectations on a percentage scale, in 10% intervals. Throughout, i indexes an individual and c the contribution level between nil and 12,000, in 3,000 intervals. $Language_{ic}$ is a dummy variable for Luganda and Lugisu and $Place_{ic}$ a second dummy variable for Kampala and Mbale. $Language_{ic} * Place_{ic}$ is the interaction effect. While norms and expectations are distinct the interaction between place and language is theoretical focus of family 3. Thus, the above presents two p values for correction in line with my FWER strategy.

Family 4 looks to test whether exposure can be conceptualised as learning about norms. In order to maintain power, I will compare the average norms of the 1/3 most recent arrivals to Kampala with the 2/3 least recent arrivals. This is consistent with the pre-registered design (Clist and Hill, 2019). In this case time in Kampala is a proxy for exposure. This comparison will be conducted at each contribution level using Wilcoxon rank-sum tests (as with families 1 and 2), before controlling for multiple tests as per my FWER strategy. Family 5 adopts the same approach as family 4, but instead compares the average expectations of the 1/3 most recent arrivals to Kampala against the 2/3 least.

5.2.4 Experiment 2 (2024 Mbale) - Empirical Approach

Hypotheses and Family-Wise Error Rate (FWER) strategy

All analysis was pre-registered (Clist and Hill, 2024). The family groups have been set out below, referencing the specific hypothesis to which they relate. Each group (contributions, norms and expectations) is comprised of two families. The specifications for contributions and norms are singular in nature, and thus I do not propose any Family-Wise Error Rate correction. For expectations analysis is conducted by level and so, for each family of hypotheses, I will control for the number of tests conducted using the Romano-Wolf stepdown p-values for multiple hypothesis testing procedure implemented in Stata by Clarke (2021). I will use the 5% significance level for all decisions, using the appropriate q values where applicable.

For each group (contributions, norms and expectations) there are two separate hypotheses:

H1: Actual & Hypothetical contributions (H1.1), expectations (H2.1) and norms (H2.1) for Wall Street Game (pooling treatment 1 & 2).

- **Family 1:** Test within a regression framework such that H_0 is that contributions (1.1), expectations (2.1) or

norms (3.1) for both the hypothetical and actual Wall Street Game are equal against H_a that they are not.

H2: Actual & Hypothetical contributions (H1.2), expectations (H2.2) and norms (H3.2) for Community Game (pooling treatment 1 & 2).

- **Family 2:** Test within a regression framework such that H_0 is that contributions (1.2), expectations (2.2) or norms (3.2) for both the hypothetical and actual Community Game are equal against H_a that they are not.

Group 1: Contributions

The first family of tests explores whether differences exist between hypothetical and actual contributions across treatments, but within each game: Wall Street Game (*H1.1*) and Community Game (*H1.2*). Throughout, contributions refer to either actual or hypothetical contributions made to the common pot in the Public Goods Game, between UGX 0 and UGX 8,000, in multiples of 2,000.

Families 1.1 & 1.2: Actual vs Hypothetical: Wall Street & Community Game (between treatments)

The specification will be a regression, with β_1 being the coefficient of interest to test *H1.1* and *H1.2*. Only Wall Street (1.1) and Community Game (1.2) responses are included, using actual responses from treatment 1 and hypothetical responses from treatment 2 for *H1.1*, and the opposite for *H1.2*. A significant result would mean that the H_0 is rejected, meaning there is evidence to support H_a that actual and hypothetical contributions do differ for the Wall Street or Community Game. The regression will use the basic set up, where all variables are as previously defined:

$$Y_i = \alpha + \beta_1 ActHyp_i + controls + \epsilon_i \tag{5.3}$$

where Y_i is the contribution level in multiples of UGX 2,000 to the common pot in the PGG, by participant i. $ActHyp_i$ is coded 0 = actual and 1 = hypothetical. Controls will comprise of age and gender.

Group 2: Expectations

The second family of tests explores whether differences exist between hypothetical and actual expectations across treatments, but within each game: Wall Street Game (*H2.1*) and Community Game (*H2.2*). Throughout expectations are between 0% and 100%, with the percentage representing the expected in session contribution made at each of the five contribution levels (UGX 0 - UGX 8,000, in UGX 2,000 intervals). Thus, for each participant, there are ten

responses (five for actual and five for hypothetical). As expectations are measured as the percentage of participants in that session who are expected to contribute at the five contribution levels, within a treatment, the sum of actual and hypothetical responses will total 100% a piece.

Families 2.1 & 2.2: Actual vs Hypothetical: Wall Street & Community Game (between treatments)

The specification will be a regression, with β_1 being the coefficient of interest to test *H2.1* and *H2.2*. Note, for this specification, only Wall Street (2.1) and Community Game (2.2) expectations are being included, using actual responses from treatment 1 and hypothetical responses from treatment 2 for *H2.1*, and the opposite for *H2.2*. A significant result would mean that the H_0 is rejected, meaning there is evidence to support H_a that actual and hypothetical expectations do differ for the Wall Street or Community Game. Standard errors will be clustered at the participant level, as each participant provides five actual and five hypothetical responses for expectations. To avoid issues of collinearity, regressions will be run at each contribution level, meaning five regressions per family.

$$X_i = \alpha + \beta_1 ActHyp_i + controls + \epsilon_i \tag{5.4}$$

where X_i is the expectation between 0% and 100%, by participant i, with data long. $ActHyp_i$ is coded 0 = actual and 1 = hypothetical. Controls will comprise of age and gender.

Group 3: Norms

The third family of tests explores whether differences exist between hypothetical and actual norms across treatments, but within each game: Wall Street Game (*H3.1*) and Community Game (*H3.2*). Throughout norms are coded on a four point scale, as set out in Krupka and Weber (2013): -1 for very socially unacceptable, -1/3 for socially unacceptable, 1/3 for socially acceptable and 1 for very socially acceptable. Thus, for each participant, there are ten responses (five for actual and five for hypothetical norms, all on the four point scale, across the five contribution levels).

Families 3.1 & 3.2: Actual vs Hypothetical: Wall Street & Community Game (between treatments)

The specification will be a regression, with β_1 being the coefficient of interest to test *H3.1* and *H3.2*. Note, for this specification, only Wall Street (3.1) or Community Game (3.3) norms are being included (though will be pooled at that level), using actual responses from treatment 1 and hypothetical responses from treatment 2 for *H3.1*, and the opposite for *H3.2*. A significant result would mean that the H_0 is rejected, meaning there is evidence to support H_a that actual

and hypothetical norms do differ for the Wall Street or Community Game. Standard errors will be clustered at the participant level, as each participant provides five actual and five hypothetical responses for norms. The regression will use the basic set up:

$$Z_i = \alpha + \beta_1 ActHyp_i + \beta_2 Level_i + controls + \epsilon_i \tag{5.5}$$

where Z_i is norms coded between -1 and 1, in 2/3 intervals as set out in Krupka and Weber (2013), by participant i, with data long. $ActHyp_i$ is coded 0 = actual and 1 = hypothetical. $Level_i$ is coded 0 = UGX 0, 1 = UGX 2,000, 2 = UGX 4,000, 3 = UGX 6,000 and 4 = UGX 8,000. Controls will comprise of age and gender.

5.3 Research Design

5.3.1 Experiment 1 (2019 Kampala)

This section outlines the experiment, as summarised in figure 5.1. A full experimental script is included in appendix 1. Once subjects enter the room, they hear an explanation in the randomly allocated language of that session.

Location, Sample and Language Screening

I follow Clist and Verschoor (2017) by limiting the sample to those that speak Lugisu and Luganda. I depart by conducting the experiment in Uganda's capital Kampala, as opposed to rural areas in Bugisu (the homelands of the Gisu - see figure 5.4). There are clear socio-economic differences between Mbale (within the Bugisu sub-region) and Kampala, and the evolutionary trajectories of the social norms within these regions is likely to have varied substantially. Ethnographic literature documents that the Gisu are highly autonomous and individualistic (Heald, 1989). In the Bugisu region, land shortage and resource scarcity give rise to competition amongst male Gisu kin, who compete for early partible inheritance as they approach manhood (Hargreaves Heap et al., 2012). This apportionment of land is a key source of conflict and shapes the economic prosperity of the next generation. In this context, it is unsurprising that the primary reason cited for migration from rural eastern Uganda to Kampala is economic, with the majority of migrants being under 25 years old (Mukwaya and Bamutaze, 2012).

Figure 5.4: Regions of Uganda

Source: https://dhsprogram.com/pubs/pdf/SR245/SR245.pdf accessed 28.01.2022

Greater economic opportunities exist in urban areas such as Kamapala relative to the largely rural Bugisu region. Even Mbale, a small city of circa 70,000 inhabitants, will present significantly fewer economic opportunities when compared to the capital city. Thus, migrants to Kampala are less reliant on inheritance for economic independence and therefore are likely to have different psychological associations relative to the Bugisu region. For this reason I sample students who have migrated from Bugisu region to Kampala.

The experiment is run across 14 sessions, with an average size of 25.7 subjects. Each session is randomly assigned one of the two languages (Luganda or Lugisu) and there is an equal number of sessions for each language. Luganda and Lugisu are both Bantoid languages in Uganda. In terms of relatedness, they are as alike as English and German, or French and Spanish (Hammarström et al., 2021). Luganda was originally spoken by the Ganda people of central Uganda, whilst Lugisu was spoken by the Gisu people of eastern Uganda (also known as 'Masaba' - see figure 5.5). Both are spoken at home by over a 750,000 people, with Luganda an offical language in southern Uganda. Participants are ethnically Gisu (or have a strong connection to the area, e.g. through their mother) but live in Kampala where Luganda is the Lingua Franca. Additionally as university students also all speak English.

Figure 5.5: Language Groups of Uganda

Source: https://commons.wikimedia.org/wiki/ accessed 08.11.2019

A constant set of facilitators deliver a script that has been carefully translated and backtranslated to ensure consistency. All facilitator roles are maintained across both session languages to ensure comparability, though all are proficient in both languages. The assigned session language is used in *all* communication of that session (between facilitators, and from the first interaction to the last). The sample of 360 comprises 200 over 8 sessions from Makerere University and 160 over 6 sessions from Kyambogo University. Both have a sizeable Gisu population. Subjects are pre-screened before the experiment to gauge their ability to partake in the experiment (in either language).

The experiment is conducted by the Field Lab (https://thefieldlabuganda.com), which is based in Mbale (a city in the Bugisu region), naturally aligning with the target population. Facilitators explain this when delivering the script and highlight that the experiment is interested in how people that are Gisu, or from the Bugisu region, make decisions. Though English is the national language of Uganda, Luganda is the primary language spoken in the Buganda region, which includes Kampala. Local informants suggest that if a subject in Kampala speaks Lugisu, they will also

be fluent in Luganda. This was confirmed during the pilot. Lugisu is the language spoken at home by 84% of the sample though all reside in Kampala. As such, all recruitment activities are in Lugisu.

To recruit ethnically Gisu subjects (or subjects who have a strong link to the area) I initially use the semi-formal tribal social groups. These groups maintain membership lists which represent the most accurate data of the local ethnic Gisu population for each campus. Willing subjects register in advance and are randomly allocated to a session/language. From the survey data, 96% of subjects were born in the Bugisu region (111 from urban Mbale and 236 from rural areas). The remaining 4% (13) identify as being ethnically Gisu.

Controls and Public Goods Game (PGG)

I employ a standard linear one-shot two player PGG, which captures each participant's preference for cooperation. Subjects must trade off their own payoff and the social benefit, which is explained to subjects as a choice of how much to allocate to the 'common basket' and 'private envelope'. Both individuals in the pair are endowed with 12,000 Ugandan Shillings. At the time of the experiment this is equivalent to around £2.70 or $3.25, about two days income for an average Ugandan. The experiment lasts approximately an hour. Player 1's payoff (V_1) is calculated as:

$$V_1 = 12,000 - C_1 + 0.75(C_1 + C_2) \tag{5.6}$$

where C_1 is their contribution to the joint fund and C_2 is the contribution from player 2. Each player can earn between 9,000 and 21,000 Ugandan Shillings, with the minimum and maximum pay out per pair totalling 24,000 and 36,000 Ugandan Shillings respectively. The 'common basket' represents the group fund, $(C_1 + C_2)$ multiplied by 1.5. 'Private envelopes' represent the amount of the initial endowment retained (12,000 — C_1).

In order to ensure subjects have fully understood instructions, the three corner solutions are used as the basis for control questions (C_1/C_2: 12,000/12,000, nil/nil and 12,000/nil). The first two control questions are solved jointly to aid understanding, and the third individually. Eleven subjects (3%) answered less than 30% of the third control question steps correctly, and are excluded from the analysis (as outlined in the pre-analysis plan). On completion of the control questions, subjects make their contribution decision: splitting 12,000 between their private envelope and a common basket (in an increment of 3,000).

Expectations, norms (by place) and survey

After the contribution selection, the expectations and norms for each contribution level are measured. Subjects are incentivised in both instances, being able to earn a bonus 4,000 Ugandan Shillings if their answer is correct. Before answering the survey, a random question from the expectation/norm sections is selected to determine which question is used for payment.

For the expectation question, subjects are asked to guess the percentage of people in their session that gave the various amounts (0 to 12,000 in 3,000 increments) to the common basket. Subjects earn the bonus if that question was selected and their guess was within 10% of the actual answer. For norms, subjects are asked to rank how socially acceptable each of the contributions are, using the now standard Krupka and Weber (2013) method. I employ a four-point scale; very socially unacceptable (- -), somewhat socially unacceptable (-), somewhat socially acceptable (+) and very socially acceptable (++). I use the whole distribution of possible choices. Subjects earn the bonus if they choose the most popular option. Note that while Krupka and Weber (2013) uses distinct samples, I follow d'Adda et al. (2016) and Erkut et al. (2015) in using the same sample for norms and contribution decisions, as they show an insignificant difference in outcome. Subjects are only eligible for this bonus if they only select one option for each scenario.

Subjects are also asked to complete an exit survey. Each question is read aloud, giving time for subjects to write their answers. Payment is made to subjects after completion of the survey and marks the end of the experiment. A full English script, answer sheet and survey questions, including encoding, are included in appendix 1.

Summary statistics

Table 5.1 reports summary statistics, including variables from the survey data. The median participant is a 22 year-old male, who was born in Mbale and learnt Lugisu first. They have since spent 4 years in Kampala, speak Lugisu at their home, and know 2 other people in the session. I see some variation in where people are from, though only 3% are from non-Bugisu homelands. I note less variation in language with 85% having learnt Lugisu first, and 84% speaking Lugisu in their current home. The sample has more men than women, though this is broadly in line with the latest data on gender inequality in tertiary education in Uganda.[1] The random allocation mostly worked well, but there is one sign of a lack of balance. Those allocated to Lugisu sessions previously spent 1.74 more years in Kampala than those allocated to Luganda sessions. This variable is transformed into a categorical variable to avoid the pull of outliers (as pre-registered), and there is also a lack of balance in the transformed variable. Essentially, the randomisation allocated more morning sessions to Luganda (5/7). Students that spent less time in Kampala were more likely to be able to

[1] World Bank data, available at `https://data.worldbank.org/indicator/SE.ENR.TERT.FM.ZS?locations=UG`, provides a gender parity index. The last observation is from 2014, and implies 42% of university students are female.

attend in the morning (presumably as they lived on campus), and so more recent arrivals ended up being statistically more likely to be allocated to a morning session. I was not aware of this, and had no control of this. The variable is only relevant for one pre-registered test which considers exposure. This is further discussed in 5.5, where all of the aggregated analysis on exposure across all specifications is insignificant.

Table 5.1: Summary Statistics, by Treatment

	Session Language			Balance Tests
Variable	Luganda	Lugisu	Total	*p*
Female	0.378 (0.486)	0.367 (0.483)	0.372 (0.484)	0.837
Age (capped at 30)	22.279 (2.084)	22.271 (2.322)	22.275 (2.205)	0.973
# people known in session	3.483 (3.841)	3.113 (3.021)	3.295 (3.449)	0.318
Years in Kampala, Raw	4.267 (3.384)	6.006 (5.178)	5.149 (4.466)	<0.001***
Years in Kampala, Transformed				0.017**
< 2.5 years	59 (34.3%)	41 (23.2%)	100 (28.7%)	
2.5-5.5 years	72 (41.9%)	72 (40.7%)	144 (41.3%)	
> 5.5 years	41 (23.8%)	64 (36.2%)	105 (30.1%)	
Main Language Used At Home				0.971
Lugisu	143 (83.1%)	149 (84.2%)	292 (83.7%)	
Luganda	11 (6.4%)	12 (6.8%)	23 (6.6%)	
English	15 (8.7%)	13 (7.3%)	28 (8.0%)	
Other	3 (1.7%)	3 (1.7%)	6 (1.7%)	
First Language Learnt				0.406
Lugisu	151 (87.8%)	145 (81.9%)	296 (84.8%)	
Luganda	7 (4.1%)	13 (7.3%)	20 (5.7%)	
English	8 (4.7%)	9 (5.1%)	17 (4.9%)	
Other	6 (3.5%)	10 (5.6%)	16 (4.6%)	
Birth Location				0.418
North Bugisu Region	53 (30.8%)	51 (28.8%)	104 (29.8%)	
South Bugisu Region	33 (19.2%)	48 (27.1%)	81 (23.2%)	
Central Bugisu Region	23 (13.4%)	18 (10.2%)	41 (11.7%)	
Mbale	58 (33.7%)	53 (29.9%)	111 (31.8%)	
Other	5 (2.9%)	7 (4.0%)	12 (3.4%)	
N	172 (49.3%)	177 (50.7%)	349 (100.0%)	

Note: The first four p values are for binary or continuous variables, and come from regressions. The second four come from Pearson tests. In each case *** denotes significance at 1%, ** at 5% and * at 10%. Standard deviations, or percentages, are provided in parenthesis. As pre-registered, I capped age at 30, which affects three people, aged 32, 33 and 47. The other languages used at home comprise Swahili, Lugere and Lusoga. The other languages first learnt include Swahili, Lugere, Lusoga, Luyankole, Ruqiso and Rutoro.

5.3.2 Experiment 2 (2024 Mbale)

This section sets out the key elements of the experiment, as summarised in Figure 5.2.

Screening & Sampling

This study was conducted using a lab-in-the-field approach, running 14 sessions with between 26 and 48 participants (average N=34). A total of 481 participants took part, 153 participants failed the control questions, leaving 328 to be included in the analysis. This is a much higher control failure rate than the 2019 experiment for two key reasons. The first is that the sample was recruited randomly from three rural parishes from near Mbale, rather than solely from universities in urban Kampala. Secondly, because of the lower expected mathematical competence, participants are only included in the analysis if 100% of control questions were answered correctly (compared to just 70% in the 2019 experiment). This experiment was conducted in the Bugisu Sub Region, Uganda, using the Gisu as the cultural group of focus. The Gisu are typified by an out-group bias and low in-group cooperation, which provides fertile opportunity to explore the affect of applying different frames in a Public Goods Game (Hargreaves Heap et al., 2012; D'Exelle and Verschoor, 2015).

Recruitment adopted the following process and was facilitated by The Field Lab ('TFL' - https://thefieldlabuganda.com), who have expertise and local connections in the region:

a) Identify a suitable research site and identify appropriate parishes (in the Bugisu Region).

b) Obtain a sampling frame (electoral register or similar).

c) Randomly select participants from the sampling frame.

d) Visit these with the help of the LC1 chairperson (a local elected leader) or somebody nominated by the LC1 chairperson and elicit willingness to participate.

e) If the selected individual is willing to participate TFL would provide the individual with the details of the lab-in-the-field session, including time and location.

f) If they are unwilling or unavailable to participate, then TFL would randomly replace them with somebody else on the sampling frame. All participants will be >18 years old and have multiple opportunities to express if they are unwilling/uncomfortable to participate. Though the local elected leader will support, they will not coerce participation; this will be solely the individual's choice. All participants will sign a written local language consent form prior to any participation, as well as having this explained verbally.

All subjects will need to be screened for language proficiency and to ensure they are an ethnic Gisu (a fundamental part of identity - see chapter 4). The following questions were asked prior to candidates being invited to participate:

1. Which tribe/ethnic group do you most identify with?

2. Where were you born?

3. Are you fluent in Lugisu?

Upon assessing this evidence, coupled with spoken language proficiency, enumerators then determined whether the individual is an ethnic Gisu.

Controls and 'Actual' Public Goods Game ('PGG') - Community [Wall Street] Game

A participation fee of UGX 1,500 was paid to participants irrespective of how many questions were answered by participants. This is in addition to any payout relating to the Public Goods Game (initial endowment UGX 8,000) and a potential bonus payout of UGX 3,000 for the norms and expectations elicitation.

I employ a standard linear one-shot two player PGG, which captures each participant's preference for cooperation, across two treatments: the 'Community Game' or the 'Wall Street Game'. All participants played both games, with the treatment determining the order. In treatment 1, participants played the Community game, providing responses to the PGG (contribution), expectations and norms. They were then asked to visualise instead that they are playing a Wall Street Game. They were then asked to re-answer the questions 'hypothetically'. In both cases norms and expectations are incentivised, attracting a potential UGX 3,000 bonus payout. Treatment 2 is identical, but participants started with the Wall Street Game, and then visualised the Community Game. Exploring whether the visualised 'hypothetical' treatments can be accurately predicted is the primary focus of this study.

Considering the mechanics of the PGG, participants must trade off their own payoff and the social benefit, which was explained to subjects as a choice of how much to allocate to the 'common basket' and 'private envelope'. Both individuals in the pair are endowed with 8,000 Ugandan Shillings (UGX). At the time of the experiment this is equivalent to around £1.70 or $2.20, about half a days income for an average Ugandan. The experiment lasted approximately an hour and a half. Player 1's payoff (V_1) is calculated as:

$$V_1 = 8,000 - C_1 + 0.75(C_1 + C_2) \tag{5.7}$$

where C_1 is their contribution to the joint fund and C_2 is the contribution from player 2. Each player can earn between 6,000 and 14,000 Ugandan Shillings, with the minimum and maximum pay out per pair totalling 16,000 and 24,000 Ugandan Shillings respectively. The 'common basket' represents the group fund, (C_1 + C_2) multiplied by 1.5. 'Private envelopes' represent the amount of the initial endowment retained (8,000 — C_1).

In order to ensure subjects have fully understood instructions, the three corner solutions were used as the

basis for control questions (C_1/C_2: 8,000/8,000, nil/nil and 8,000/nil). The first two control questions were solved jointly to aid understanding, and the third individually. Incorrect responses to the control question are excluded from the analysis. On completion of the control questions, subjects make their contribution decision: splitting 8,000 between their private envelope and a common basket (in an increment of 2,000).

'Actual' Expectations & Norms - Community [Wall Street] Game

After the contribution selection, the expectations and norms for each contribution level were measured. Subjects were incentivised in both instances, being able to earn a bonus 3,000 Ugandan Shillings if their answer was correct. Before answering the survey, a random question from the expectation/norm sections was selected to determine which question was used for payment. For the expectation question, subjects were asked to guess the percentage of people in their session that gave the various amounts (UGX 0 to UGX 8,000 in UGX 2,000 increments) to the common basket. Subjects earnt the bonus if that question was selected and their guess was within 10% of the actual answer. For norms, subjects were asked to rank how socially acceptable each of the contributions were, using the now standard Krupka and Weber (2013) method. I employ a four-point scale; very socially unacceptable (- -), somewhat socially unacceptable (-), somewhat socially acceptable (+) and very socially acceptable (++). Subjects earnt the bonus if they choose the most popular option. Note that while Krupka and Weber (2013) uses distinct samples, I follow d'Adda et al. (2016) and Erkut et al. (2015) in using the same sample for norms and contribution decisions, as they show an insignificant difference in outcome. Subjects are only eligible for this bonus if they only select one option for each scenario.

'Hypothetical' PGG, Expectations & Norms - Wall Street [Community] Game

This section repeats the actual game, but instead asked participants to visualise the alternative treatment, and to re-answer as though there were playing for the first time. Again all choices were incentivised. Participants were informed that either the actual or hypothetical PGG would be used to determine payouts. Additionally, the bonus question selected was to be across both actual and hypothetical expectations and norms; one of twenty questions selected.

Survey

Before answering the survey, a question was randomly selected by one of the participants to determine the bonus payouts. Subjects were then asked to complete an exit survey. The questions, as well as their intended use in the empirical analysis, are set in table 5.2. Each question was read aloud, giving time for subjects to write their answers. The entire experiment was conducted in Lugisu, the tribal language of the Gisu. Payment was made to subjects

after completion of the survey and marked the end of the experiment. A full English script, answer sheet and survey questions are included in appendix 3.

Table 5.2: Survey Questions

Question	**Nature**
How old are you?	Control variable
What is your gender?	Control variable
What is your highest education level?	Descriptive
Are you a member of any political party?	Descriptive
If so, which?	Descriptive

Note:'Nature' refers to the intended use of the data collected. 'Control variable' means that age or gender data will be used as a control in regressions. 'Descriptive' means that the data has been collected purely to provide summarised information about the sample.

Table 5.3 presents a range of summary statistics for Experiment 2, showing details for participants who passed the control questions. Of the 328 participants, 144 were males and 168 were female. The average age was 27.8 years, with the gender split being with 0.9 years. Only Age and $Female$ are used as controls across all specifications, the remaining data is informational only. Christianity was the primary religion across the participants (at 75%), with the remaining 25% identifying as Muslim. Political membership was broadly consistent across the genders, with 24% of all participants having a political party membership.

Table 5.3 shows that all variables, both those used as controls (age and gender) and other descriptive factors are balanced across the two treatments (being the exogenous variable across all specifications). Treatments only differ the order in which game frames are applied, with Treatment 1 applying the Community Game first, and then the Wall Street Game the second. Treatment 2 reverses this order.

Table 5.3: Summary Statistics - Experiment 2

	Treatment 1 $CG \rightarrow WSG$	Treatment 2 $WSG \rightarrow CG$	Total	Balance Test p
N	144 (43.9%)	184 (56.1%)	328 (100.0%)	
Age	28.694 (10.492)	27.038 (8.907)	27.765 (9.654)	0.123
# people known in session	4.424 (4.865)	3.967 (4.289)	4.168 (4.549)	0.368
Gender				0.726
Male	68 (47.2%)	76 (45.2%)	144 (46.2%)	
Female	76 (52.8%)	92 (54.8%)	168 (53.8%)	
Religion				0.123
Christian	102 (70.8%)	144 (78.3%)	246 (75.0%)	
Muslim	42 (29.2%)	40 (21.7%)	82 (25.0%)	
Education Level				0.315
Primary	23 (16.0%)	21 (11.4%)	44 (13.4%)	
Secondary	90 (62.5%)	113 (61.4%)	203 (61.9%)	
Tertiary	31 (21.5%)	50 (27.2%)	81 (24.7%)	
Political Membership				0.471
No	107 (74.3%)	143 (77.7%)	250 (76.2%)	
Yes	37 (25.7%)	41 (22.3%)	78 (23.8%)	

Note: Age, # people known in session and political membership p values are for binary or continuous variables, and come from regressions. The other three come from Pearson tests. In each case *** denotes significance at 1%, ** at 5% and * at 10% where relevant. Standard deviations, or percentages, are provided in parenthesis.

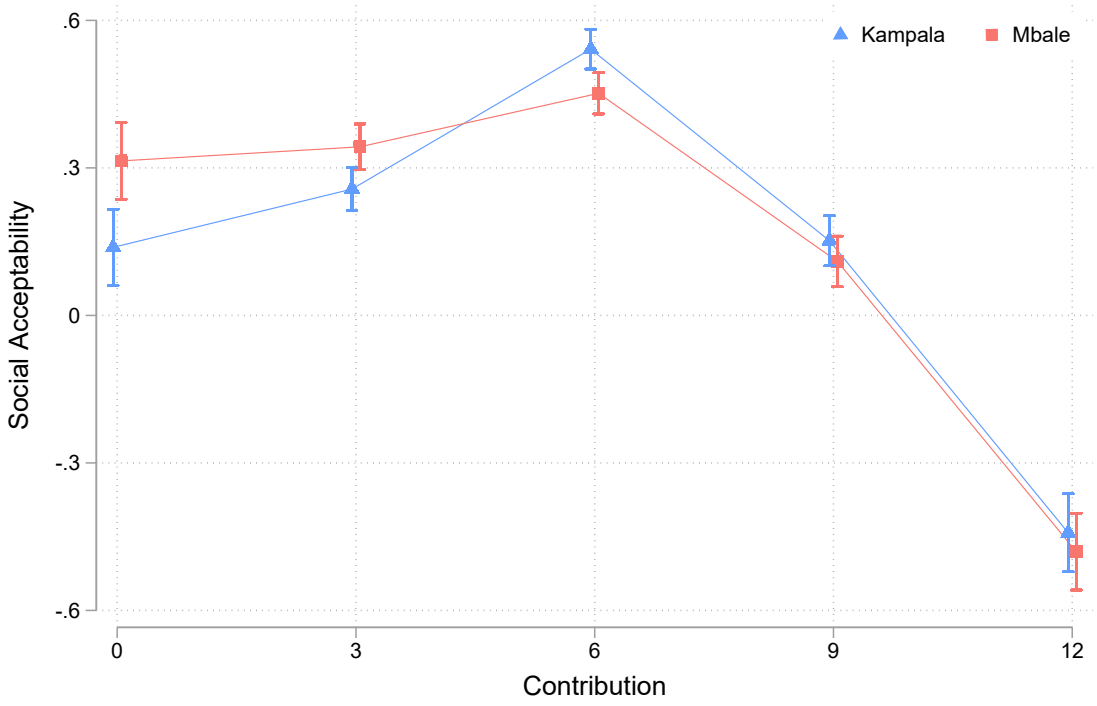
5.4 Results Experiment 1 - Social Norms by Place and Language

5.4.1 Norms by Place

Norms and separately expectations for both Kampala and Mbale were elicited by each subject. Both norms and expectations were elicited at the five contribution levels permitted when playing the Public Goods Game (where contributions to common pot could be either UGX 0, 3,000, 6,000, 9,000 or 12,000). The Krupka and Weber (2013) methodology incentivises subjects and acts as coordination device. For Kampala norms, subjects sought to coordinate the actual in-session norms (needing to correctly guess the session average in order to get the bonus payout). For Mbale, all subjects in the session were asked to imagine that they were playing in Mbale and coordinating on that basis in order to get the bonus payout. Thus the Mbale norms are hypothetical in nature.

From figure 5.6 there is a clear difference in the distribution of norms by place at the 0, 3,000 and 6,000 contribution levels. The difference is less pronounced at the 9,000 level and the averages are almost identical at 12,000. Table 5.4 presents this data numerically, showing that it is more socially acceptable to contribute lower amounts (0 and 3,000) to the common pot when hypothetically in Mbale than when in Kampala. At contribution level 6,000 (the focal point) this relationship switches. This is consistent with chapter 4 and Clist and Verschoor (2017). This is also the case at contribution levels 9,000 and 12,000, though less markedly.

Figure 5.6: Average Social Acceptability (Norms) by Contribution by Place



Note: 5% confidence intervals shown. Norms: 1 / -1 very socially acceptable/unacceptable, 1/3 / -1/3 somewhat socially acceptable/unacceptable.

Table 5.4 shows that there is a significant difference by place for norms (social acceptability), though only at contribution levels 0, 3,000 and 6,000 at the 1% level. After the application of the FWER strategy, only contribution level 0 survives multiple hypothesis correction at 1%, with 3,000 and 6,000 remaining significant at 5%. There is little evidence to suggest that contribution levels 9,000 and 12,000 differ by place.

Table 5.4: Norms (Social Acceptability) by Place

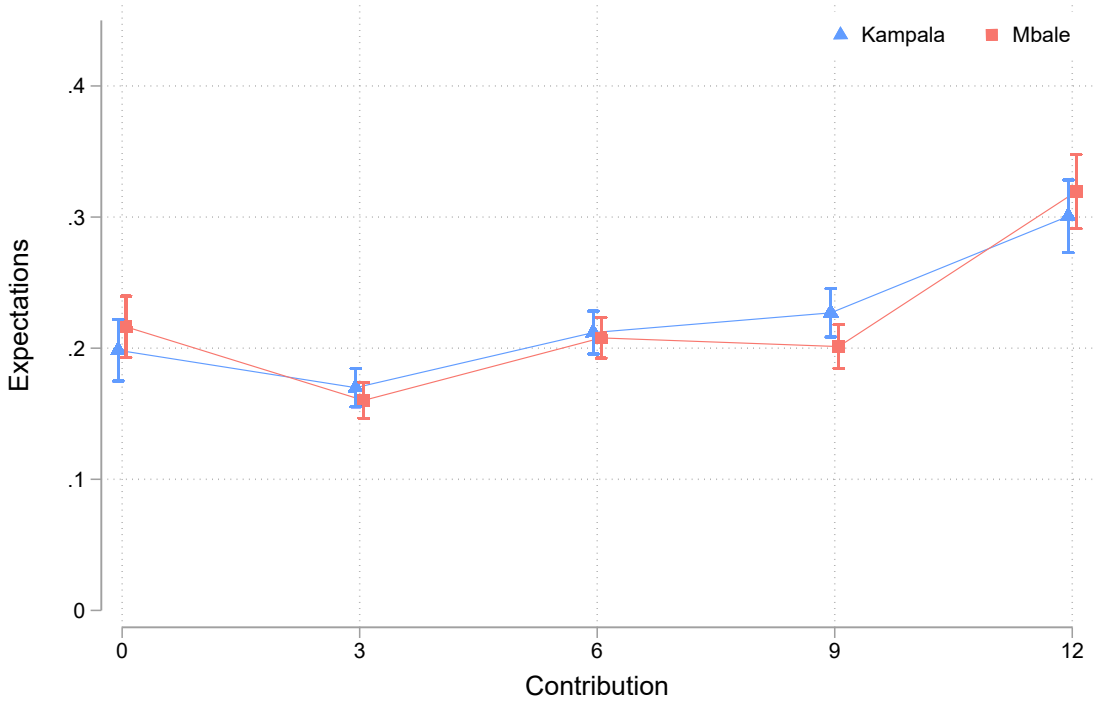
	Norm Rating										Test Stats.	
	Kampala					Mbale					Rank Sum	
Contribution	-1	-1/3	1/3	1		-1	-1/3	1/3	1		p value	q value
0	65	85	86	113	0.138	54	59	79	157	0.314	0.001***	0.005***
					(0.739)					(0.743)		
3,000	6	72	227	44	0.257	7	53	217	72	0.343	0.004***	0.01**
					(0.416)					(0.442)		
6,000	1	18	201	129	0.542	3	28	222	96	0.452	0.003***	0.01**
					(0.386)					(0.400)		
9,000	7	132	159	51	0.152	9	145	149	46	0.110	0.24	0.49
					(0.486)					(0.490)		
12,000	202	57	35	55	-0.442	209	63	22	55	-0.480	0.52	0.52
					(0.756)					(0.745)		

Note: Mean values and standard deviations (in parentheses) are shown. The q value refers to the multiple hypothesis corrected p-values, using the Hochberg procedure. *** p<.01, ** p<.05, * p<.1. N=349. Norms coded: 1 / -1 very socially acceptable/unacceptable, 1/3 / -1/3 somewhat socially acceptable/unacceptable.

5.4.2 Expectations by Place

From figure 5.7 it is clear that the distributions of expectations by place are more closely aligned. Expectations at the 3,000 and 6,000 are almost identical and table 5.5 shows that all contribution levels are close with the exception of 9,000 which is significant at the 5% level. However this does not survive the multiple hypothesis correction. All other contribution levels are insignificant and thus there is little evidence that expectations vary by place. While norms do vary by place, expected contributions do not, suggesting that norm adherence is different by place.

Figure 5.7: Average Expectations by Contribution by Place



Note: 5% confidence intervals shown. Expectations coded between 0 and 1 in 0.1 increments.

Table 5.5: Expectations by Place

	Kampala		Mbale		Test Stat.	
Contribution	Mean	sd	Mean	sd	p value	q value
0	0.198	0.214	0.216	0.213	0.21	0.78
3,000	0.170	0.134	0.160	0.125	0.39	0.78
6,000	0.212	0.149	0.208	0.140	0.92	0.92
9,000	0.227	0.168	0.201	0.152	0.05**	0.25
12,000	0.301	0.252	0.319	0.258	0.38	0.78

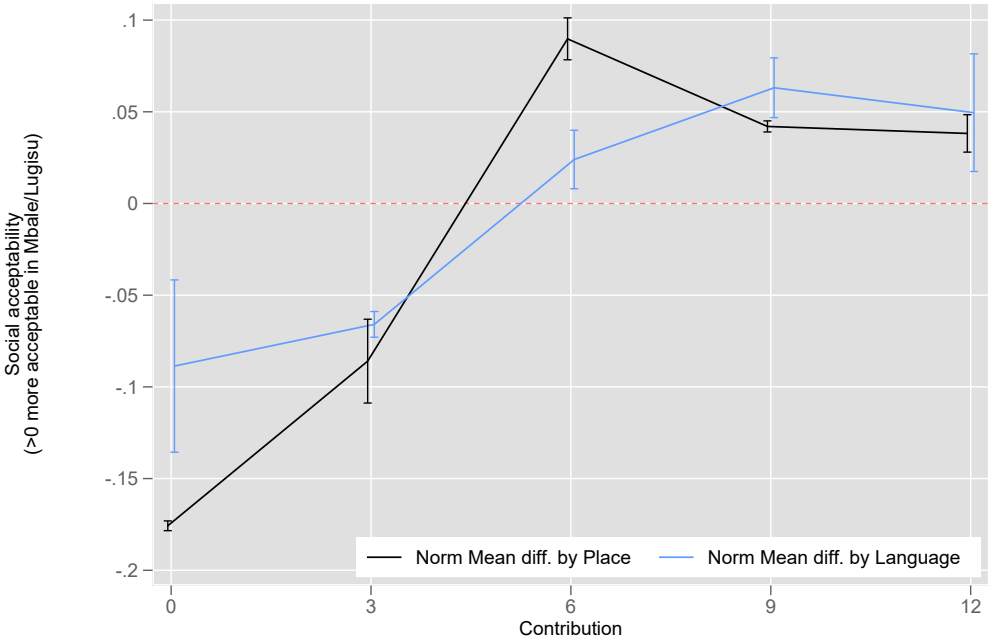
Note: Expectations are coded between 0 and 1 in 0.1 increments. The test statistic comes from a rank sum test. The q value refers to the Hochberg multiple hypothesis testing procedure. N=319 after exclusions for outliers.

5.4.3 Norms: Interactions Place & Language

Figure 5.8 shows the relative effect sizes of language and place on norms across the five contribution levels. Visually there is less variability in language than for place, particularly at the 0, 3,000 and 6,000 contribution levels, consistent

with table 5.4. In chapter 4 norms were not found to significantly vary by language.

Figure 5.8: Relative Effect Size on Norms (Social Acceptability): Place and Language

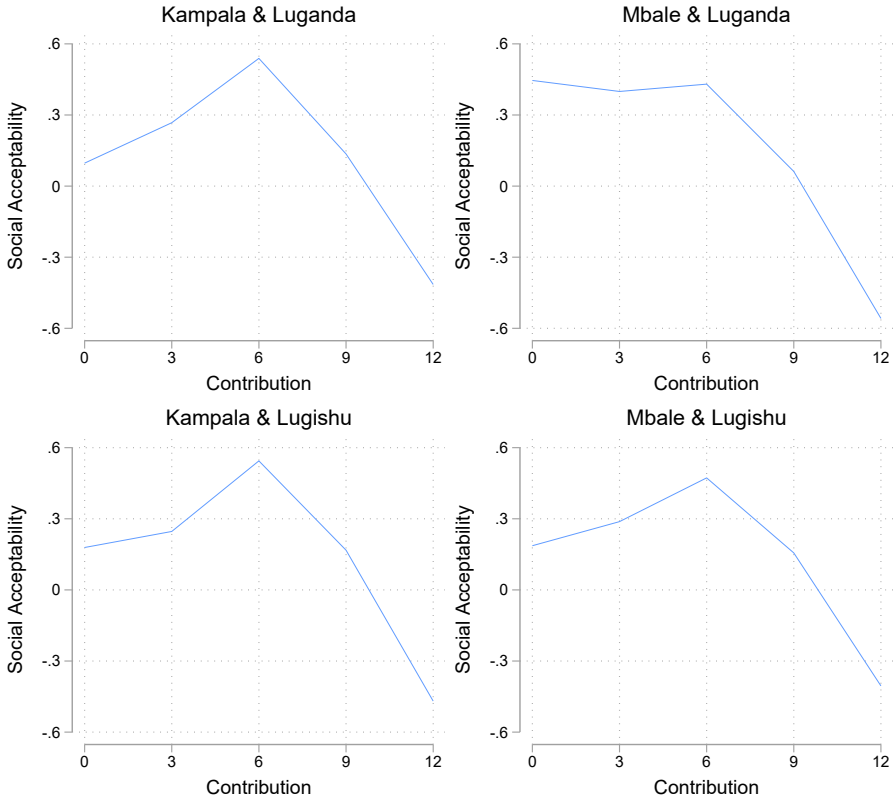


Note: 5% confidence intervals shown. Norms: 1 / -1 very socially acceptable/unacceptable, 1/3 / -1/3 somewhat socially acceptable/unacceptable. Differences calculated as Kampala/Luganda means less Mbale/Lugisu means. Negative difference in mean caused by a flip in the by place/language mean by contribution level. For example, table 5.4 shows Kampala > Mbale at contribution level 0, but Kampala < Mbale at contribution level 9.

Figure 5.9 presents a comparison of the the interactions between place and language on norms. This is also presented numerically in table 5.6. It is evident that the largest difference in the distributions is by place, consistent with figure 5.8. Language has very little impact on norms in Kampala, but there is a noticeable difference by language in Mbale.

Table 5.7 shows that contributions and Place*Contributions are significant at the 1% level. Contribution*Place*Language (being in Mbale and speaking Lugisu) is significant at the 5% level, both before and after the FWER Hochberg adjustment (q value). The dependent variable is norms coded 1 / -1 very socially acceptable / unacceptable, 1/3 / -1/3 somewhat socially acceptable / unacceptable. To interpret, when speaking Lugisu in Mbale, there is an increase of 0.111 in social acceptability relative to speaking Luganda in Kampala. Table 5.7 includes two specifications: Ref 1 is the regression as pre-registered, and Ref 2 which is an additional specification. As is evident from figure 5.6, the distribution of norms is non-linear and instead has a 'humped' shape. Reg 2 includes a *Contributions Squared* term to address this, which is significant at the 1% level. However, across both specifications, all coefficients of interest are consistently significant at the levels discussed.

Figure 5.9: Norm Interactions: Place and Language



Note: Interactions between place and language on norms shown. Presented numerically in table 5.6.

Table 5.6: Norm Interactions Summary

Contribution	Kampala/Luganda	Mbale/Luganda	Kampala/Lugisu	Mbale/Lugisu
0	0.097	0.446	0.179	0.186
3,000	0.267	0.339	0.247	0.288
6,000	0.539	0.430	0.544	0.473
9,000	0.136	0.062	0.168	0.156
12,000	(0.415)	(0.558)	(0.469)	(0.405)

Note: Interactions between place and language on norms shown. Presented visually in figure 5.9.

Table 5.7: Norms, Place and Language Interaction

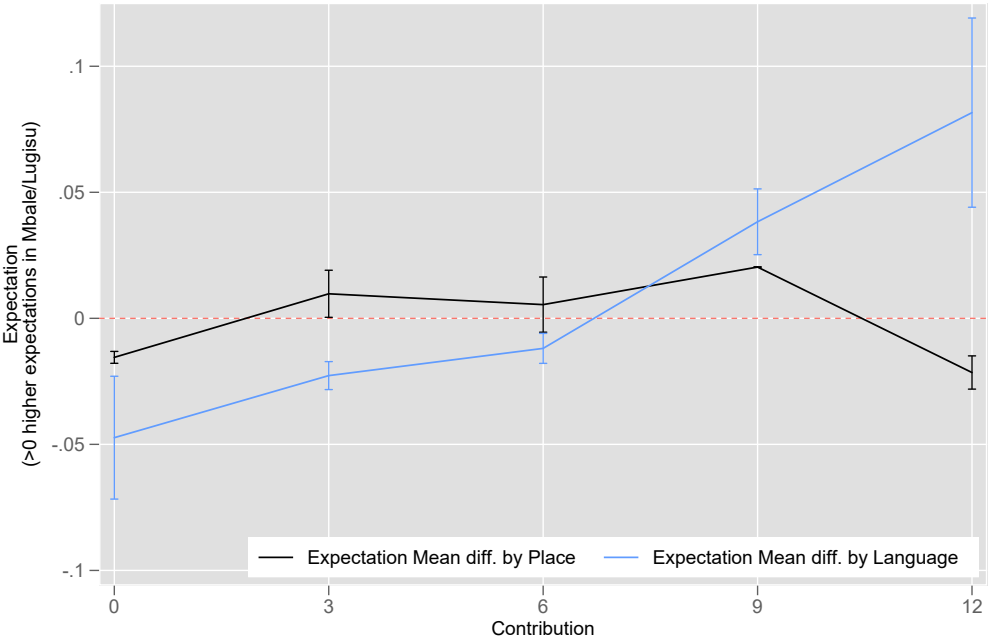
		Reg 1	*q* value	Reg 2	*q* value
Contribution		-0.049***		0.131***	
		(-9.90)		(13.12)	
Contribution Squared				-0.015***	
				(-20.98)	
Contribution*Place	(Mbale)	-0.010***		-0.010***	
		(-2.65)		(-2.65)	
Contribution*Language	(Lugisu)	-0.001		-0.001	
		(-0.35)		(-0.35)	
Contribution*Place*Language	(Mbale & Lugisu)	0.111**	0.028**	0.111**	0.028**
		(2.21)		(2.21)	
Constant		0.447***		0.179***	
		(16.07)		(5.66)	
N		349		349	

Note: For regression column, coefficients are shown with *t* statistics in parentheses. *** p<.01, ** p<.05, * p<.1. The dependent variable is norms. Clustered by subject id. Place is coded as 0=Kampala and 1=Mbale. Language is coded as 0=Luganda and 1= Lugisu. Reg 1 as pre-registered, Reg 2 is an additional specification to address non-linearity.

5.4.4 Expectations: Interactions Place & Language

Figure 5.10 shows the relative effect sizes of language and place on expectations across the five contribution levels. Visually there is less variability in place than for language, with the largest effect differences occurring at the 12,000 contribution level. Table 5.8 shows that contribution and Contribution*Language are statistically significant but have small coefficients. There is little evidence to support an interaction between place and language for expectations. As norms do differ by place but expectations do not, there is a clear difference between being aware of multiple sets of social norms and believing people will adhere to them.

Figure 5.10: Relative Effect Size on Expectations: Place and Language



Note: 5% confidence intervals shown. Expectations coded between 0 and 1 in 0.1 increments. Negative difference in mean caused by a flip in the by place/language mean by contribution level. For example, table 5.5 shows Kampala<Mbale at contribution level 0, but Kampala>Mbale at contribution level 9.

Table 5.8: Expectations, Place and Language Interaction

	Regression	q value
Contribution	0.007***	
	(4.78)	
Contribution*Place	-0.001	
	(-1.14)	
Contribution*Language	0.003***	
	(2.69)	
Contribution*Place*Language	0.002	0.148
	(1.45)	
Constant	0.170***	
	(20.96)	
Hochberg		0.148
N	319	319

Note: For regression column, coefficients are shown with *t* statistics in parentheses. *** p<.01, ** p<.05, * p<.1. The dependent variable is expectations. Clustered by subject id. Place is coded as 0=Kampala and 1=Mbale. Language is coded as 0=Luganda and 1=Lugisu.

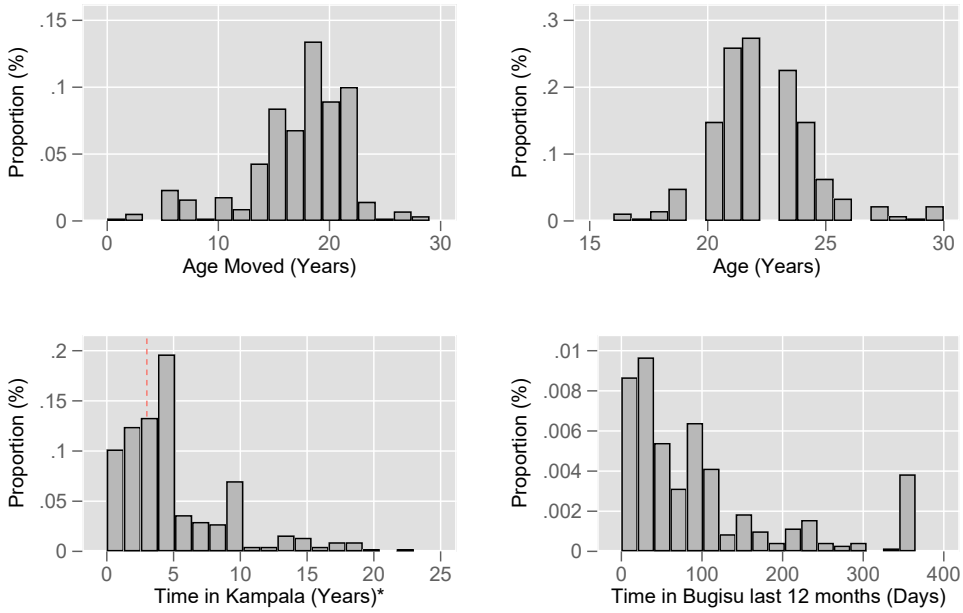
5.5 Results Experiment 1 - Social Norms by Exposure

5.5.1 Norms by Exposure

From section 5.4.1 norms differ by place significantly at the 0, 3,000 and 6,000 contribution levels. Section 5.4.2 shows that this is not true for expectations, meaning that subjects are aware that different norms exist across the two geographies, but they do not expect other Gisus in the session to adhere to those norms any differently. Subjects are aware that there are differences in social norms by place and have multiple sets of social norms. As all subjects are rural-to-urban migrants (having moved at some point in their lifetime from the Bugisu region to Kampala), it should be possible to detect the 'learning' of norms and expectations. Previous studies have shown that one year is sufficient for behaviours to start to change (Berge et al., 2018; Cameron et al., 2015).

Figure 5.11 provides a summary of the temporal statistics of the sample. The average age is 22.28 years and average time in Kampala 5.15 years. The average age at the time of migration to Kampala is 17.10 years.

Figure 5.11: Summary of Temporal Variables



Note: *1/3 vs 2/3 data cut shown. N=349. Data sourced from survey questions.

Table 5.9: Learning Norms: Kampala Norms by Time in Kampala

	1/3 Most Recent				2/3 Least Recent				Rank Sum	
Contribution	-1	-1/3	1/3	1	-1	-1/3	1/3	1	p value	q value
0	23	30	25	38	42	55	61	75	0.54	0.97
3,000	2	22	75	17	4	50	152	27	0.18	0.88
6,000	0	4	66	46	1	14	135	83	0.67	0.97
9,000	3	43	54	16	4	89	105	35	0.48	0.97
12,000	65	22	10	19	137	35	25	36	0.97	0.97

Note: Time represents split in sample based on 1/3 most and 2/3 least recent arrivals. The q value refers to the multiple hypothesis corrected p-values, using the Hochberg procedure. *** p<.01, ** p<.05, * p<.1. N=349.

Table 5.9 shows that time in Kampala does not significantly affect norms at any contribution level, before or after the application of the FWER strategy. This is further illustrated by table 5.10, which shows that average norms and expectations (by contribution level) for the 1/3 and 2/3 sample split are broadly comparable. The average time in Kampala of the 1/3 most recent arrivals was 1.59 years, compared to 6.90 years for the least recent 2/3.

It should be noted that the expectations do not total to 100% in either language treatment, or when the data is cut between by time in Kampala (i.e. by 1/3 and 2/3). As set out in the experimental script in appendix 1, participants were not asked to ensure that their responses total 100%, but rather to state how many people in the room contributed the amounts shown in 10% intervals. As I am interested in the relative difference by treatment (for language experiments) or by time in Kampala at a contribution level, this is not believed to affect the analysis. As the approach was incentivised, each 'guess' by participant could potentially attract a bonus payout. Forcing participant to ensure their total expectations equalled 100% would have potentially interfered with the attempt to accurately guess the expectations at each contribution level. Allocating their estimation across the five contribution levels simultaneously is conceptually more difficult than doing the same task at individual contribution levels. This incentive was consistent across treatments and thus responses across treatments are comparable. Forcing expectations to total 100% for statistical testing (even if the proportions between the contribution levels were maintained) would change the relative values by treatment at each contribution level. However, for robustness, the test performed in table 5.12 was conducted where expectations were proportionally forced to total 100%. This empirical approach did not yield a different result once the FWER Hochberg procedure was implemented; expectation were not found to significantly vary by exposure (i.e. time in Kampala).

Table 5.10: Average Norms and Expectations in Kampala by Time

	Norms		Expectations	
Contribution	1/3	2/3	1/3	2/3
0	0.156	0.103	18.2%	20.2%
	(0.732)	(0.757)	(0.219)	(0.234)
3,000	0.250	0.270	17.3%	17.9%
	(0.409)	(0.431)	(0.150)	(0.145)
6,000	0.525	0.575	22.6%	22.8%
	(0.396)	(0.367)	(0.170)	(0.167)
9,000	0.153	0.149	21.6%	25.7%
	(0.487)	(0.486)	(0.166)	(0.190)
12,000	-0.454	-0.420	33.1%	30.3%
	(0.751)	(0.769)	(0.271)	(0.252)
N	349		349	

Note: Split by time spent in Kampala. 1/3 represents the 116 subjects who have spent the least amount of time in Kampala and 2/3 representing the 233 subjects who have spent the most.

This analysis has been performed with alternative data cuts (for example 1/4 most recent compared to 3/4 least recent) with results shown in table 5.11. None of the alternative data splits provide significant results and this additional analysis was not included in the registered pre-analysis plan.

Table 5.11: Kampala Norms by Time in Kampala

	1/2 - 1/2		2/3 - 1/3		3/4 - 1/4		4/5 - 1/5		5/6 - 1/6		6/7 - 1/7		7/8 - 1/8	
Contribution	p	q	p	q	p	q	p	q	p	q	p	q	p	q
0	0.66	0.71	0.35	0.8	0.17	0.82	0.52	0.81	0.34	0.93	0.86	0.86	0.46	0.92
3,000	0.36	0.71	0.8	0.8	0.82	0.82	0.29	0.81	0.39	0.93	0.13	0.66	0.09	0.44
6,000	0.46	0.71	0.28	0.8	0.64	0.82	0.45	0.81	0.93	0.93	0.63	0.86	0.92	0.92
9,000	0.71	0.71	0.7	0.8	0.41	0.82	0.81	0.81	0.81	0.93	0.74	0.86	0.32	0.92
12,000	0.41	0.71	0.46	0.8	0.34	0.82	0.21	0.81	0.11	0.57	0.43	0.86	0.49	0.92
N	175	174	233	116	262	87	280	69	291	58	300	49	306	43
	8.1	2.2	6.9	1.6	6.4	1.3	6.1	1.2	6	1	5.9	0.8	5.8	0.7

Note: The q value refers to the multiple hypothesis corrected p-values, using the Hochberg procedure. *** p<.01, ** p<.05, * p<.1. N=349. At 1/3 time in urban was 3 years, with an 1/3 average 1.6 years, 2/3 average 6.9 years and overall average 5.1 years.

5.5.2 Expectations by Exposure

Moving to expectations, table 5.12 shows that time spent in Kampala has a significant impact on the 9,000 contribution level at the 10% significance level, but this does not survive the FWER correction. Table 5.10 shows the summary

statistics by contribution level and data split (being the comparison between 1/3 (N=116) participants who have spent the least time in Kampala, against the 2/3 (N=233) who have spent the most time in Kampala). There is little evidence to support that expectations are influenced by exposure. This is consistent with section 5.4.2 which shows that expectations do not differ by place. If expectations had differed by exposure, this would have contradicted section 5.4.2, suggesting a factor was missing from the analysis. I find no evidence of this.

Table 5.12: Learning Expectations: Kampala Expectations by Time in Kampala

	Expectations		Rank Sum	
Contribution	1/3	2/3	p value	q value
0	18.2%	20.2%	0.17	0.52
3,000	17.3%	17.9%	0.13	0.52
6,000	22.6%	22.8%	0.81	0.81
9,000	21.6%	25.7%	0.07*	0.35
12,000	33.1%	30.3%	0.45	0.81

Note: The q value refers to the multiple hypothesis corrected p-values, using the Hochberg procedure. *** p<.01, ** p<.05, * p<.1. N=349.

5.6 Results Experiment 2 - Community & Wall Street Game

5.6.1 Group 1: Contributions

This section sets out the results for contributions group, with two specific tests. Figure 5.12 summarises the overall relationship between treatment and actual/hypothetical. In treatment 1, the Community Game was played first and the Wall Street Game second, with treatment 2 played in the reverse order. In both treatments, the first game played is termed 'actual' and the second 'hypothetical'. Table 5.13 presents the mean of responses by game, by treatment. The first observation here is that the Community Game has higher average responses than the Wall Street Game overall. The second observation is that average contributions by game, whether actual or hypothetical, appear to be approximately equal: Community Game UGX 4,153 vs UGX 4,348, Wall Street Game UGX 3,924 vs UGX 3,653 (table 5.14 sets out whether these differ statistically).

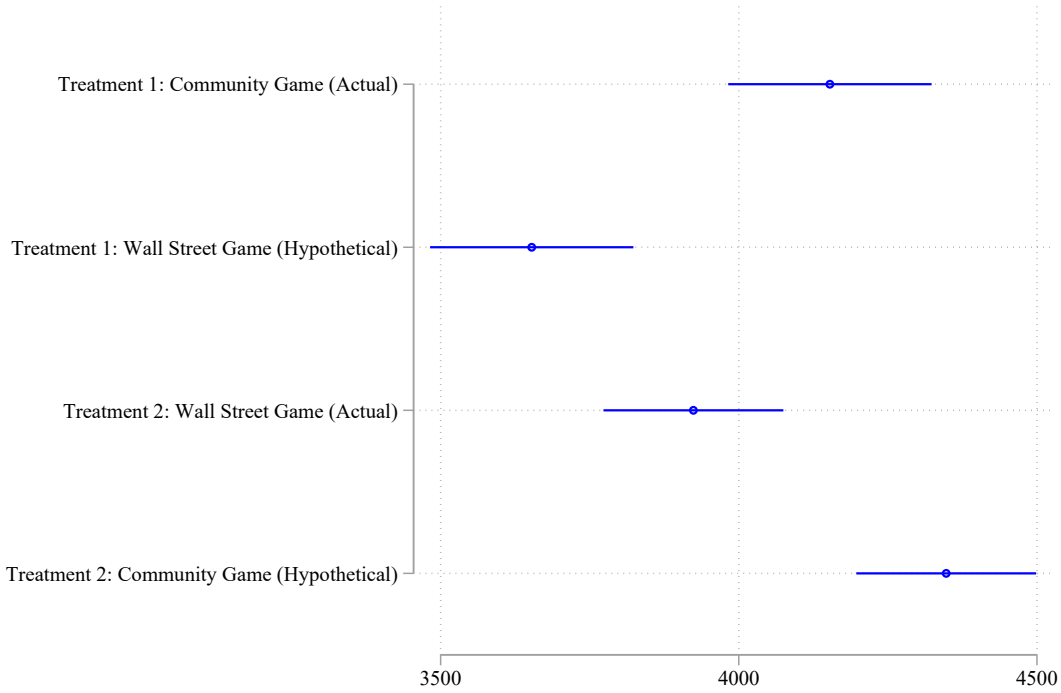
The experiment involved 481 participants. Of these, 153 failed to answer all of the control questions correctly. This is a higher ratio than seen chapter 4, though this experiment randomly sampled in a more rural setting and did not preferentially recruit university students. Thus, the level of mathematical competence can reasonably be expected to be lower and indeed, from table 5.3, only 25% of participants have tertiary level education. This provides a sample size of N=328 (excluding the control failures), split between treatment 1 and 2, 144 and 184 respectively. Each participant provides both an actual and hypothetical answer, data is arranged long so that N doubles for each treatment.

Table 5.13: Summary of Mean Contributions by Actual and Hypothetical

	Actual	Hypothetical
Community Game	4,153	4,348
N	144	184
Std Dev	*2,463*	*2,624*
Wall Street Game	3,924	3,653
N	184	144
Std Dev	*1,823*	*2,386*

Note: All amounts shown in Uganda Shillings (UGX).

Figure 5.12: Contributions: Interaction between Treatment & Actual vs Hypothetical



Note: Treatment 1 participants played the Community Game first ('Actual') and the Wall Street Game second ('Hypothetical'), with the reverse for treatment 2. For Treatment 1 N=144 and for Treatment 2 N=184. Graph shows the means and 95% confidence intervals for each estimate. Contributions on x axis in Ugandan Shillings (UGX).

Table 5.14 sets out the results for families 1.1 to 1.2, which test whether there is a difference between actual and hypothetical responses by game frame, across subject (as each participant only played each game frame once, either actual or hypothetical). For example, whether or not hypothetical responses can broadly approximate actual responses, or vice versa. For family 1.1 and 1.2 there is insufficient evidence to reject the null that actual and hypothetical responses are equal. Thus, this suggests that hypothetical responses, for both frames, are statistically accurate proxies for what participants actually answer (across subjects), within 0.15 and 0.08 standard deviations for the Wall Street and

Community Game respectively.

Table 5.14: Group 1 Contribution Results, by Family

Family	1.1	1.2
	Wall Street	Community
ActHyp	-348.657	270.170
	(1.45)	(0.95)
Age	22.500*	49.751***
	(1.85)	(3.79)
Female	399.446	156.985
	(1.68)	(0.54)
Constant	3144.984***	2728.441***
	(8.11)	(6.15)
N	318	318

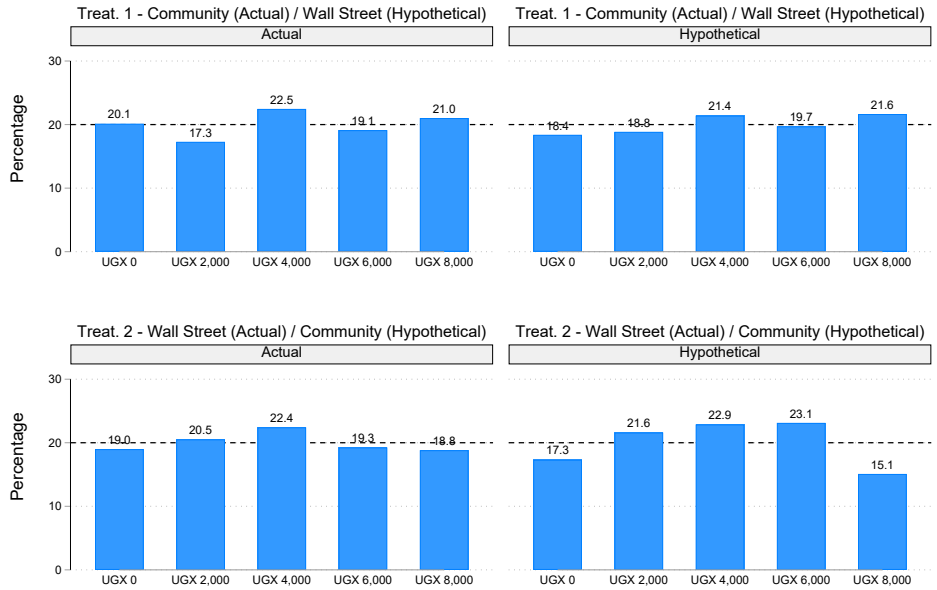
Note: *t* statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is the contribution to the common pot. N=318 after control failures excluded. $ActHyp$ is a dummy coded as 0 if actual and 1 if hypothetical.

Result 1: There is evidence that actual and hypothetical contributions differ within treatment, and insufficient evidence to reject the null that actual and hypothetical responses are equal for each game. This suggests that participants are both able to identify different levels of contributions by game, and that hypothetical responses are able to approximate actual contributions within 0.15 and 0.08 standard deviations for the Wall Street and Community Game respectively.

5.6.2 Group 2: Expectations

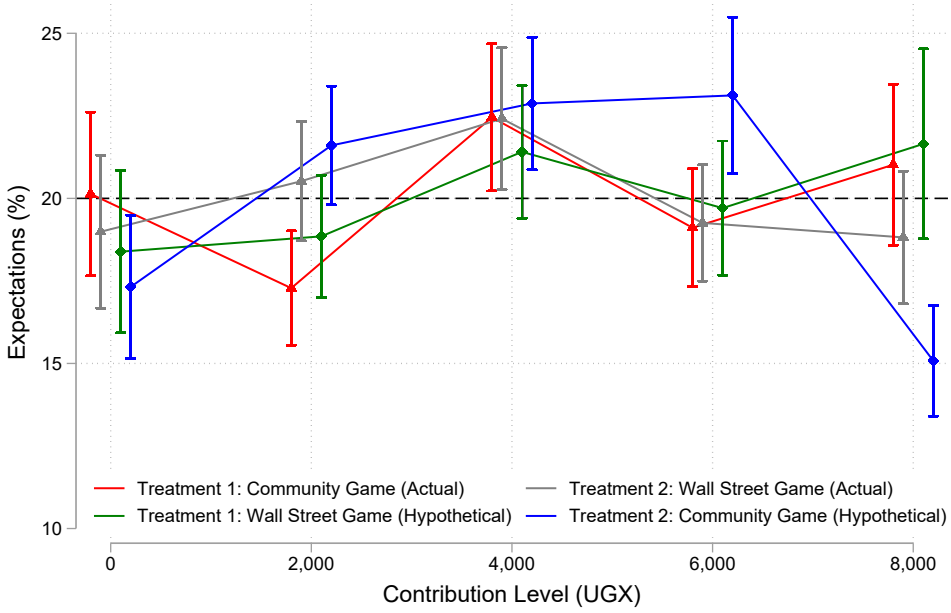
This section sets out the results for the expectations group of hypotheses, with four specific families of tests. Figure 5.13 presents the distribution of expectations, by contribution level, across treatment and actual/hypothetical. Figure 5.14 shows the four distributions overlayed with 95% confidence intervals. Participants were asked how many individuals in their session they expected to contribute at each of the five contribution levels. As session size varied, responses have simply been divided by the number of actual participants to arrive at a percentage. In situations where the sum of responses was higher or lower than the actual number of participants, responses were proportionally scaled so that the total response across all levels, by participant, totals 100%. Given that there are five contribution levels, the 20% line is shown as the baseline. For expectations a regression pooling data across levels is not possible as the mean will always total 20% by $ActHyp$ (sum of expectations is always 100%, divided by the five contribution levels).

Figure 5.13: Expectations: by Treatment, by Actual/Hypothetical



Note: Treatment 1 played the Community Game first ('Actual') and the Wall Street Game second ('Hypothetical'), with the reverse for treatment 2. Graphs show the distribution of expectations at each of the five contribution levels.

Figure 5.14: Expectations: by Treatment, by Level, Combined



Note: Error bars show 95% confidence intervals.

Visually, there is little differentiation between actual and hypothetical for treatment 1. There two humps at UGX 4,000 and UGX 8,000, which are broadly approximate, though there is some variation between UGX 2,000. For treatment 2 the the distributions are less consistent, particularly at UGX 6,000 (which is higher for the Community Game) and UGX 8,000 (which is higher for the Wall Street Game). From figure 5.14, there appears to be relatively little consistency across treatments and it is noticeable that the error bars are large and frequently overlap. This suggests that expectations of others are not well established.

Families 2.1 and 2.2 consider whether hypothetical expectations statistically different from actual expectations, which is a diagonal comparison of the distributions presented in figure 5.13, and blue vs red/green vs grey in figure 5.14. Family 2.1 assesses the Wall Street Game, which compares the top-right and bottom-left distributions. Visually these distributions appear to be broadly approximate, and indeed, referencing the full results in table 5.15 there is insufficient evidence to reject the null that the two distributions are equal. However this is misleading and does not mean that hypothetical expectations can predict actual expectations; both distributions are simply relatively flat, with responses typically near or close to 20% throughout. Considering the Community Game (top-left panel compared to the bottom-right), visually there is significantly more variation here, with significant results at the 1% level for UGX 2,000 and UGX 8,000 and at the 5% level for UGX 6,000. As the distributions vary significantly, this suggests that hypothetical expectations for the community game are unable to predict actual expectations.

Table 5.15: Families 2.1 & 2.2: Expectations, by Game

	Wall Street Game (Family 2.1)					Community Game (Family 2.2)				
	0	2,000	4,000	6,000	8,000	0	2,000	4,000	6,000	8,000
ActHyp	-0.648	-2.358*	0.167	0.440	2.398	-1.984	4.991***	-0.035	2.664**	-5.636***
	(1.703)	(1.316)	(1.432)	(1.371)	(1.797)	(1.652)	(1.266)	(1.507)	(1.419)	(1.518)
Age	0.079	0.249***	-0.031	-0.153***	-0.144*	0.200**	0.144**	-0.151*	-0.108*	-0.085
	(0.079)	(0.088)	(0.067)	(0.059)	(0.083)	(0.093)	(0.071)	(0.078)	(0.055)	(0.068)
Female	-3.092*	3.922***	0.797	-0.188	-1.439	-1.868	1.349	-1.971	2.748*	-0.258
	(1.762)	(1.267)	(1.469)	(1.384)	(1.762)	(1.735)	(1.291)	(1.542)	(1.464)	(1.485)
Constant	18.402***	11.979***	21.715***	23.756***	24.148***	15.415***	12.462***	27.819***	20.726***	23.578***
	(2.731)	(2.551)	(2.309)	(2.163)	(2.889)	(2.667)	(2.210)	(2.681)	(1.959)	(2.575)
q value	0.931	0.059*	0.931	0.931	0.931	0.178	0.009***	0.951	0.020**	0.009***
N	312	312	312	312	312	310	310	310	310	310

Notes: Each cell reports the coefficient (rounded to three decimal places) for the specified variable and level. T-values are reported in rounded brackets, rounded to two decimal places.

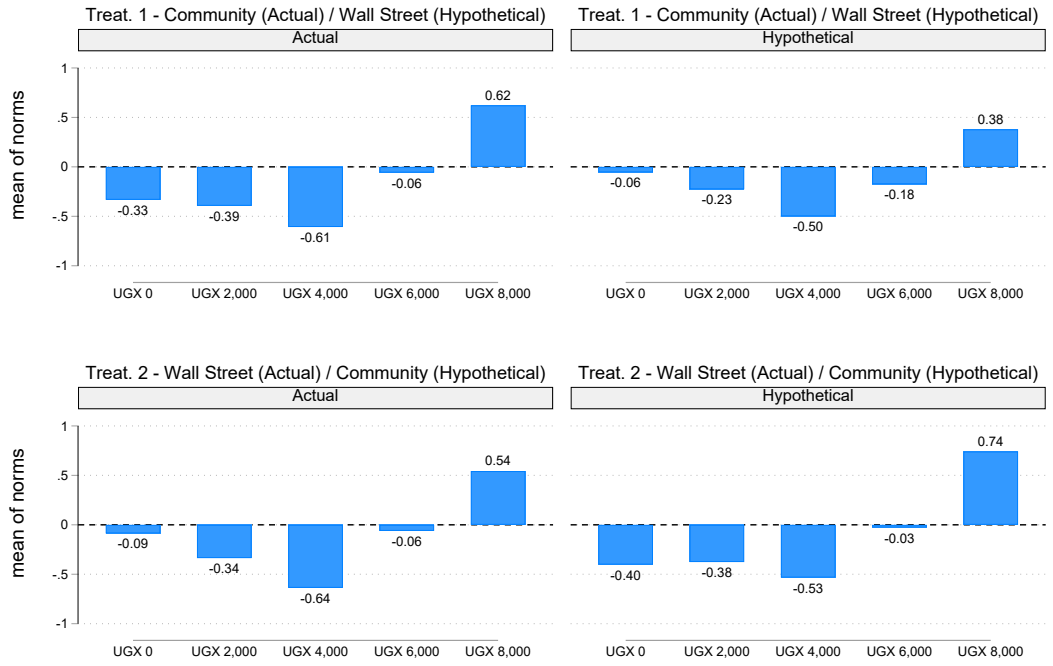
Result 2: Expectations do not differ within treatment (visually), suggesting that distinct distributions of expectations do not exist by game/frame. For the Community Game, expectations differ significantly across the majority of the distribution. For the Wall Street Game, hypothetical and actual expectations are broadly consistent and there is generally insufficient evidence to reject the null the two distributions are equal. However, as both distributions are relatively flat, they appear to be equal because of the lack of general variations within treatments as established

expectations appear not to exist.

5.6.3 Group 3: Norms

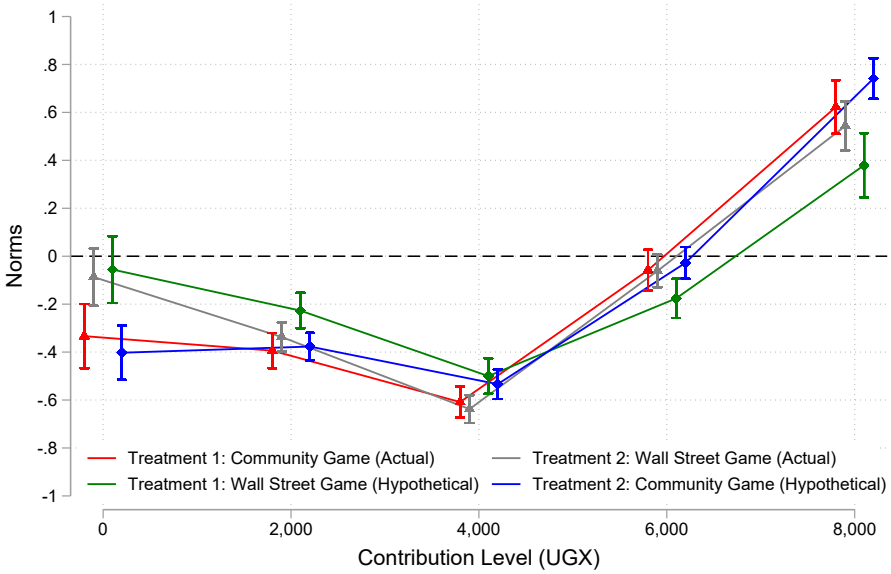
This section sets out the results for the norm group of hypotheses, with two specific families of tests. Figure 5.15 presents the distribution of norms, by contribution level, across treatment and actual/hypothetical. Participants were asked to rank norms on a scale of social acceptability, where -1 is 'very socially unacceptable', 1 is 'very socially acceptable' and 1/3 and -1/3 are 'socially acceptable' and 'unacceptable' respectively. Figure 5.16 presents a combined distribution with 95% error bars. Comparing this graph to figure 5.14, it is immediately apparent that the shape of the distribution of norms is more consistent than that of expectations. Secondly, there is significant variation at the extreme: red and blue (which represent the Community Game, actual and hypothetical respectfully) show a higher social acceptability when being more generous (UGX 8,000) and lower social acceptability when being less generous (UGX 0) compared to the Wall Street Game (green and grey).

Figure 5.15: Norms (Social Acceptability): by Treatment, by Actual/Hypothetical



Note: Average social acceptabilities shown based on the Krupka and Weber (2013) scale, where -1 / 1 references very socially unacceptable / acceptable respectively.

Figure 5.16: Norms: by Treatment, by Level, Combined



Note: Average social acceptabilities shown based on the Krupka and Weber (2013) scale, where -1 / 1 references very socially unacceptable / acceptable respectively. Error bars show 95% confidence intervals.

Moving to the detailed results in table 5.16, $ActHyp$ is the variable of interest. For families 3.1 and 3.2, there is insignificant evidence to reject the null that the populations are equal.

Table 5.16: Regression Results by Treatment and Group.

Family	3.1	3.2
	Wall Street	Community
ActHyp	0.001	0.040
	(0.06)	(1.76)
Level	0.117***	0.239***
	(6.57)	(15.13)
Age	0.001	0.001
	(0.41)	(0.72)
Female	-0.020	-0.005
	(-0.87)	(-0.21)
Constant	-0.356***	-0.660***
	(-6.83)	(-12.80)
N	1,560	1,552

Result 3: There is insufficient evidence to reject the null that the populations are equal when comparing actual and hypothetical norms within game frames. The distribution of norms follows a consistent 'U' shape for all treatments throughout, suggesting that hypothetical norms can be used as a statistically accurate proxy for actual norms.

5.7 Discussion and Conclusion

Social norms are ubiquitous across human society and have been shown to influence cooperative behaviour (Fehr and Schurtenberger, 2018). By nature, social norms are highly susceptible to context. As social norms evolve along different stochastic pathways, local conformity and global diversity emerges (Young, 2015). An individual will likely have multiple overlapping social norm sets as they will have membership to multiple cultural groups. Despite the well documented impact of social norms on decision making, the interaction between context, culture and identity is often difficult to disentangle. This chapter offers methodological improvements to the Krupka and Weber (2013) approach, namely through the inclusion of hypothetical norm elicitation and a demonstration of its ability to provide a statistically accurate proxy for actual responses. Secondly, there has been relatively little research exploring how rural-to-urban migration differs to the substantially researched international migration. The existence of 'linkages' and the ease by which a national migrant can return to their original community means there are likely key differences (Stites et al., 2014).

I initially contribute to this literature by applying the Krupka and Weber (2013) methodology by eliciting hypothetical social norms by place in a pre-registered experiment (Experiment 1). This permits within subject analysis of social norm sets. Applying the Bicchieri (2010) distinction between normative beliefs ('norms') and empirical expectations ('expectations'), I find a significant difference in the Kampala and Mbale norms, but little evidence to support a difference in expectations. The difference in norms is in line with anthropological literature. This suggests that subjects are aware that different norms exist across the two geographical regions, but that they don't believe other Gisus would adhere to those norms.

As Experiment 1 only measured actual and hypothetical norms and expectations in one location (Kampala), Experiment 2 further contributes by measuring actual and hypothetical across two frames. This enables a test of whether hypothetical contributions, expectations and norms can act as a statistically accurate proxy for actual values across subject in an incentive compatible design. I find that there is insufficient evidence to reject the null that hypothetical contributions and norms are equal to actual responses. Considering both Experiment 1 and 2 results jointly, norms vary significantly across both geographical place (Experiment 1) and game frame (Experiment 2) and expectations are consistently insignificant. This provides evidence that in two different contexts individuals hold multiple sets of norms, which can be accessed when the appropriate frame is applied. Secondly, this is consistent despite a significantly different sample populations and the profile of social norms. The Experiment 1 was conducted in Kampala with university students, displaying a 'humped' norm distribution. Experiment 2 used a more diverse sample of Gisu, from more rural districts and display a 'U' shaped norm distribution. Despite these differences, the different contexts applied led to measurably different norms. This suggests that actual norms can be accurately hypothesised (i.e. which is socially acceptable), but the degree to which others are expected to adhere to those norms cannot.

A further contribution to the literature comes from exploring the interaction between social norms and identity. As set out by Hoff and Stiglitz (2016), social context informs mental modes which in turn determine behaviour. I find that language significantly interacts with place for norms, but is insignificant for expectations. This means that a Gisu speaking Lugisu in Mbale will have different norms to a Gisu speaking Luganda in Kampala, despite language individually being insignificant. However, the expected behaviour of others is thought to be the same. Language is a store of cultural information and it is intuitive that complementing factors, such as being physically present in the place where a language originates, could trigger a different perception or decision making.

The final contribution of this chapter is with regards to exposure in the context of rural-to-urban migration. The 2019 experiment exclusively recruits ethnic Gisus, all of which have lived within Bugisu region and have subsequently migrated to Kampala. Theories of international migration (see Algan et al. (2012) for overview) suggest that migrants should become aware of different cultural factors, even if they do not themselves assimilate. Though I find that norms do differ by place, there is no evidence to support that these differences vary by the level of exposure. Previous studies have shown that awareness of different cultures occurs quickly, often within a year (Cameron et al., 2015; Berge et al., 2018). Though alternative data cuts were tested, one explanation is that learning happens more quickly, but there is insufficient power in this experiment to detect this. An alternative explanation is that, unlike a different country, it is possible Gisus already had an awareness of the cultural norms of Kampala (the capital city) prior to migration. This would have been bolstered by common linkages with those who have already migrated (Stites et al., 2014).

The main contributions of this chapter are methodological. Policy makers could adopt the approach taken in these experiments to determine the degree to which migrants (international or rural-to-urban) are likely to assimilate and cooperate with their new communities post migration. The inclusion and interaction of other identity, narrative or worldview variables within this analysis would also be of use to policy makers. Further, knowing that a specific language in one place stimulates more cooperative norms, but another language is more effective in a second place, is potentially a powerful tool in promoting public good provisioning. Being able to accurately elicit and measure expected behaviour and/or social norms hypothetically is also likely to be of value to policy makers and researchers alike.

This chapter has a number of limitations which provide opportunities for further future research. The first is, because of the within subject design, the 2019 experiment was not able to determine whether place affects contributions as the data for Mbale could not be collected. While this has partially been mitigated/further investigated by the 2024 experiment, this adopts an alternative approach due to budgetary and logistical constraints (the 2024 experiment was run at the same time, and in addition to, the main 'misperceptions' experiment as per chapter 6). Two experiments would need to be run conterminously across the two geographies, potentially with some of the same participants, in order to fully investigate the differences by place. Another area for further research is regarding exposure. This experiment did not collect sufficient survey data (to ascertain whether social norms were known prior to migration) and had insufficient sample of less than one year migrants. A follow-up experiment would include a more detailed survey

and be more targeted in recruiting individuals that had moved to Kampala within the last couple of years. This would permit a monthly exposure analysis.

Lastly, when comparing the shape of the norm distribution across Experiment 1 and Experiment 2, it is clear that these are different. The norms (social acceptability distributions) elicited in Kampala in 2019 show a $\cap$ shape, whereas norms elicited in Mbale in 2024 show a $\cup$ shape. These experiments recruited very different participants; Experiment 1 used students in Kampala, all of who had migrated from the Bugisu region, whereas Experiment 2 selected from a sampling frame in the region surrounding Mbale. Only 25% of this sample had a tertiary level of education and the demographics of the sample vary significantly. Additionally, between 2019 and 2024 Uganda (as well as the rest of the world) experienced the COVID 19 pandemic, which may have had a significant impact on cooperative norms. Participants in both experiments (who passed the control questions) achieved a greater than 50% bonus payout rate (which is consistent with other experiments), thus understanding was strong in both samples. Lastly, while the general design of both experiments followed a Public Goods Game, the frames applied by each experiment differed. However, it is not possible to conclude on the reason for the difference, and this is an area for further research.

Chapter 6

Chapter 6 - Social Norms, Values & Misperceptions

6.1 Introduction

Social change typically requires a critical mass to enable a 'tipping point', where sufficient momentum is present for a transformational cascade to occur. The correction of misperceptions is the most commonly cited mechanism driving this process (Legros and Cislaghi, 2020). This involves updating personal (mis)perceptions about what peers really think is socially acceptable (i.e. injunctive norms as in Cialdini et al. (1991)) or the values that they hold (Schwartz, 1992, 2012). Misperceptions are essentially a misalignment between private (first order) and normative (second order) beliefs, where individuals underestimate the extent to which second order beliefs actually align to their privately held first order beliefs. The correction of these mis-calibrations has been attributed to significant behavioural change, particularly in the short-term (Bursztyn and Yang, 2022).

Within the literature, a relatively small number of papers elicit both in-group and out-group beliefs (Cullen and Perez-Truglia, 2022; Mildenberger and Tingley, 2019; Ahler and Sood, 2018; Bordalo et al., 2016; Levendusky and Malhotra, 2016; Wiswall and Zafar, 2015; Ahler, 2014; Graham et al., 2012), though these are almost exclusively focused on the United States political domain. When grouping this data, the evidence suggests that: 1) an individual's own attitudes and beliefs are strongly, positively correlated with perceptions of the in-group, but negatively correlated with out-groups; and 2) the accuracy of in-group perceptions tends to be higher when compared to out-groups (Bursztyn and Yang, 2022). Though there have been suggestions, the mechanisms supporting these findings are not understood. A better understanding of the underlying mechanisms will be particularly valuable in promoting pro-sociality in socially fragmented regions, as well as all manner of behavioural and policy change.

This chapter addresses three core research questions: 1) How do individuals' second-order beliefs about their in-group compare to those about out-groups? 2) What mechanisms drive misperceptions in these inter-group contexts? 3) How accurate are these perceptions and what factors contribute to inaccuracy? To explore these questions, I conduct a two-phase experimental study in Uganda, focusing on three ethnic groups (Gisu, Ganda, and Acholi) across five domains: Politics, Religion, Deception, Violence and Adultery. Phase One involves a regional survey to elicit beliefs and perceptions, while Phase Two uses a lab-in-the-field design to test mechanisms such as motivated reasoning, cultural identity, projection bias, and the curse of knowledge. This approach allows for a detailed analysis of both the direction and accuracy of misperceptions, as well as an exploration in to differences across both culture and domain.

Within a cross disciplinary literature, interventions to correct misperceptions have taken many forms, including socialisation, informational campaigns and mass media messaging (Chung and Rimal, 2016; Lapinski and Rimal, 2005; Cialdini and Trost, 1998). Interventions can be particularly effective when initial perceptions are shown to be wrong (Morris et al., 2015). Beyond misperceptions, there are a number of other factors that can also drive social change. These factors can be structural or institutional (Morris et al., 2015; Jackson, 2011), legal (Sunstein, 1996) or socio-political (most notably the rise of role models or lobby groups - see Young (2015)), though these mechanisms

are not the focus of this paper. There has been an increasing focus on the determinants of misperceptions, and thereby how these might be corrected. Bursztyn and Yang (2022) conducted a broad cross-disciplinary meta-analysis and find that misperceptions are widespread across all domains (i.e. not simply measurement error), and that attempts to re-calibrate misperceptions generally work as intended. Additionally, they note that large behavioural changes are common immediately after misperception corrections, though with a reduction over time.

One mechanism which can explain both positive and negative correlations with own beliefs is *cultural identity*, as set out within Bonomi et al. (2021). This builds on the psychology of identity and how culture influences beliefs across many domains (Haidt, 2013; Turner et al., 1987; Tajfel et al., 1979). Social Identity Perspective (SIP) is the leading theory of identity and inter-group relations, and combines Social Identity Theory (SIT, Tajfel et al., 1979), and Self Categorisation Theory (SCT, Turner et al., 1987). Within SIP, when a social cleavage becomes salient, individuals identify with one of the relevant groups. Identity then anchors beliefs about self, others, and behaviour. This causes beliefs to become polarised along features relevant to those groups and stereotypes can manifest. If an individual's identity (for example, their ethnicity) is made salient, the individual will anchor their identity to that group, leading to a positive correlation with that individual's own beliefs. Equally, opposing stereotypes about other identities (or ethnicities) could potentially be the source of the observed negatively correlated beliefs.

A further mechanism which can explain both positive and negative correlations is *motivated reasoning* (Kunda, 1990). This theory suggests that reasoning and cognitive processes (such as memory encoding, information selection & evaluation, attitude formation, judgment and decision-making) are influenced by goals, motivations or vested interests. This is most commonly noted in politics, where individuals support a political affiliation (from which they gain utility), even when it is not in their economic interests (Berinsky, 2017). Beyond this, motivated reasoning has been noted in a range of contexts, from economics to health (Epley and Gilovich, 2016). Applying the theory to perceptions, an individual could rationalise that their association with their in-group drives the highest utility, even if that's not necessarily true or they actually disagree with a number of in-group behaviours (i.e. cognitive dissonance). The individual will thus ignore the undesirable factors and focus only on those which support this argument. Framed another way, an individual may perceive their in-group to be more like them than they really are, with the opposite for out-groups. Kahan (2015) summarise this within their Poltically Motivated Reasoning Paradigm, where Bayesian updating (specifically the likelihood ratio) is biased towards predispositions. Though this focuses mainly on political reasoning, the theory could equally be applied to other domains.

I move now to consider mechanisms which can only explain positive correlations with in-groups. The *curse of knowledge* is one such mechanism (Camerer et al., 1989). Within economics, better informed agents are assumed to be capable of producing the judgements of less informed agents. Within the context of group-based misperceptions, an individual is ethnically X and also an individual Y. When forecasting beliefs of less informed agents (being those that are ethnically X only), the individual should be able to ignore Y (i.e. there should be no correlation between Y

and perceptions of X). This is also known as the law of iterated expectations (Chow and Teicher, 2003). However, experimental findings suggest that better informed agents are unable to ignore private information, even when it is in their interests to do so (see Camerer (2011) for overview). For out-groups, as the individual does not share group characteristics, the curse of knowledge is not *directly* applicable.

An alternative and complementary mechanism for positive correlations with an in-group could be *projection bias* (Madarász, 2012). This proposes that individuals too often act as if others could guess their private information correctly. When projecting information, a person misperceives the distribution of information available to others. The model ties together a set of empirically well-documented social mispredictions, including hindsight bias (Fischhoff, 1975), curse-of-knowledge (Camerer et al., 1989) or the illusion of transparency (Gilovich et al., 1998). Under full information projection, $\rho = 1$, an individual believes that all the information they know is shared with others. In the case of partial information projection, $0 < \rho < 1$, an individual believes that the probability that their information is available to others is between the truth and the full projection case. Finally, $\rho = 0$ corresponds to correct expectation, where all private information is unknown by others. However, this theory does not convincingly extend to out-groups; projection bias is only applicable when the third party could conceivably have gained private information about the individual.

Moving now to the accuracy of prediction across in and out-groups, there are clear interpretations once the mechanisms are understood. Inaccuracies can be summarised into two distinct categories. The first is rational inattention; individuals are incorrect because they are not correctly calibrated, meaning greater information is needed about other groups (Bursztyn and Yang, 2022). This suggests that established misperception correction techniques are likely be effective at reducing inaccuracies in the short-term. The second category is identity-based errors. In this scenario, simply providing additional information alone will not work, and may even lead to conflict (Bonomi et al., 2021; Iyengar et al., 2019). Bridging misunderstandings of this nature will require a more nuanced approach.

Existing research on in vs out-group perceptions has been almost exclusively focused along the political domain, predominantly in the US (Cullen and Perez-Truglia, 2022; Mildenberger and Tingley, 2019; Ahler and Sood, 2018; Bordalo et al., 2016; Levendusky and Malhotra, 2016; Wiswall and Zafar, 2015; Ahler, 2014; Graham et al., 2012). It is likely that the underlying mechanism(s) is also contingent on the domain. For this reason, I explore a range of domains: Politics, Religion, Deception, Violence and Adultery. Further, the existing research typically only considers perceptions about one or two Western groups. It is highly likely that social proximity will also be a factor in both perceptions and their accuracy. For this reason, I intend to select a range of different ethnics groups from Uganda, collecting data from each group's geographical home: the Gisu (Mbale), Ganda, (Kampala), and Acholi (Gulu), each being increasingly more distant from Gisu culture (the primary group of focus).

I have three main results. The first is that the belief of others (hereon referred to as second order beliefs)

is significantly influenced by own beliefs (hereon first order beliefs), and this is more extreme for the in-group than the out-group. This effect is not consistent across all domains, and there was insufficient evidence for politics and adultery to reject the null that in and out-group beliefs are equal. This is consistent with the meta-analysis presented by Bursztyn and Yang (2022), though it is surprising to find politics does not vary significantly. The second result is that, for both the in and out-groups, there is strong evidence that Motivated Reasoning and Cultural Identity are underlying mechanisms driving the beliefs of others. Motivated Reasoning is consistent across all domains, whereas Cultural Identity is more varied, specifically for Adultery which is not significant. For the other in-group mechanisms, Curse of Knowledge is also significant across all domains, whereas Projection Bias is only significant for Politics.

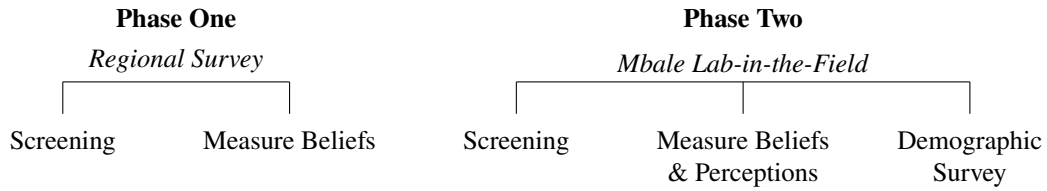
Concerning inaccuracy, the third result is that inaccuracy significantly differs across all domains between the in and out-group, however inaccuracy is actually highest for the in-group, which is the opposite relationship observed in US based political studies. This highlights the additional complexity added to the analysis when in and out-groups are from different ethnic groups. Inaccuracy also differs significantly by self-reported confidence level, where an increase in confidence reduces accuracy. When considering social proximity, inaccuracy differs significantly by cultural distance, across all domains, though an increase in cultural distance actually reduces inaccuracy. Lastly, inaccuracy differs between the in and out-group, with the in-group, on average, being more (erroneously) confident.

This chapter proceeds as follows. Section 6.2 sets out the research design, section 6.3 presents the empirical overview, sections 6.4 to 6.7 present the results and section 6.8 concludes.

6.2 Research Design

Figure 6.1 summarises the elements of the experiment. The phases run chronologically, as data collected during phase one will be used in phase two. The Gisu are the primary ethnic group of interest and form the 'in-group' for this study. The Ganda and Acholi form the 'out-group'. A full experimental script has been set out in appendix 3. Funding for this study has been provided by the South East Network for Social Sciences (SeNSS) (`https://senss-dtp.ac.uk/`) as part of the author's SeNSS ESRC-funded doctoral studentship award. No other funding is attached to the project. Ethics approval for the study has been granted by the International Development Ethics Committee, University of East Anglia, and the Research Ethics Committee, Uganda. For phase one, consent was obtained verbally and appendix 4 contains the written (Lugisu) consent form used during phase two.

Figure 6.1: Experimental Design



6.2.1 Phase One - Regional Survey

Screening & Beliefs Survey

This experiment was pre-registered with the RCT Registry in August 2024 in advance of any fieldwork (Hill, 2024). Phase one of the experiment comprises of a survey asking four questions across five domains (politics, religion, deception, violence and adultery) and across the three ethnic groups (at three specific geographical locations, central to each ethnic group): the Gisu (Mbale), Ganda (Kampala) and Acholi (Gulu). The questions have been set out in table 6.1. For context, figure 6.2 presents a map of Uganda, showing (in blue) the three geographical areas where the surveys were conducted. The sample size is N=100 for each ethnic group (thus N=300 for Phase One). Phase Two involves only Gisu participants (N=481) and was been conducted near Mbale (Bugisu Sub-Region).

Table 6.1: Phase One - Beliefs Survey Questions

Question
QS.1 Do you think politics is important?
QS.2 Do you think religion is the most important aspect of identity?
QS.3 Do you think it is acceptable to deceive others?
QS.4 Do you think the use of violence is acceptable?
QS.5 Do you think cheating on your spouse is acceptable?

All questions were answered on a scale between 1-10.

Figure 6.2: Map of Uganda Showing Regions of Interest

Adapted from https://www.nationsonline.org/oneworld/map/uganda-map.htm, accessed 14/01/2024

For each group, the survey was conducted in the local language of each ethnic group. Questions and scripts were carefully cross-translated to ensure consistency. Enumerators randomly sampled passers-by in each of the three regions. This experiment (both phases one and two) was facilitated by The Field Lab (`https://thefieldlabuganda.com/`), an organisation based in Mbale which has expertise running experiments of this nature. As data on specific groups is required, the following questions will be asked in advance to screen participants:

1. Which tribe/ethnic group do you most identify with?

2. Where were you born?

3. Are you fluent in [local language]?

Based on this evidence, coupled with spoken language proficiency, enumerators are able to determine whether the individual is an ethnic Gisu, Ganda or Acholi, as appropriate. Otherwise, the applicant was thanked for their interest and informed that they do not meet the criteria. The survey took about five minutes and participants receive UGX 5,000 for their time (about £1, or the cost of good meal). Participants were asked to complete their answers anonymously on a separate answer sheet as the questions were read verbally. Once completed, answer sheets were folded and posted into a sealed box to reassure participants that their responses were truly anonymous. For Phase One, no demographic or identifying data was collected to ensure responses could not be attributed to any one participant.

6.2.2 Phase Two - Mbale Region Lab-in-the-Field

Screening & Sampling

Phase Two requires only ethnic Gisu participants, recruited in or near to Mbale. In order to investigate the underlying mechanisms, it is important that all subjects have similar cultural identities/exposures. Recruitment adopted the following process and was facilitated by The Field Lab, who have expertise and local connections in the region:

a) A suitable research site was identified, and three parishes in the Bugisu Sub-County (near Mbale) were selected.

b) A sampling frame was obtained from the electoral register.

c) Participants were randomly selected from the sampling frame, split evenly across the three parishes.

d) The Field Lab then visited each participant with the help of the LC1 chairperson (a local elected leader, or somebody nominated by the LC1 chairperson) and elicit willingness to participate. During this step, the same questions set out in phase one were asked to ensure the potential participant is an ethnic Gisu.

e) If the selected individual was willing to participate The Field Lab then provided the individual with the details of the lab-in-the-field session, including time and location. The participant's identity was verified before they were able to join the session.

f) If they were unwilling or unavailable to participate, they were randomly replaced with somebody else on the sampling frame. All participants were >18 years old and had multiple opportunities to express if they are

unwilling/uncomfortable to participate (including providing written consent forms in advance of participation - see appendix 4). Though the local elected leader will support, they will not coerce participation; this will be solely the individual's choice.

All subjects were screened for language proficiency and to ensure they are an ethnic Gisu during step d). Consistent with phase one, the following questions will be asked prior to candidates being invited to participate.

1. Which tribe/ethnic group do you most identify with?
2. Where were you born?
3. Are you fluent in Lugisu?

Based on this evidence, coupled with spoken language proficiency (Lugisu), enumerators are able to determine whether the individual is an ethnic Gisu.

Perceptions & Beliefs

This stage is an incentivised survey which builds on Phase One and includes only Gisu (N=481) participants. There are two groups of perceptions questions; group A (table 6.2) collects perception and accuracy data, where group B (table 6.3) collects additional data to derive the mechanisms. The questions are asked sequentially, as as set out in the script included in appendix 3. The payout for this section is contingent on how close participant's responses are for the incentivised questions.

In group A, some of the questions ask the participant's own opinions (first order beliefs), while others ask the participant to record their belief of 'others' (second order beliefs). 'Others' (referenced as 'X' throughout) refers to the Gisu, Ganda and Acholi (i.e. this will include the participant's own group (Gisu), plus the other two).

Table 6.2: Phase Two - Group A Questions

	Domain	Question
Q1.1	Politics	Do you think politics is important? [1-10, where 1 is complete disagreement with the statement, 10 is complete agreement with the statement]
Q1.2*	Politics	We asked 100 'X' "do you think politics is important?" on a scale 1-10. How many do you think answered 6 or higher?
Q1.3	Politics	Where 1 = not at all confident and 10 = completely confident, how confident do you feel about your answer to Q.1.2?
Q1.4*	Politics	The handout shows the actual responses for a randomly selected 10% of all 'X' who answered 6 or higher in Q.1.2. Given this additional information, what would your answer to Q.1.2 be now, for each group?
Q2.1	Religion	Do you think religion is the most important aspect of identity? [1-10, where 1 is complete disagreement with the statement, 10 is complete agreement with the statement]
Q2.2*	Religion	We asked 100 'X' "do you think religion is the most important aspect of identity?" on a scale 1-10. How many do you think answered 6 or higher?
Q2.3	Religion	Where 1 = not at all confident and 10 = completely confident, how confident do you feel about your answer to Q2.2?
Q2.4*	Religion	The handout shows the actual responses for a randomly selected 10% of all 'X' who answered 6 or higher in Q2.2. What would your answer to Q2.2 be now?
Q3.1	Deception	Do you think it is acceptable to deceive others? [1-10, where 1 is complete disagreement with the statement, 10 is complete agreement with the statement]
Q3.2*	Deception	We asked 100 'X' "do you think it is acceptable to deceive others?" on a scale 1-10. How many do you think answered 6 or higher?
Q3.3	Deception	Where 1 = not at all confident and 10 = completely confident, how confident do you feel about your answer to Q3.2?
Q3.4*	Deception	The handout shows the actual responses for a random 10% of all 'X' who answered 6 or higher in Q3.2. What would your answer to Q3.2 be now?
Q4.1	Violence	Do you think the use of violence is acceptable? [1-10, where 1 is complete disagreement with the statement, 10 is complete agreement with the statement]
Q4.2*	Violence	We asked 100 'X' "do you think the use of violence is acceptable?" on a scale 1-10. How many do you think answered 6 or higher?
Q4.3	Violence	Where 1 = not at all confident and 10 = completely confident, how confident do you feel about your answer to Q4.2?
Q4.4*	Violence	The handout shows the actual responses for a random 10% of all 'X' who answered 6 or higher in Q4.2. What would your answer to Q4.2 be now?
Q5.1	Adultery	Do you think cheating on your spouse is acceptable? [1-10, where 1 is complete disagreement with the statement, 10 is complete agreement with the statement]
Q5.2*	Adultery	We asked 100 'X' "do you think cheating on your spouse is acceptable?" on a scale 1-10. How many do you think answered 6 or higher?
Q5.3	Adultery	Where 1 = not at all confident and 10 = completely confident, how confident do you feel about your answer to Q5.2?
Q5.4*	Adultery	The handout shows the actual responses for a random 10% of all 'X' who answered 6 or higher in Q5.2. What would your answer to Q5.2 be now?

*These questions are incentivised. In the script, Q1-5.4 will be asked after question Q7 (see below), to allow for initial responses to be collected before providing the additional information and further answer sheets.

Within the survey, the order of the tribal groups will be randomised to measure and control for order effects. Table 6.2 sets out the the first 20 questions, four per domain, to be answered three times each (once for each culture 'X'), totalling 60 questions. These questions are partially incentivised (questions 2 and 4 in each domain), and measure:

1. The subject's opinion on the issue, measured on a ten point scale (where 10 is complete agreement with the statement, 1 being no agreement with the statement).
2. The subject's guess (or perception) of the proportion of the three ethnic groups who would agree with the same statement (would rate greater or equal to 6 on the scale), as a percentage. This is from the subject's perspective. This is incentivised, with a payout being given if participants correctly guess within 10% of the actual answer. One of the 15 questions (five domains by three cultures) will be selected to determine the payout at the end of the session.
3. How certain the subjects are about each of their 15 guess (for question 2, five domains by three cultures), on a scale 1 - 10 (with 10 being absolute certainty).
4. After providing additional information showing what 10% of phase one subjects actually answered, subjects are then asked to provide an updated answer to question 2. This is incentivised in the same way, with subjects trying to correctly guess the actual answer given.

Moving to group B questions (table 6.3), this part of the survey will measure additional characteristics which can be used to determine the mechanisms. Question six will have a total of 30 responses (ten characteristics by three groups), whereas question seven will require only one answer (for Gisu only). These questions are not incentivised. Question six also contains a number of irrelevant stereotypes to reduce demand effects.

Survey

Before answering the survey, a random question will be selected by one of the participants to determine the bonus payouts. While participant payoffs are being calculated, participants will be asked to complete an exit survey. This section will collect key demographic information, including age, gender, education level, religion and political affiliations. The questions, as well as their intended use in the empirical analysis, are set in table 6.4. Each question will be read aloud, giving time for subjects to write their answers.

Phase 2 was comprised of 14 sessions in total, with an average session size of 35. In all sessions, the order of the cultural groups was randomised to counter order effects.

Table 6.3: Phase Two - Group B Questions

Questions	
Cultural Identity ‡	
Q6.1	The X* are faithful?
Q6.2	The X* are intelligent†?
Q6.3	The X* are stubborn†?
Q6.4	The X* are aggressive?
Q6.5	The X* are efficient†?
Q6.6	The X* are very religious?
Q6.7	The X* are arrogant†?
Q6.8	The X* are deceitful?
Q6.9	The X* are straight forward†?
Q6.10	The X* are political?
Projection Bias	
Q7	How well do you feel other Gisu understand you and your beliefs?

† I include these irrelevant stereotypes to address demand effects. ‡ "Where 1 = completely disagree and 10 = completely agree, as a Gisu how much do you agree the following statements reflect stereotypes of X?"

Table 6.4: Survey Questions

Question	**Nature**
How old are you?	Control variable
What is your gender?	Control variable
What is your highest education level?	Descriptive
Are you a member of any political party?	Descriptive
If so, which?	Descriptive

Note:'Nature' refers to the intended use of the data collected. 'Control variable' means that age or gender data will be used as a control in regressions. 'Descriptive' means that the data has been collected purely to provide summarised information about the sample.

6.3 Empirical Overview

6.3.1 Theoretical Framework and Model

Bursztyn and Yang (2022) identify four potential mechanisms to explain the noted differences between in and out-group perceptions. These mechanisms are summarised below:

Mechanism	**Scope**	**Specific driver**
(MR) Motivated Reasoning	In and Out-Group	Bias - cognitive processes
(CI) Cultural Identity	In and Out-Group	Salient group stereotypes
(PB) Projection Bias	In-Group only	Bias - information asymmetry
(CoK) Curse of Knowledge	In-Group only	Bias - iterated expectation

It is possible that all four mechanisms may be concurrently contributing to the in-group effect, though

only Cultural Identity and Motivated Reasoning can convincingly explain the out-group effect. In addition to these mechanisms, an individual's degree of uncertainty must also be considered. An individual may have a perception which they feel is inaccurate, either because of lack of exposure or awareness (social distance and self-reported confidence levels are considered in the experiemental design). Thus to test the mechanisms I will need two basic models. These models are presented to aid understanding of the empirical approach and are not intended to be exhaustive:

$$IN_{Di} = f(U_{Di} + CI_{Di} + MR_{Di} + CoK_{Di} + PB_{Di}) \tag{6.1}$$

Where: IN_{Di} is the perception about the Gisu only (i.e. the primary reference group used throughout, forming the in-group) for individual i across the five domains D. U_{Di} is the individual's uncertainly level, which will influence the overall belief (the higher the level of uncertainty, the wider the range of possible beliefs). CI, MR, CoK and PB are the four mechanisms that may act to bias or impact the overall perception.

$$OUT_{DCi} = f(U_{DCi} + CI_{DCi} + MR_{DCi}) \tag{6.2}$$

OUT_{DCi} is the perception about out-groups, where C references the Ganda (Kampala) and the Acholi (Gulu).

6.3.2 Deriving the Mechanisms

Motivated reasoning (MR) is typically measured in the lab by first measuring a set of beliefs without any incentive, and then comparing these results to incentivised beliefs (for recent literature review see Amelio and Zimmermann, 2023). For example, there is a significant body of research looking at how financial advisors select stocks for the clients (Gneezy et al., 2020; Jonas and Frey, 2003). The basic approach is to present the profiles of two stocks, asking players to make a recommendation with and without incentives. In a second treatment, players are told about the incentives ex-ante, where there is an incentive to recommend the weaker stock. The difference between the two treatments has been used as a proxy for the degree of motivated reasoning at work.

Kahan (2015) summarise a more appropriate model for my research questions, the Politically Motivated Reasoning Paradigm, which builds on motivated reasoning. In this model, which is based on Bayesian updating, the likelihood ratio is biased towards a predisposition. In other words a prior is not accurately updated when new information is provided. This bias is a measurement of motivated reasoning. The difference between the answer to Q1-5.2 and Q1-5.4 is the 'update' following new information (what a randomly selected 10% of surveyed participants in phase one actually answered). Terming this new information as the 'signal', if Bayesian updating was working perfectly,

the difference between the signal (which is given before Q1-5.4 are answered) and answer to Q1-5.2 should be equal to difference between Q1-5.2 and Q1-5.4. This is because, as the signal is the most accurate information available, participants are expected to re-calibrate to this information[1]. I intend to test this within a regression framework.

Cultural identity (*CI*) can be measured by stereotypes. There are three primary techniques for measuring stereotypes (with a number of variations) within the literature. The first is the checklist technique, developed by Katz and Braly (1933). The checklist (of 84 different adjectives) seeks to reveal the consensus of one group's view of another, by selecting the adjectives they believe best describes that group. The second is the percentage, as set out by Brigham (1971), which seeks to determine the prevalence of a set of traits for a given group using percentage rankings. The third technique is the diagnostic ratio (see Martin, 1987) where subjects indicate the percentage of people in general who have each trait, and then for the group of interest. The ratio between the general population and the group of interest determines the degree of stereotype. Stephan et al. (1993) compare the three techniques, and find a high degree of agreement, suggesting (irrespective of the technique used), the same cognitive processes were at play. Further, Devine and Elliot (1995) suggest improvements which make clear the distinction between personal beliefs and actual stereotypes when framing questions.

For my experiment, the percentage technique is the most applicable to my design. I intend to ask ten questions, Q6.1 $\rightarrow$ 6.10, to measure the relative strength of stereotypes across the five domains, in addition to five 'dummy' questions designed to avoid demand effects. Q6.1 "faithful" will measure adultery, Q6.4 "aggressive" will measure violence, Q6.6 "very religious" will measure religion, Q6.8 "deceitful" will measure deception and Q6.10 "political" will measure politics. The higher (or lower) the value of these variables, the stronger the presence of a stereotype (either positively or negatively). This can be tested within a simple regression framework. These adjectives are either directly taken from the original Katz and Braly (1933) 84 adjectives, or slightly adapted. It's also particularly important to make the Gisu identity salient before taking measurements.

Projection Bias (*PB*) is only relevant for the in-group, so Q7 will only be asked for the Gisu directly. As per Madarász (2012), under full information projection, $\rho = 1$, an individual believes that all the information they know is shared with others. Conversely, $\rho = 0$ corresponds to the correct expectation, where all private information is unknown by others. Clearly the response to Q7 (on a scale 1-10) is a direct proxy for ρ; if projection bias is a mechanism, a higher value of *PB* should, on average, increase the proportion of Gisu the participant thinks agrees with them. I am not aware of any empirical papers which use/test this theory directly, but it is cited within Bursztyn and Yang (2022) as a potential mechanism. As ρ is a literal belief, it feels appropriate to simply ask subjects what their belief is.

[1]Participants were informed of the design of Stage 1: "We recently interviewed 100 Gisu, Baganda and Acholi and asked them to answer these questions. These interviews were completed on the streets of either Mbale, Kampala or Gulu, where passers-by were randomly selected. All responses were collected blind and placed in a sealed box so that responses could not be traced back to individual interviewees". The signal provided should be interpreted as random and unbiased (as the data collection was unobserved and anonymous) by participants. Thus, the signal should typically provide additional information which is perceived to be accurate and thus act to pull the average update towards it.

I will use Q1.1 → Q5.1 and Q1.2 → Q5.2 to calculate the participant's level of *agreement*. Q1.2 → Q5.2 asks the participant to state how many of the 100 responses (for each culture) were above 5.5. I can then use the participants response to Q1.1 → Q5.1 to calculate their level of 'agreement': if Q1.1 → Q5.1 is less than 5.5, agreement will be equal to 1 minus the response to Q1.2 → Q5.2 (where Q1.2 → Q5.2 is expressed as a percentage); if Q1-5.1 is greater than 5.5, then the agreement is equal to the response to Q1-5.2. I can then regress *agreement* (dependent variable) against responses to Q7, plus appropriate controls.

Curse of Knowledge (CoK**)**, in the initial Camerer et al. (1989) paper, a simple investment game was used. This game has two rounds, where predictions of company payouts by subjects in round 1 (group 1) determine the payouts (dividends) in round 2 (for group 2). Group 2 know that the asset dividend was determined by group 1's predictions, but they do not know the exact amount of the dividend. Group 1 received 'value line' reports about the companies, which group 2 also received, as well as additional information. In this experiment, group 2 should be able to closely predict what group 1 predicted, disregarding the additional information. However, due to the curse of knowledge, there were persistent 'errors' in the estimates.

For this experiment, the curse of knowledge is only a potential mechanism for the in-group. For Q1.2 → Q5.2 I have data from the 100 surveyed Gisu in phase one, in addition to the 481 Gisu who participated in the lab based sessions in Mbale (phase two), and thus N=581. This allows me to compute Q1.2 → Q5.2 actual. The difference between Q1.2 → Q5.2 and Q1.2 → Q5.2 actual can be defined as the $ErrorRate$. If the Curse of Knowledge is present, there should be a correlation between the error rate and the participant's own belief (Q1.1 → Q5.1). I will regress the error rate (dependent variable) against Q1.1 → Q5.1, plus controls as appropriate.

Uncertainty (U**): Deriving associated variables**

In addition to the mechanisms, I am also interested in the **inaccuracy of predictions**, and what might be driving these. As discussed, misperceptions (or inaccuracies in the perceptions of others) can be categorised as either *irrational inattention* or *identity-based errors*. Cultural stereotypes, and other aspects of identity associated with culture, fall firmly within the identity-based error category. These are more difficult to 'correct'. Motivated Reasoning, Curse of Knowledge and Projection Bias are all essentially types of cognitive bias which, with carefully designed interventions, could be 're-calibrated'. For this reason, I classify these as irrational inattention mechanisms.

Phase One collects data of each culture, enabling *inaccuracy* to be calculated. As previously defined, the difference between participant i's measured belief (in Phase Two, $QX2$) and the actual average response of those surveyed (Phase One, $QX2Actual$) is the $ErrorRate$. Inaccuracy can thus be calculated as:

$$Inaccuracy_{Di} = |QX2Actual_{Di} - QX2_{Di}| \tag{6.3}$$

Hence, $Inaccuracy$ is the absolute of the $ErrorRate$ and thus can only be a positive value between zero and 99 (as this is the scale of QX2).

Cultural distance is another metric of interest, particularly to enable a better understanding how uncertainty (U) impacts misperceptions. Uganda has a large number of tribal groups (c. 43, Logan 2022), some of which vary significantly in attitudes and beliefs. Uganda was selected for this experiment because it has significant variation in cultural groups. As the Gisu are the focal group of Phase Two, an assessment of variation has been made relative to this group. The two tribes selected are both well known to the Gisu (particularly important when eliciting stereotypes), and also divergent across a range of domains. The table below presents ten diverse questions from the Afrobarometer R8 2022 data (Logan, 2022). While not exhaustive, the average value (indexed where Gisu is set to 1), is indicative of the overall cultural distance of each group to the Gisu (i.e. how culturally close each group is to the Gisu). For the empirical specifications, $Cultdist$ will be computed as 1 - the values shown in the table below. Thus, Gisu will be nil (i.e. no cultural distance), Ganda is 0.17 and Acholi 0.42.

Table 6.5: Cultural Distance, Afrobarometer Data

Afrobarometer R8 Question	**Acholi**	**Ganda**	**Gisu**
Q10A Free to say what think	0.46	0.91	1.00
Q10C Free to vote without pressure	0.59	0.84	1.00
Q43B Report corruption without fear	1.01	1.35	1.00
Q54A Fear violence in neighbourhood	1.28	1.12	1.00
Q72A Aware of climate change	-0.12	1.00	1.00
Q82A Ethnic group treated fairly	0.59	0.55	1.00
Q83 Most people can be trusted	0.92	1.04	1.00
Q86E Neighbour: different political party	-0.23	0.07	1.00
Q86C Neighbour: homosexual	0.83	0.97	1.00
Q86A Neighbour: different religion	0.44	0.48	1.00
Average	**0.58**	**0.83**	**1.00**

Index created from R8 2022 Afrobarometer data (Logan, 2022). Questions selected to cover a range of themes, including those relevant for the domains investigated. All questions standardised on a scale -1 to 1, with intermediary questions scaled on linear intervals (three options coded -1, 0, 1, four options coded -1, -0.33, 0.33, 1, etc). Gisu indexed to 1 in each case, thus all other responses divided by Gisu value. Negative values means reponses were at the oppositive end of the scale to Gisu. Average based on a equal value weighting.

6.3.3 Family-Wise Error Rate (FWER) strategy

Within each family of hypotheses, I will control for the number of tests conducted, controlling the FWER using the Romano-Wolf stepdown p-values for multiple hypothesis testing procedure implemented in Stata by Clarke (2021). I will use the 5% significance level for all decisions, using the appropriate q values. There are three testing groups,

which are set in turn below:

- Group 1: Replication
- Group 2: Mechanisms for beliefs
- Group 3: Exploring inaccuracy

Typically there are multiple families within each group, as set out in each section below. Families are typically defined by hypotheses, in/out-group and mechanism, which are all distinct in nature. Corrections are typically required as domains are tested both jointly and individually.

6.3.4 Group 1: Conceptual Replication

H1: Measured own beliefs are more related to those of the in-group than out-group (across the five domains).

The first family of tests explores whether measured own beliefs are more consistent with the in-group than the out-group. This is a replication and extension of previous studies across a number of new domains and ethnicities in Sub-Saharan Africa. The original studies found that an individual's own attitudes and beliefs are strongly, positively correlated with perceptions of the in-group, but negatively correlated with out-groups (Bursztyn and Yang, 2022). The existing evidence is primarily in the political domain in the United States, whereas this test will also consider Religion, Deception, Violence and Adultery, across a number of ethnic groups. While existing evidence suggests a positive/negative correlation (in a meta-analysis), I will explore the relative strength of the relationship between own beliefs (first order beliefs) and beliefs of the in and out-groups (second order beliefs), as not all domains may follow the positive/negative distinction.

This testing group will be comprised of the following families for my FWER strategy:

- **Family 1:** test jointly (all domains) and then each domain individually

The specification will be a regression, with β_1 being the coefficient of interest to test *H1*. $Answer > 5.5_{Di}$ is a dummy equal to 1 if for QX1 (being first order beliefs) if a participant answered six or higher. This approach allows me to test three key things simultaneously. The first is whether second order beliefs differ by $ingroup$, with the coefficient β_2 providing this insight. Secondly, if β_3 is positive, this would suggest that on average first order beliefs are important in determining second order beliefs generally. Lastly, the coefficient β_1 indicates whether there is an interaction between being both a member of the in-group and answering above 5.5 in QX1 (first order beliefs). If β_1 is positive, this would mean that a Gisu participant's second order beliefs are affected by their first order beliefs.

Standard errors will be clustered at the participant level. Each regression will use the basic set up:

$$Y_{Di} = \alpha + \beta_1 ingroup * Answer > 5.5_{Di} + \beta_2 ingroup_{Di} + \beta_3 Answer > 5.5_{Di} + controls + \epsilon_{Di}$$

where Y_{Di} is the second order belief (i.e. responses to Q2 across each domain), by each domain D, by participant i. Controls will comprise of age and gender. In line with the FWER strategy, the p-values (by domain) will be corrected using the Romano-Wolf stepdown p-values Clarke (2021). This procedure allows for dependence among p-values (by bootstrap resampling), and also allows wild cluster bootstrapping within the specification natively within Stata.

6.3.5 Group 2: Mechanisms for Beliefs

This group of tests explores whether measured **in-group** beliefs (across each of the five domains) are significantly affected by Motivated Reasoning, Cultural Identity, Projection Bias or Curse of Knowledge:

H2.1: Measured in-group beliefs (across each of the five domains) are significantly affected by Motivated Reasoning, Cultural Identity, Projection Bias or Curse of Knowledge.

This group of tests also investigates whether measured **out-group** beliefs (across each of the five domains) are significantly affected by Motivated Reasoning or Cultural Identity:

H2.2: Measured out-group beliefs (across each of the five domains) significantly affected by Motivated Reasoning or Cultural Identity.

As set out in the FWER strategy, each mechanism will form it's own family, with five p-values to be corrected in each (by domain). Each mechanism is distinct, with unique dependent and independent variables in each case. However, each uses a regression framework, so the Romano-Wolf stepdown p-values Clarke (2021) is used in each case. Throughout this group, reference is made to survey responses provided during phase two of the experiment. Full details of these questions are provided in section 6.2.

This testing group will be comprised of the following families for the FWER strategy:

For **in-group:**

- **Family 2.1.1:** test Motivated Reasoning jointly (across all domains) and then each domain individually
- **Family 2.1.2:** test Cultural Identity jointly (across all domains) and then each domain individually

- **Family 2.1.3:** test Projection Bias jointly (across all domains) and then each domain individually
- **Family 2.1.4:** test Curse of Knowledge jointly (across all domains) and then each domain individually

For **out-group:**

- **Family 2.2.1:** test Motivated Reasoning jointly (across all domains) and then each domain individually
- **Family 2.2.2:** test Cultural Identity jointly (across all domains) and then each domain individually

Family 2.1.1 & 2.2.1: Motivated Reasoning (In & Out-Groups)

As discussed, I will be applying the Politically Motivated Reasoning Paradigm (Kahan, 2015) to test this model. In this model, Bayesian updating (specifically the likelihood ratio) is biased towards a predisposition. The difference between the answer to Q1.2 → Q5.2 and Q1.4 → Q5.4 is the 'update' following new information (what a randomly selected 10% of surveyed participants in phase one actually answered). Terming this new information as the 'signal', if Bayesian updating were working perfectly, the difference between the signal and answer to Q1.2 → Q5.2 should be related to difference between Q1.2 → Q5.2 and Q1.4 → Q5.4. Each regression will use the basic set up:

$$Update_{Di} = \alpha + \beta_1 signal - firstguess_{Di} + controls + \epsilon_{Di}$$

where $Update_{Di}$ is the difference between Q1.4 → Q5.4 and Q1.2 → Q5.2 (second order belief after and before the signal), by each domain D, by participant i. β_1 is the coefficient of interest, and $signal - firstguess_{Di}$ is calculated as the difference between the signal and responses to Q1.2 → Q5.2. Thus there will be six regressions, one for each domain, plus one taking into account all of the data across domains. Controls will comprise of age and gender. The approach for in and out-groups is identical except that in-group is comprised only of the Gisu and the out-group of the Ganda and Acholi.

Family 2.1.2 & 2.2.2: Cultural Identity (In & Out-Groups)

As discussed, given there is little difference in the outcome of each stereotype elicitation methodology (Stephan et al., 1993), I adopt the percentage technique. Q6.1 → Q6.10 measure the relative strength of stereotypes across the five domains, in addition to five 'dummy' questions designed to avoid demand effects. Q6.1 "faithful" will measure adultery, Q6.4 "aggressive" will measure violence, Q6.6 "very religious" will measure religion, Q6.8 "deceitful" will measure deception and Q6.10 "political" will measure politics. The closer to either 1 or 10, the stronger the presence

of a stereotype (captured in the independent variable CI), as a low value will signal a stereotype of an opposite nature. Each regression will use the basic set up:

$$Y_{Di} = \alpha + \beta_1 CI_{Di} + controls + \epsilon_{Di}$$

where Y_{Di} is the response to Q1.2 → Q5.2 (second order beliefs before the signal), by each domain D, by participant i. β_1 is the coefficient of interest, and CI_{Di} are the five adjectives of interest. Thus there will be six regressions, one for each domain, plus one taking into account all of the data across domains. Controls will comprise of age and gender. The approach for in and out-groups is identical except that the in-group is comprised only of the Gisu and the out-group of the Ganda and Acholi.

Family 2.1.3: Projection Bias (In-Group)

Projection Bias is only relevant for the in-group, so Q7 will only be asked for Gisu directly. Q7 acts as direct proxy for ρ, but scaled 1-10 (Madarász, 2012). If projection bias is a mechanism, a higher value of PB should, on average, correspond with a significant movement in Q1.2 → Q5.2.

I will use the relationship between Q1.1 → Q5.1 and Q1.2 → Q5.2 to calculate the participant's level of *agreement*. Q1.2 → Q5.2 asks the participant to state how many of the 100 responses (for each culture) were above 6. I can then use the participant's response to Q1.1 → Q5.1 to calculate their level of 'agreement': if Q1.1 → Q5.1 is less than 6, agreement will be equal to 1 minus the response to Q1.2 → Q5.2 (where Q1.2 → Q5.2 is expressed as a percentage); if Q1.1 → Q5.1 is greater than/or equal to 6, then the agreement is equal to the response to Q1.2 → Q5.2. I can then regress *agreement* (dependent variable) against responses to Q7 (PB), plus appropriate controls. I will also include additional specifications to check the sensitivity of this cut-off. Hence, each regression uses the basic set up:

$$Y_{Di} = \alpha + \beta_1 PB_{Di} + controls + \epsilon_{Di}$$

where Y_{Di} is *agreement*, by each domain D, by participant i. β_1 is the coefficient of interest. Thus there will be six regressions, one for each domain, plus one taking into account all of the data across domains. Controls will comprise of age and gender.

Family 2.1.4: Curse of Knowledge (In-Group)

Curse of knowledge is only a potential mechanism for in-groups. For Q1.2 → Q5.2 I have the actual data for not only the initial 100 surveyed in Phase One, but also the 481 Gisu who participated in Phase Two (N=581). Table 6.6 summarises the proportion of Gisu (Phase One and Two), Ganda and Acholi (Phase One), that answered six or higher (inclusive) for the first order belief question for each domain (i.e. Q1.1 → Q5.1). The Ganda and Acholi data is relevant for Group 3 - Exploring Inaccuracy only.

Table 6.6: % of Participants Answering Six or Higher, by Domain & Ethnicity - Phase One & Two

	Politics	**Religion**	**Deception**	**Violence**	**Adultery**
Gisu (N=581)	93%	77%	11%	5%	13%
Ganda (N=100)	77%	65%	17%	6%	23%
Acholi (N=100)	74%	67%	12%	13%	11%

Thus, the above becomes Q1.2 → Q5.2 *actual*. The difference between Q1.2 → Q5.2 and Q1.2 → Q5.2 actual can be defined as the 'error' rate, akin to the error rate set out in Camerer et al. (1989). If the curse of knowledge is present, there should be a correlation between the error rate and the participant's first order belief (Q1.1 → Q5.1). I regress the error rate (dependent variable) against Q1.1 → Q5.1, plus controls as appropriate. Hence, each regression uses the basic set up:

$$Y_{Di} = \alpha + \beta_1 QX1_{Di} + controls + \epsilon_{Di}$$

where Y_{Di} is the error rate, by each domain D, by participant i. β_1 is the coefficient of interest, and $QX1$ is first order beliefs (i.e. Q1.1 → Q5.1). Thus there will be six regressions, one for each domain, plus one taking into account all of the data across domains. Controls will comprise of age and gender.

6.3.6 Group 3: Exploring Inaccuracy

This group explores inaccuracy across three hypotheses, in addition to a number of additional exploratory specifications:

H3.1: Inaccuracy of predictions (across each of the five domains) is lowest for in-group.

H3.2: Inaccuracy of predictions (across each of the five domains) can be predicted by self-reported confidence level.

H3.3: Cultural distance from Gisu can predict inaccuracy of predictions (across all five domains).

This testing group will be comprised of the following families for the FWER strategy:

- **Family 3.1:** test inaccuracy of predictions jointly between in and out-group (across all domains) and then each domain individually
- **Family 3.2:** test inaccuracy against self-reported confidence levels jointly (across all domains) and then each domain individually
- **Family 3.3:** test inaccuracy against cultural distance jointly (across all domains) and then each domain individually

Family 3.1: Inaccuracy: In vs Out-Group

This family tests whether the inaccuracy of predictions (across each of the five domains) is lowest for in-group. As discussed, uncertainty U, should be lowest for the in-group, and increase as cultural distance and awareness increases. As set out in equation 6.3, inaccuracy is calculated as the difference between second order beliefs (Phase Two), and the actual responses collected during Phase One (and Phase Two for Gisu). Hence, each regression will use the basic set up:

$$Inaccuracy_{Di} = \alpha + \beta_1 inout_{Di} + controls + \epsilon_{Di}$$

where $Inaccuracy_{Di}$ is inaccuracy, by each domain D, by participant i. β_1 is the coefficient of interest, and *inout* is a dummy variable which equals 1 if Gisu and 0 otherwise. Thus there will be five regressions, one for each domain, plus one taking into account all of the data across domains. Controls will comprise of age and gender.

Family 3.2: Inaccuracy: self-reported confidence

This family investigates whether the inaccuracy of predictions (across each of the five domains) can be predicted by self-reported confidence level. Self-reported confidence levels ($QX3$) are taken from Q1.3 $\rightarrow$ Q5.3. Hence, each regression will use the basic set up:

$$Y_{Di} = \alpha + \beta_1 QX3_{Di} + controls + \epsilon_{Di}$$

where Y_{Di} is inaccuracy, by each domain D, by participant i. β_1 is the coefficient of interest, and $QX3$ is the self reported confidence level, on a scale 1-10. Thus there will be five regressions, one for each domain, plus one taking into account all of the data across domains. Controls will comprise of age and gender.

Family 3.3: Inaccuracy: cultural distance

This family tests whether cultural distance from Gisu can predict the inaccuracy of predictions (across all five domains). As set out in Section 3, R8 2022 Afrobarometer (Logan, 2022) data has been used to construct a measure of cultural distance ($cultdist$), where Gisu = 1, and all other cultures are scaled between 0 and 1. Hence, each regression will use the basic set up:

$$Y_{Di} = \alpha + \beta_1 cultdist_{Di} + controls + \epsilon_{Di}$$

where Y_{Di} is inaccuracy, by each domain D, by participant i. β_1 is the coefficient of interest, and $cultdist$ is index of cultural distance calculated from R8 2022 Afrobarometer data (Logan, 2022). Thus there will be five regressions, one for each domain, plus one taking into account all of the data across domains. Controls will comprise of age and gender.

Family 3: Other exploratory specifications

Below are three additional exploratory specifications, which are ancillary to the main hypotheses. All three specifications are at a population level, but a control has been added to capture domain effects. $\beta_2 Domain_{Di}$ is a categorical variable coded Politics = 1, Religion = 2, Deception = 3, Violence = 4 and Adultery = 5. $Controls$ are comprised of age and gender.

The first specification explores whether uncertainty differs across the in and out groups, where variables are as previously defined:

$$QX3_{Di} = \alpha + \beta_1 inout_{Di} + \beta_2 Domain_{Di} + controls + \epsilon_{Di}$$

The second specification explores whether the $\beta_1 UpdateRatio_{Di}$, being the ratio of the difference between the original first order beliefs (QX2) and the updated first order beliefs (QX4) and the difference between the original first order beliefs (QX2) and the signal, is affected by in/out-group membership. If the ratio is 1, it would represent perfect alignment, and the absence of any bias. All other variables are as previously defined.

$$UpdateRatio_{Di} = \alpha + \beta_1 inout_{Di} + \beta_2 Domain_{Di} + controls + \epsilon_{Di}$$

Lastly, the third specification investigates whether the $UpdateRatio_{DI}$ is affected by self-reported certainty $QX3_{Di}$. In other words, if a participant is more confident in their original answer, is this met by a smaller update? All other variables are as previously defined:

$$UpdateRatio_{Di} = \alpha + \beta_1 QX3_{Di} + \beta_2 Domain_{Di} + controls + \epsilon_{Di}$$

6.4 Summary Statistics

Table 6.7 presents a range of summary statistics for Phase Two participants. Demographic data was not collected for participants of Phase One. Of the 481 participants, 229 were males and 252 were female. The average age was 28.4 years, with the gender split being almost identical. Only Age and $Female$ are used as controls across all specifications, the remaining data is informational only. Christianity was the primary religion across the participants (at 78%), with the remaining 22% identifying as Muslim. Political membership was broadly consistent across the genders, with 25% of all participants having a political party membership. There is no evidence of imbalance across the sample.

Table 6.7: Summary Statistics - Phase Two

	Gender			Balance Test
	Male	Female	Total	p
Age	28.175 (9.938)	28.540 (10.613)	28.366 (10.288)	0.698
# people known in session	3.913 (4.730)	4.183 (4.326)	4.054 (4.520)	0.514
Religion				0.175
Christian	161 (70.3%)	191 (75.8%)	352 (73.2%)	
Muslim	68 (29.7%)	61 (24.2%)	129 (26.8%)	
Education Level				0.808
Primary	42 (18.3%)	52 (20.6%)	94 (19.5%)	
Secondary	138 (60.3%)	149 (59.1%)	287 (59.7%)	
Tertiary	49 (21.4%)	51 (20.2%)	100 (20.8%)	
Political Membership				0.161
No	163 (71.2%)	198 (78.6%)	361 (75.1%)	
Yes	66 (28.8%)	54 (21.4%)	120 (24.9%)	
N	229 (47.6%)	252 (52.4%)	481 (100.0%)	

Note: Age, # people known in session and political membership p values are for binary or continuous variables, and come from regressions. All others come from Pearson tests. In each case *** denotes significance at 1%, ** at 5% and * at 10% where relevant. Standard deviations, or percentages, are provided in parenthesis

Moving now to a summary of the experimental results, figure 6.3 provides a summary of first order beliefs (QX1 - left panel) and self-reported confidence levels (QX3 - right panel). First order beliefs, where Politics and Religion are considered in terms of importance, and where Deception, Violence and Adultery in terms of acceptability, is highest for Politics and Religion. Adultery is considered to be more socially acceptable than both Deception and Violence, which has the lowest rating. Considering now the self-reported confidence data, participants felt most confident in their responses to Politics and Religion and, on the whole, were most confident about the in-group (Gisu).

Figure 6.3: Summary of First Order Beliefs and Self-Reported Confidence, by Domain, by Ethnic Group

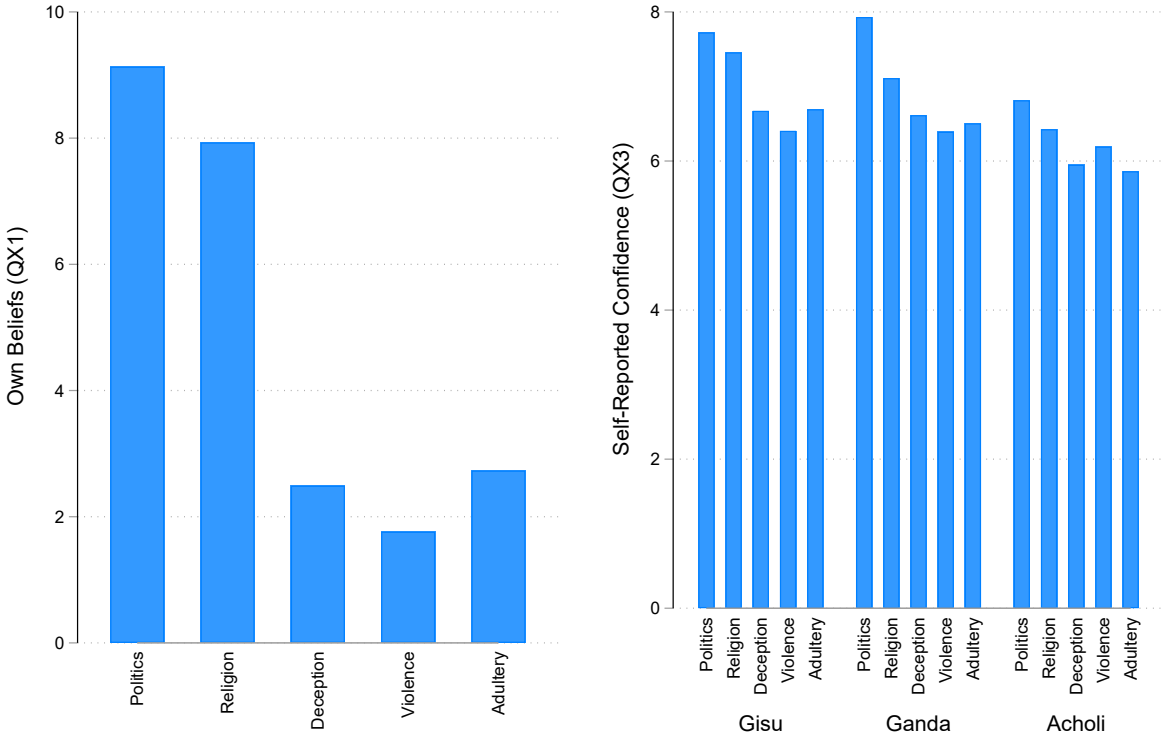
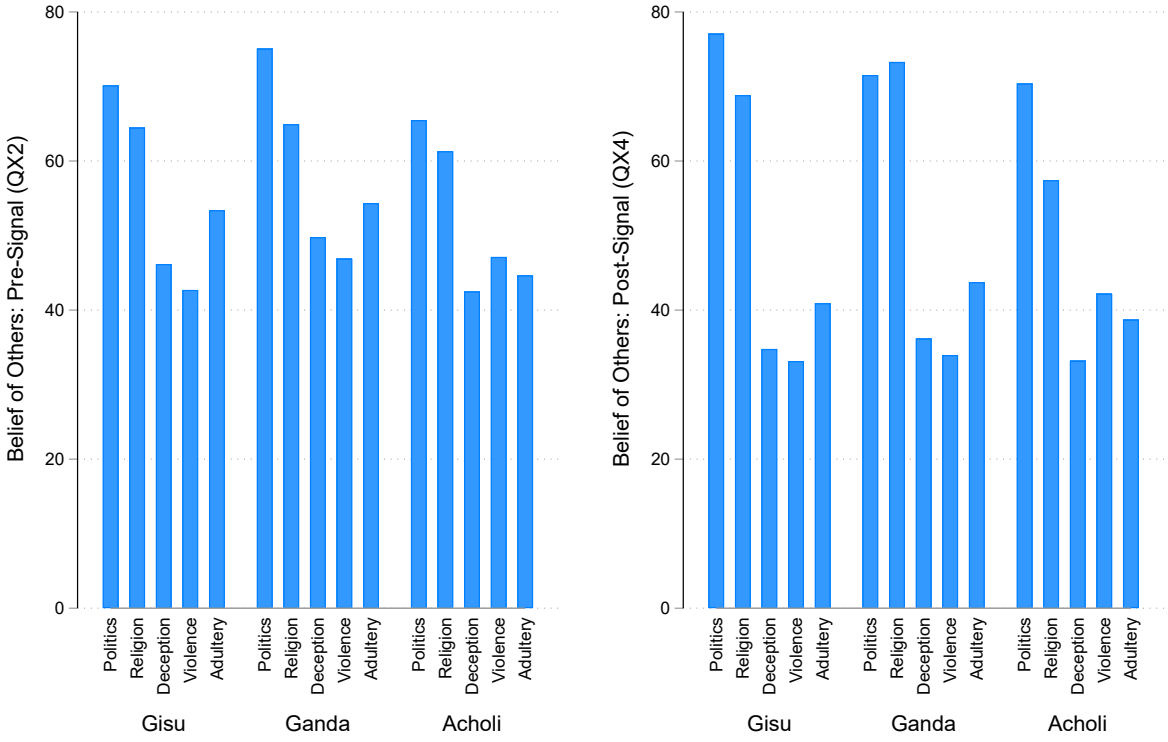


Figure 6.4 presents the data on second order beliefs pre (QX2) and post (QX4) the signal, which is summarised in table 6.8. Note the signal represents 10% of the responses actually provided in Phase One, so the answer of 9 for Q1.2 Gisu, if representative, would infer that participants should be answering 90 for Q1.4. Looking at the results from QX2 in the left panel of figure 6.4, it is clear that for all ethnic groups Politics and Religion are expected to have the highest number of participants answering six or higher. The ordering of the other three domains varies by ethnic group. Now considering the right panel (QX4), typically both Politics and Religion have slightly increased (though not for the Ganda), and Deception, Violence and Adultery have all decreased.

Table 6.8: 10% Phase One Signal, by Culture

Domain	**Question**	**Gisu**	**Ganda**	**Acholi**
Politics	Q1.2	9	7	8
Religion	Q2.2	8	9	5
Deception	Q3.2	2	2	2
Violence	Q4.2	1	1	5
Adultery	Q5.2	2	3	3

Figure 6.4: Summary of Second Order Beliefs: Pre and Post-Signal, by Domain, by Ethnic Group



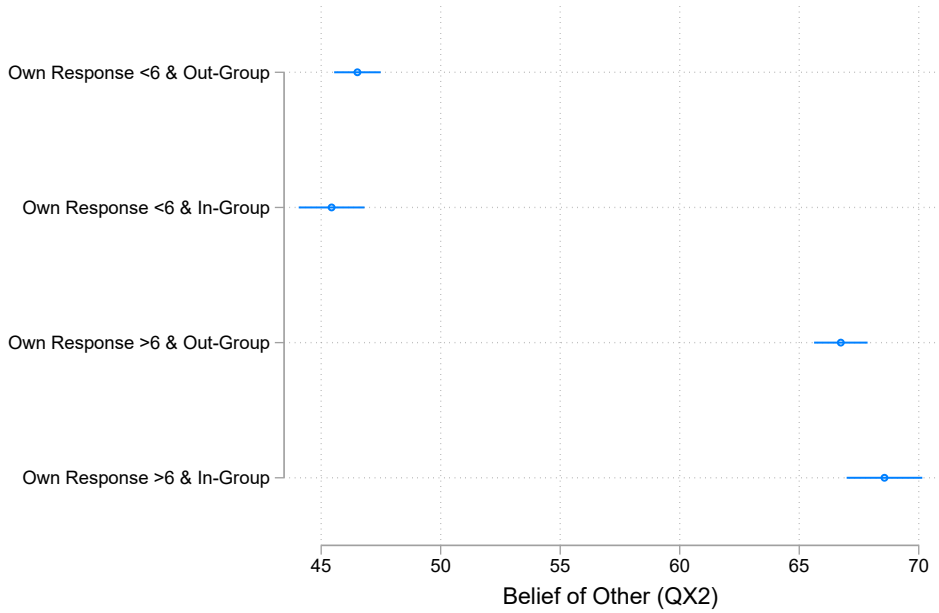
6.5 Results - Group 1: Conceptual Replication

The first family of tests explores whether measured own beliefs are more consistent with the in-group than the out-group. This conceptually replicates and expands the initial findings, which are predominantly in the political domain in the United States. The Gisu's first order (QX1) and second order beliefs (QX2) were measured across five domains (including politics) and three cultural groups. $Ingroup$ is a dummy equal to one if the in-group (Gisu) and zero otherwise (Ganda or Acholi). $Answer > 5.5$ is a dummy equal to one if the personal belief is above six (scale 1-10) and zero otherwise. $Interaction$ interacts $Ingroup$ and $Answer > 5.5$. The dependent variable for all specifications is second order beliefs (QX2) by domain, where responses represent the number of Phase One respondents participants believe answered six or higher for each domain (out of 100). As data is pooled, by five domains and then by three cultural groups, the original N=481 increases to N=7,215 for this group (N=1,443 for each domain family).

6.5.1 Family 1: Conceptual Replication

Figure 6.5 presents this visually, with the x axis showing the expected response to QX2 (second order beliefs) out of 100, pooled across all five domains. The y axis shows the four possible interactions. Graphically, it is evident that the largest driver of difference is $Answer > 5.5$, where the second order beliefs answering six or higher is substantially higher when the participant themselves believes the answer is higher than six. Visually, it can also be seen that, when the in-group equals one, the results become more extreme in both directions.

Figure 6.5: Replication: In vs Out-Group



Note: The dependent variable is QX2 pooled to include all five domains (second order beliefs answering six or above, out of 100). $Answer > 5.5$ equals one when the participant's answer to QX1 (first order belief) is above six (answers on a scale between 1-10). Ingroup equals one if Gisu, zero otherwise. Thus, a typical Gisu participant believes that almost 69 Gisu will answer six or higher (i.e. agree with them) whereas the out-group (Ganda and Acholi pooled) will only have around 66. Conversely, where the participant answered less than six for QX1, they typically believe only about 45 Gisu will answer six or higher in QX2, whereas they expect around 47 for the out-group.

Table 6.9 presents the full results. When the data is pooled (across all five domains), $Interaction$ is significant at the 1% level. This means that a participant answering higher or lower than the midpoint of 5.5 has a significant impact on the participant's second order beliefs. Considering the coefficients, where $Answer > 5.5$ equals one, participants on average believe 20.26 more people answered six or higher. Additionally, where $Interaction$ equals one (i.e. Gisu and own belief higher than six), participants on average believe 2.80 more participants answered six or higher for QX2. However, when disaggregating across the five domains, Politics and Adultery are not significant, while Religion, Deception and Violence are significant at the 5% level. This result is consistent after correction for multiple hypotheses. This is surprising given that previous studies in the US have almost exclusively focused on the

political domain. Further, it can be noted that the coefficients vary significantly across domains, with violence having the largest $Interaction$ coefficient (8.12).

Table 6.9: Replication, by Domain

	All	Politics	Religion	Deception	Violence	Adultery
Interaction	2.905***	2.465	5.868**	6.091**	8.115**	3.205
	(3.88)	(0.50)	(2.40)	(2.45)	(2.23)	(1.43)
Answered>5.5	20.259***	9.152**	14.040***	18.817***	10.712**	8.701***
	(17.13)	(2.12)	(6.15)	(6.26)	(2.39)	(3.57)
Ingroup	-1.080**	-2.500	-3.519	-0.778	-4.815***	3.370***
	(-2.09)	(-0.52)	(-1.55)	(-0.92)	(-5.58)	(3.57)
Female	1.414	-0.251	1.943	3.763	0.652	-0.386
	(1.08)	(-0.18)	(1.20)	(1.68)	(0.26)	(-0.18)
Age	0.085	-0.022	0.076	0.0646	0.245*	0.035
	(1.13)	(-0.29)	(0.76)	(0.55)	(1.89)	(0.30)
Constant	43.340***	62.282***	48.279***	39.840***	39.051***	47.257***
	(19.14)	(14.10)	(14.21)	(11.03)	(9.64)	(12.97)
q values		0.663	0.049**	0.029**	0.049**	0.168

Note: For regression column, coefficients are shown with *t* statistics in parentheses. The dependent variable is second order beliefs (QX2). *** p<.01, ** p<.05, * p<.1.

Result 1: Second order beliefs are significantly influenced by own beliefs, and this is more extreme for the in-group than the out-group. This effect is not consistent across all domains, with only three of the five domains being significant.

6.6 Results - Group 2: Mechanisms for Belief

This section sets out the results for both the in and out-group mechanisms for second order beliefs. As per equation 6.1 Motivated Reasoning and Cultural Identity are potential mechanisms for both the in and out-group, whereas Projection Bias and Curse of Knowledge are only potential mechanisms for the in-group (as per equation 6.2).

6.6.1 Motivated Reasoning

For Motivated Reasoning, the dependent variable is the total $Update$ from QX2 to QX4 following the signal provided after QX3. This signal provides information about what 10% of Phase One participants actually answered (i.e. how many answered six or higher for each domain). The variable of interest is $signal - firstguess$, which is the difference between QX2 and the signal. If a participant perfectly updated (without bias), then $Update$ and $signal - firstguess$ should be equal, and thus the coefficient should be one. Table 6.10 provides a summary of the average $Update$ and

$signal - firstguess$ for both in and out-group, by domain and pooled. Two results are apparent. First, the actual update is consistently smaller than the update predicted by a pure Bayesian updating. Secondly, for Politics and Religion the original response to QX2 is lower than QX4, meaning an understatement in the number of people participants expected to answer six or higher. The opposite is true for Deception, Violence and Adultery. This relationship is consistent for both in and out-groups.

Table 6.10: Summary of Update and Signal-firstguess, by Domain

	Pooled	Politics	Religion	Deception	Violence	Adultery
In-Group						
Update	-4.35	6.98	4.42	-11.28	-9.47	-12.41
	(25.68)	(20.98)	(22.74)	(24.66)	(26.95)	(25.78)
signal-firstguess	-11.39	19.85	15.51	-26.18	-32.71	-33.21
	(35.56)	(20.00)	(23.47)	(28.96)	(30.45)	(27.44)
Original Belief (QX2)	55.39	70.15	64.49	46.18	42.71	53.21
	(28.33)	(20.00)	(23.47)	(28.96)	(30.45)	(27.43)
Updated Belief (QX4)	51.04	77.13	68.91	34.90	33.24	40.80
	(30.50)	(20.14)	(21.75)	(25.43)	(27.97)	(25.80)
Signal	44.00	90.00	80.00	20.00	10.00	20.00
N	2,405	481	481	481	481	481
Out-Group						
Update	-5.09	0.73	2.31	-11.34	-8.88	-8.25
	(24.19)	(19.30)	(22.46)	(24.01)	(27.65)	(23.61)
signal-firstguess	-10.22	4.71	6.87	-26.15	-17.02	-19.51
	(31.95)	(22.65)	(28.59)	(28.49)	(36.37)	(27.25)
Original Belief (QX2)	55.22	70.29	63.13	46.15	47.02	49.51
	(27.85)	(20.98)	(22.13)	(28.49)	(30.44)	(27.25)
Updated Belief (QX4)	50.13	70.97	65.44	34.81	38.14	41.26
	(28.11)	(18.88)	(22.59)	(24.97)	(27.80)	(23.67)
Signal	45.00	75.00	70.00	20.00	30.00	30.00
N	4,810	962	962	962	962	962

Note: Averages shown with standard deviation shown in parentheses. *N* shown are consistent for in and out-group specifications.

Family 2.1.1: In-Group

Considering the in-group results in table 6.11, we can see that the coefficient on the $signal - firstguess$ is significantly different to zero at the 1% level for all domains (N=481) and where data is pooled (N=2,405). Thus I can reject the null that the coefficient of $signal - firstguess$ is zero. Further, though not pre-registered, when running a post-estimation specification to test whether the coefficient of $signal - firstguess$ is equal to one, for all the domains and pooled data, there is evidence at the 1% level that $signal - firstguess$ is significantly different to one. For all domains, the coefficient is less than one, meaning that the update is consistently smaller (on average only 0.427 the size) of what would be expected by $signal - firstguess$. This provides evidence to support a systematic bias in how beliefs are being updated.

Table 6.11: Motivated Reasoning In-Group, by Domain

	Pooled	Politics	Religion	Deception	Violence	Adultery
signal-firstguess	0.427***	0.549***	0.536***	0.738***	0.474***	0.499***
	(19.10)	(9.59)	(10.37)	(12.03)	(11.54)	(11.71)
Age	0.050	-0.235***	-0.085	0.145	0.243**	0.188*
	(0.75)	(-2.69)	(-0.94)	(1.43)	(2.06)	(1.78)
Female	-1.673	2.305	-0.001	-3.350	-3.001	-3.685
	(-1.55)	(1.43)	(-0.00)	(-1.82)	(-1.45)	(-1.86)
Constant	-0.427	1.208	-1.544	-1.279	0.706	0.839
	(-0.03)	(0.43)	(-0.51)	(-0.44)	(0.21)	(0.25)
q values		0.009***	0.009***	0.009***	0.009***	0.009***

Note: For regression column, coefficients are shown with *t* statistics in parentheses. The dependent variable is $Update$. *** p<.01, ** p<.05, * p<.1.

Family 2.2.1: Out-Group

Moving to the out-group results in table 6.12, we can see that the coefficient of the $signal - firstguess$ is significantly different to zero at the 1% level for all domains (N=962) and where data is pooled (N=4,810). Further, though not pre-registered, when running a post-estimation specification to test whether the coefficient of the $signal - firstguess$ is equal to one, for all the domains and pooled data, there is evidence at the 1% level that the coefficient on the $signal - firstguess$ is significantly different to one. For all domains, the coefficient is less than one, meaning that the update is consistently smaller (on average only 0.444 the size) of what would be expected by $signal - firstguess$. There is strikingly little difference between the in and out-group, across all domains. This provides evidence that Motivated Reasoning is a consistent cognitive bias when considering the second order beliefs.

Table 6.12: Motivated Reasoning Out-Group, by Domain

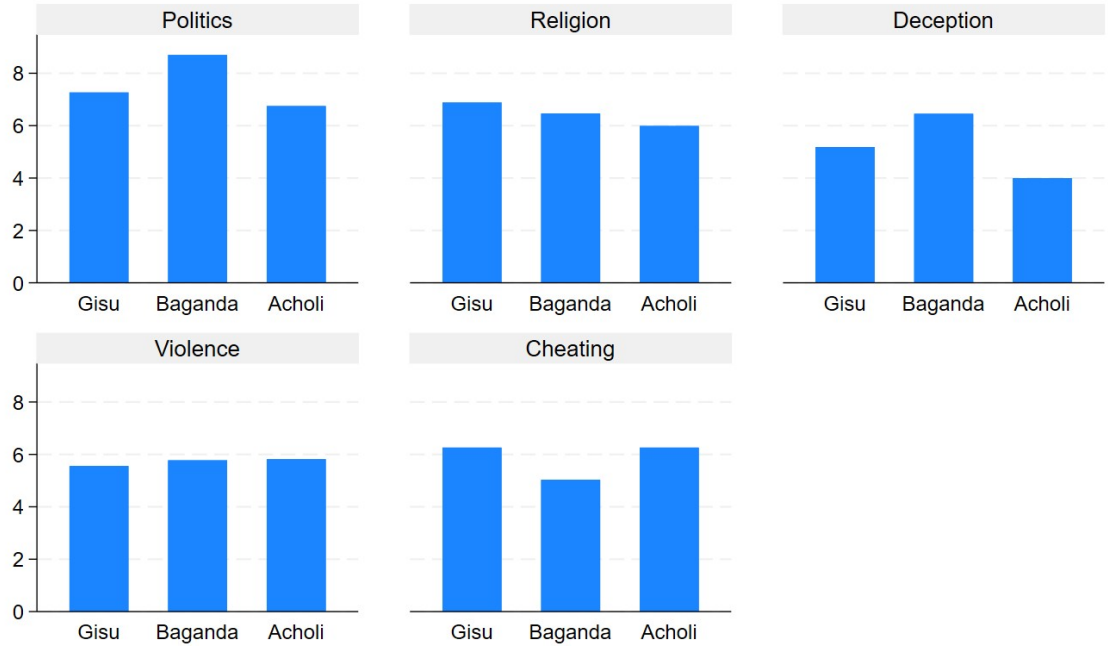
	All	Politics	Religion	Deception	Violence	Adultery
signal-firstguess	0.444***	0.487***	0.444***	0.469***	0.411***	0.501***
	(20.21)	(14.78)	(14.45)	(13.41)	(14.98)	(14.54)
Age	0.006	-0.177***	-0.051	-0.073	0.129	0.047
	(0.11)	(-3.04)	(-0.77)	(0.87)	(1.38)	(0.52)
Female	-1.274	-0.614	-0.374	-2.272	0.320	-3.473
	(-1.29)	(-0.53)	(-0.29)	(-1.36)	(0.18)	(-2.21)
Constant	-0.079	3.742	0.861	-0.012	-5.755**	1.999
	(-0.05)	(2.11)	(0.44)	(-0.00)	(-1.97)	(0.76)
q values		0.009***	0.009***	0.009***	0.009**	0.009***

Note: For regression column, coefficients are shown with *t* statistics in parentheses. The dependent variable is $Update$. *** p<.01, ** p<.05, * p<.1.

6.6.2 Cultural Identity

For Cultural Identity, the dependent variable is QX2, being second order beliefs. Within the specification, the coefficient on CI is the variable of interest, which measures the perception of stereotypes across five adjectives, closely associated with each of the five domains. Figure 6.6 sets out the average response, where the scale is 1-10, by domain and by ethnic group. There is clear variation both by domain and ethnic group, with Politics, Deception and Adultery seeing significant variation, whereas religion and violence are broadly consistent across all groups.

Figure 6.6: Cultural Stereotypes: Average Response, by Domain and Cultural Group



Averages of responses to Q6.1 “Faithful" (Cheating), Q6.4 “Aggressive" (Violence), Q6.6 “Very Religious" (Relgion), Q6.8 “Deceitful" (Deception) and Q6.10 “Political" (Politics), where responses were between 1 - 10. Ganda and Acholi are grouped as out-group, and Gisu forms the in-group.

Family 2.1.2: In-Group

Moving now to the detailed results for the in-group, The coefficient on CI is significant at the 1% level where data is pooled, as well as for Politics and Religion. Deception and Violence are significant at the 5% level, though there is insufficient evidence to reject the null for Adultery. These results survive the correction for multiple hypotheses. In terms of effect sizes, on average (where data is pooled) a 1.90 increase in stereotype rating corresponds to a increase of one in terms of responses to QX2. Religion has the smallest effect size, requiring a 2.52 increase in stereotype rating per one increase in QX2, whereas Deception and Violence require only a 0.93 and 0.88 increase respectively.

Table 6.13: Cultural Identity In-Group, by Domain

	Pooled	Politics	Religion	Deception	Violence	Adultery
CI	1.900***	1.436***	2.519***	0.933**	0.880**	-0.256
	(8.10)	(3.50)	(5.64)	(1.98)	(1.87)	(-0.53)
Female	1.015	-1.414	3.788	4.238	0.393	-0.887
	(0.69)	(-0.79)	(1.83)	(1.61)	(0.14)	(-0.35)
Age	0.075	-0.121	0.177	0.103	0.210	-0.026
	(0.87)	(-1.19)	(1.46)	(0.77)	(1.52)	(-0.19)
Constant	40.900***	63.897***	40.158***	36.217***	31.654***	56.230***
	(14.30)	(15.76)	(9.24)	(7.45)	(6.21)	(11.22)
q values		0.009***	0.009***	0.049**	0.049**	0.465

Note: For regression column, coefficients are shown with t statistics in parentheses. The dependent variable is $QX2$. *** p<.01, ** p<.05, * p<.1.

Family 2.2.2 Out-Group

Similarly, for the out-group (Ganda and Acholi), the coefficient CI is significant at the 1% level where data is pooled, and for all domains except Adultery. These results survive the correction for multiple hypotheses. In terms of effect sizes, on average (where data is pooled) a 1.73 increase in stereotype rating corresponds to a increase of one in terms of responses to QX2. Politics has the smallest effect size for the out-group, requiring a 2.30 increase in stereotype rating per one increase in QX2, whereas Deception and Violence require only a 0.85 and 0.94 increase respectively.

Table 6.14: Cultural Identity Out-Group, by Domain

	Pooled	Politics	Religion	Deception	Violence	Adultery
CI	1.725***	2.296***	1.854***	0.854***	0.939***	-0.526
	(11.32)	(7.91)	(5.95)	(2.71)	(2.67)	(-1.60)
Female	0.637	-0.426	2.494	2.847	0.111	-2.113
	(0.48)	(-0.29)	(1.45)	(1.21)	(0.04)	(-0.98)
Age	0.0876	-0.076	0.081	0.061	0.302	0.039
	(1.12)	(-0.92)	(0.83)	(0.51)	(2.26)	(0.32)
Constant	41.837***	54.922***	47.993***	38.471***	33.952***	52.467***
	(17.60)	(16.67)	(14.47)	(9.69)	(7.10)	(13.20)
q values		0.009***	0.009***	0.009***	0.009***	0.111

Note: For regression column, coefficients are shown with t statistics in parentheses. The dependent variable is $QX2$. *** p<.01, ** p<.05, * p<.1.

6.6.3 Family 2.1.3 Projection Bias (In-Group)

Projection Bias is only a possible mechanism for the in-group. For this mechanism, the dependent variable is $Agreement$ which is equal to QX2 (second order beliefs) if QX1 (own belief) is greater than six, and 100 - QX2 if QX1 is less than 6. In other words, how many Gisu does the participant expect to share the same views that they

themselves hold. $ProjectionBias$ (PB) is simply the answer to Q7, on a scale 1 - 10, measuring "how well do you feel other Gisu understand you and your beliefs?". The mean response was 5.90 (standard deviation 3.06, N = 481) across all participants.

Considering the full results in table 6.15 only Politics is significant at the 5% level, which survives the multiple hypothesis correction. In terms of effect size, which is relatively small, a 0.77 increase in PB corresponds to a one unit change in $Agreement$. To conextualise this, the mean of $Agreement$ is 59.66, with a standard deviation of 26.82.

Table 6.15: Projection Bias In-Group, by Domain

	Pooled	Politics	Religion	Deception	Violence	Adultery
PB	-0.116	0.770**	-0.252	-0.449	-0.320	-0.331
	(-0.50)	(2.29)	(-0.64)	(-1.01)	(-0.70)	(-0.74)
Female	-2.048	0.163	2.492	-6.918	-0.808	-5.167*
	(-1.49)	(0.09)	(1.19)	(-2.71)	(-0.29)	(-2.05)
Age	-0.007	-0.020	0.147	-0.013	-0.164	0.014
	(-0.11)	(-0.19)	(1.11)	(-0.10)	(-1.16)	(0.10)
Constant	62.546***	65.347***	61.280***	65.151***	65.660***	55.292***
	(22.12)	(15.52)	(11.72)	(12.42)	(11.10)	(10.00)
q values		0.020**	0.713	0.564	0.713	0.713

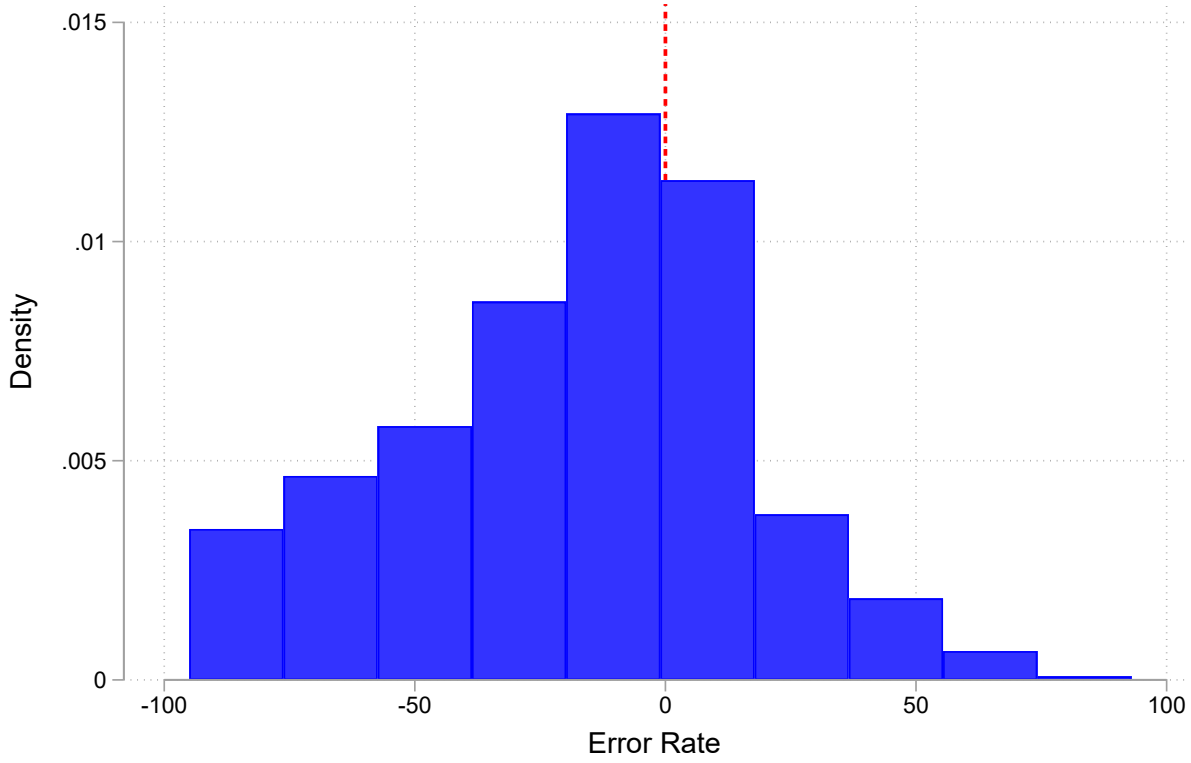
Note: Dependent variable is $Agreement$ for all specifications. For regression column, coefficients are shown with *t* statistics in parentheses. The dependent variable is $Agreement$. *** p<.01, ** p<.05, * p<.1.

6.6.4 Family 2.1.4 Curse of Knowledge (In-Group)

Curse of Knowledge is only a potential mechanism for the in-group. For this mechanism, the $ErrorRate$ is the dependent variable, which is computed as the difference between the actual answer provided to QX2 by the Gisu, Ganda and Acholi in Phase One and Two combined (see table 6.6) and QX2, by domain, by ethnic group. Figure 6.7 provides a high level distribution of the $ErrorRate$ pooled across all domains and ethnic groups (N=7,215). The mean $ErrorRate$ is -17.68 (standard deviation 34.16), meaning that across all data participants tend to overestimate the number of Phase One (and Phase Two for Gisu) responses which are above six.

The coefficient of interest for Curse of Knowledge is QX1, the participant's first order beliefs. If Curse of Knowledge is a mechanism, a participant's first order beliefs should not have any systematic impact on their error rate, which should instead be driven by uncertainty or other errors. Considering the detailed results in table 6.16, the coefficient of $QX1$ is significant at the 1% level across all domains. This result is consistent after the correction for multiple hypotheses. Note, as the Curse of Knowledge is only applicable for the in-group, these tests only include the data for the in-group.

Figure 6.7: Distribution of $ErrorRate$, Pooled



Note: Distribution of the $ErrorRate$ pooled across all ethnic groups and domains (N=7,215) where zero represents a response to QX2 which perfectly aligns to actual responses across Phase One and Two as per table 6.6. Negative deviations from zero occur where actual responses are lower than the participant's second order belief for a given domain and culture, meaning an under estimation of how many individuals answered above six for a specific domain, by ethnic group. Positive deviations are the opposite, where participants over estimate the number answering above six.

Table 6.16: Curse of Knowledge In-Group, by Domain

	Politics	Religion	Deception	Violence	Adultery
QX1	-2.193***	-2.865***	-3.442***	-2.837***	-1.631***
	(-3.26)	(-7.09)	(-7.30)	(-4.78)	(-3.80)
Age	-0.035	-0.139	-0.104	-0.156	-0.003
	(0.38)	(-1.13)	(-0.81)	(-1.12)	(-0.03)
Female	0.766	-2.666	-5.379**	-1.562	-0.851
	(0.43)	(-1.32)	(-2.13)	(-0.57)	(-0.34)
Constant	41.472***	40.576***	-20.826***	-27.478***	-35.415***
	(6.20)	(8.95)	(-4.95)	(-6.06)	(-8.01)
q values	0.009***	0.009***	0.009***	0.009***	0.009***

Note: For regression column, coefficients are shown with *t* statistics in parentheses. The dependent variable is the $ErrorRate$ for all specifications. *** p<.01, ** p<.05, * p<.1.

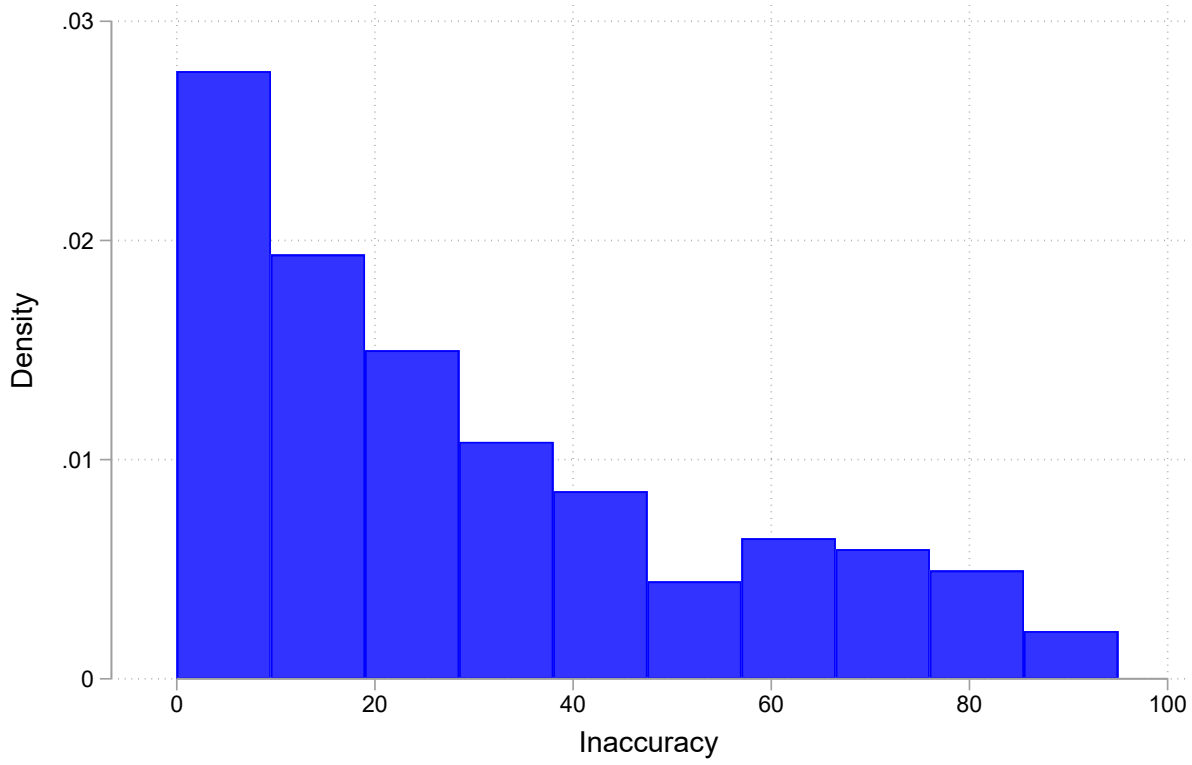
Result 2: For both the in and out-groups, there is strong evidence that Motivated Reasoning and Cultural Identity

are underlying mechanisms driving the beliefs of others. Motivated Reasoning is consistent across all five domains, whereas Cultural Identity is consistent across four domains. For the other in-group mechanisms, Curse of Knowledge is also significant across all domains, whereas Projection Bias is only significant for Politics.

6.7 Results - Group 3: Exploring Inaccuracy

As is evident in the section 6.6.4, there is significant variation in the error rate across the sample. However, $Inaccuracy$ is computed as the absolute of the difference between *actual* QX2 (from Phase One and Two for Gisu only) and QX2. As set out in figure 6.8, this means that $Inaccuracy$ has values between 0 and 99, with the mean of 29.7 and a standard deviation of 24.4.

Figure 6.8: Distribution of $Inaccuracy$, Pooled



Note: Distribution of $Inaccuracy$ pooled across all ethnic groups and domains (N=7,215) where zero represents a response to QX2 which perfectly aligns to actual responses across Phase One and Two as per table 6.6. 99 is the maximum possible level of inaccuracy as responses to QX2 were between 1 and 100. Inaccuracy is the absolute of the $ErrorRate$ as shown in figure 6.7.

Moving to the more detailed results in table 6.17, it is clear that Gisu have a higher level of inaccuracy when answering about other Gisu (i.e. the in-group) compared to their responses about the Ganda and Acholi (either

separately or jointly). The only exception is Adultery and Violence (when compared to the Ganda).

Table 6.17: Inaccuracy, by Domain, by Ethnic Group

	Pooled	Politics	Religion	Deception	Violence	Adultery
In-Group						
Gisu						
Inaccuracy	32.24	23.71	21.07	36.38	38.32	41.70
	(25.45)	(18.96)	(16.21)	(27.43)	(29.68)	(25.43)
N	2,405	481	481	481	481	481
Out-Group						
Ganda						
Inaccuracy	29.05	15.81	17.84	35.20	41.60	34.78
	(23.80)	(12.58)	(12.40)	(25.41)	(29.95)	(21.29)
N	2,405	481	481	481	481	481
Acholi						
Inaccuracy	27.93	17.49	18.55	32.19	36.71	34.71
	(23.69)	(13.99)	(13.73)	(26.13)	(26.81)	(26.21)
N	2,405	481	481	481	481	481
All Data						
Inaccuracy	29.73	19.01	19.15	34.58	38.88	37.07
	(24.39)	(15.78)	(14.26)	(26.38)	(28.90)	(24.61)
N	7,215	1,443	1,433	1,433	1,433	1,433

6.7.1 Family 3.1: Inaccuracy: In vs Out-Group

This family considers whether inaccuracy differs by the in or out-group. The dependent variable is $Inaccuracy$ for all specifications. Table 6.17 clearly shows that there is substantial variation between cultures and domains, but when values are pooled, the in-group has a higher overall average level of inaccuracy compared to the out-group (32.2 against 28.5, where zero is complete accuracy).

For all specifications, the coefficient of interest is $Ingroup$, where the in-group is coded to equal one, and the out-group to equal zero. Table 6.18 presents the detailed results, showing significant results at the 1% level for all specifications except Violence. These results are consistent after the application of the multiple hypothesis correction. For pooled data, participants are on average an additional 3.75 points more inaccurate for the in-group compared to the out-group. For all domains, the coefficient is positive and thus inaccuracy for the Gisu is higher than for the out-group (apart from Violence, which is not statistically significant). Thus, the data presented in tables 6.17 and 6.18 is at odds with the analysis presented by Bursztyn and Yang (2022), where individuals tends to be more accurate in their beliefs about in-groups.

Table 6.18: Inaccuracy: In vs Out-Group, by Domain

	Pooled	Politics	Religion	Deception	Violence	Adultery
Ingroup	3.748***	7.061***	2.871***	2.686***	-0.836	6.959***
	(10.21)	(8.72)	(4.18)	(3.45)	(-1.01)	(8.34)
Age	0.106	0.048	0.062	0.106	0.269**	0.048
	(2.00)	(0.73)	(0.98)	(0.99)	(2.21)	(0.46)
Female	0.017	-0.410	-1.371	3.114	0.167	-1.413
	(0.02)	(-0.40)	(-1.35)	(0.144)	(0.07)	(-0.75)
Constant	25.459***	15.511***	17.168***	29.048***	31.435***	34.133***
	(15.44)	(9.19)	(10.46)	(8.80)	(8.31)	(10.97)
q values		0.009***	0.009***	0.009***	0.2970	0.009***

Note: For regression column, coefficients are shown with *t* statistics in parentheses. The dependent variable is $Inaccuracy$ for all specifications. *** p<.01, ** p<.05, * p<.1.

6.7.2 Family 3.2: Inaccuracy: Self-Reported Confidence

This family considers whether inaccuracy differs by self-reported confidence (i.e. QX3, scale 1 - 10). The dependent variable is $Inaccuracy$ for all specifications. For all specifications in this family $QX3$ is the coefficient of interest, and is summarised in table 6.19. Comparing pooled data, across all domains, participants were most confident in their responses for Gisu (i.e. the in-group), with an average response of 6.99 (where 10 is complete confidence). This is closely followed by Ganda and then Acholi, with an average of 6.91 and 6.25 respectively. Across the domains there is significant variation, with participants feeling most confident in responses to Politics and Religion. Referencing table 6.17, across all data (N=7,215), Politics has the lowest level of inaccuracy (19.01), followed by Religion (19.15). Deception, Violence and Adultery are all significantly larger values, all above 30.

Moving to the detailed results in table 6.20, all specifications are significant at the 1% level, and this survives the correction for multiple hypotheses. Across all domains, an increase in confidence increases inaccuracy by an average of 0.40 basis points (scale 1 - 10) across all domains. However, there is some variation by domain, with the coefficient on $QX3$ being negative for Politics (-1.12) and Religion (-0.37). Overall Adultery has the largest coefficient at 1.73.

Table 6.19: QX3 Self Reported Confidence, by Domain, by Group

	Pooled	Politics	Religion	Deception	Violence	Adultery
In-Group						
Gisu						
Confidence	6.99	7.73	7.46	6.67	6.41	6.69
	(2.88)	(2.37)	(2.53)	(3.01)	(3.22)	(2.95)
N	2,405	481	481	481	481	481
Out-Group						
Ganda						
Confidence	6.91	7.93	7.11	6.61	6.39	6.51
	(2.83)	(2.45)	(2.55)	(2.88)	(3.01)	(2.91)
N	2,405	481	481	481	481	481
Acholi						
Confidence	6.25	6.81	6.43	5.95	6.19	5.86
	(3.00)	(2.75)	(2.82)	(3.11)	(3.11)	(3.12)
N	2,405	481	481	481	481	481
Total Out-Group						
Confidence	6.58	7.37	6.77	6.30	6.29	6.18
	(2.93)	(2.66)	(2.71)	(3.01)	(3.06)	(3.03)
N	4,810	962	962	962	962	962
All Data						
Confidence	6.72	7.49	7.00	6.41	6.33	6.35
	(2.92)	(2.57)	(2.68)	(3.02)	(3.11)	(3.01)
N	7,215	1,443	1,433	1,433	1,433	1,433

Table 6.20: Inaccuracy: Self-Reported Confidence, by Domain

	Pooled	Politics	Religion	Deception	Violence	Adultery
QX3	0.395***	-1.115***	-0.369*	1.861***	1.392***	1.727***
	(2.39)	(-5.57)	(-1.85)	(5.83)	(3.70)	(5.92)
Age	0.103*	0.071	0.068	0.088	0.274**	0.058
	(1.94)	(1.06)	(1.09)	(0.83)	(2.24)	(0.56)
Female	0.112	-0.563	-1.427	3.546	0.749	-0.981
	(0.10)	(-0.55)	(-1.41)	(1.70)	(0.32)	(-0.53)
Constant	24.093***	25.642***	20.547***	18.302***	21.907***	24.956***
	(13.33)	(11.10)	(9.73)	(5.12)	(5.11)	(6.81)
q values		0.009***	0.009***	0.009***	0.009***	0.009***

Note: For regression column, coefficients are shown with *t* statistics in parentheses. The dependent variable is $Inaccuracy$ for all specifications.
*** p<.01, ** p<.05, * p<.1.

6.7.3 Family 3.3: Inaccuracy: Cultural Distance

This family explores whether inaccuracy differs by cultural distance, as set out in table 6.5, where the Gisu are indexed at 1, the Ganda at 0.83 and the Acholi 0.58. $Cultdist$ subtracts one from these values so that Gisu are zero, the Ganda 0.17 and the Acholi 0.42. Thus, the larger $Cultdist$, the greater the cultural distance from the Gisu. The coefficient of interest is $Cultdist$. The dependent variable is $Inaccuracy$ for all specifications.

The detailed results are presented in table 6.21, where $Cultdist$ is significant for all domains at the 1% level. This survives the correction for multiple hypotheses. The coefficients are negative across all domains, with an average pooled value of -9.82. This means that an increase in cultural distance reduces inaccuracy, with the Violence (-5.00) and Adultery (-15.44) being the largest and the smallest and the domains.

Table 6.21: Inaccuracy: Cultural Distance, by Domain

	Pooled	Politics	Religion	Deception	Violence	Adultery
Cultdist	-9.822***	-13.202***	-5.341***	-10.127***	-5.006**	-15.435***
	(-9.30)	(-6.33)	(-3.01)	(-4.43)	(-2.04)	(-6.07)
Age	0.106	0.048	0.062	0.106	0.269**	0.0477
	(2.00)	(0.73)	(0.98)	(0.99)	(2.21)	(0.46)
Female	0.017	-0.410	-1.371	3.11	0.167	-1.413
	(0.02)	(-0.40)	(-1.35)	(1.46)	(0.07)	(-0.75)
Constant	28.640***	20.461***	19.176***	31.934***	32.141***	39.489***
	(16.96)	(11.95)	(11.39)	(9.56)	(8.33)	(12.56)
q values		0.009***	0.009***	0.009***	0.009***	0.009***

Note: For regression column, coefficients are shown with *t* statistics in parentheses. The dependent variable is $Inaccuracy$. *** p<.01, ** p<.05, * p<.1.

6.7.4 Family 3: Inaccuracy: Other Exploratory Specifications

Table 6.22 presents the results of additional exploratory specifications. For each, the dependent variable is shown in the column heading. The $UpdateRatio$ is the absolute ratio between (QX2 - QX4) and (QX2 - signal), where one would represent a perfect update in line with the signal. The $UpdateRatio$ is zero if no update was made following the signal. Of the pooled data across domains and groups (N=7,215), there are 816 instances (284 for in-group and 532 for the out-group) where the answer to QX2, QX4 and the signal are all equal. As these data points do not feature an update (and coding them as either zero or one would artificially skew the analysis) they have been excluded. All other variables are as previously defined in this section.

The first specification explores whether uncertainty differs across the in and out-groups, where $Ingroup$ is the coefficient of interest. This is significant at the 1% level, meaning that, on average, a Gisu will be 0.41 more certain

in their confidence level, compared to the out-group. This is consistent with the data presented in table 6.19. The second and third specifications explore whether the $UpdateRatio$ differs between the in and out-group ($Ingroup$) or by self-reported confidence level ($QX3$). Neither coefficient is significant.

Table 6.22: Inaccuracy: Other Exploratory Specifications

Dependent Variable	QX3	UpdateRatio	UpdateRatio
Ingroup	0.411***	-0.161	
	(7.23)	(-0.48)	
QX3			-0.001
			(-0.10)
Age	(0.008)	-0.003	-0.003
	(0.95)	(-1.07)	(-1.07)
Female	-0.238	0.159***	0.160***
	(-1.62)	(3.12)	(3.12)
Domain	-0.294***	-0.013	-0.013
	(-9.62)	(-1.04)	(-1.07)
Constant	7.367***	0.661***	0.661***
	(28.91)	(7.68)	(7.15)

Note: Coefficients are shown with *t* statistics in parentheses. The dependent variables are shown in column headers. *** p<.01, ** p<.05, * p<.1.

Result 3: Inaccuracy significantly differs across all domains between the in and out-group, however inaccuracy is actually highest for the in-group, which is the opposite relationship observed in US based political studies. Inaccuracy also differs significantly by self-reported confidence level, where an increase in confidence reduces inaccuracy. Inaccuracy differs significantly by cultural distance, across all domains, though an increase in cultural distance actually reduces inaccuracy. Lastly, inaccuracy differs between the in and out-group, with the in-group, on average, being more (erroneously) confident.

6.8 Discussion and Conclusion

The correction of misperceptions is the most commonly cited driver of social change, and the correction of such mis-calibrations has been attributed significant behavioural change, particularly in the short-term (Legros and Cislaghi, 2020). A recent meta-analysis, based mainly on the political domain in the United States, suggests that: 1) an individual's own attitudes and beliefs are strongly, positively correlated with perceptions of the in-group, but negatively correlated with out-groups; and 2) the accuracy of in-group perceptions tends to be higher when compared to out-groups (Bursztyn and Yang, 2022). Though there have been suggestions, the mechanisms supporting these findings are not understood, nor has the analysis been expanded to include a wider array of ethnic groups or domains. This chapter makes a number of important contributions to this literature.

The first contribution is a conceptual replication, asking whether measured own beliefs are more related to those of the in-group than out-group. The contribution of this is twofold. First, to explore whether the relationship observed in the United States, where the out-group is simply a US citizen of a different political persuasion, also holds when applied across different ethnic/cultural groups, and beyond a Western, Educated, Industrial, Rich and Demographic (WEIRD) context. Uganda offers this context and, while there is substantial variation between the Gisu, Ganda and Acholi, all groups are familiar and have a long shared history. Secondly, this chapter explores four additional domains, of increasing degrees of social taboo (i.e. violence and adultery), beyond politics. The conceptual replication confirms that second order beliefs are significantly influenced by own beliefs, and that this is more extreme for the in-group. This reinforces the importance of the more generalised finding, by providing evidence of its existence in a materially different context. The effect, however, is not uniform across all domains. Politics and Adultery do not significantly differ across in and out-group. Interestingly, Violence has the largest effect size, potentially due to the Gisu's tumultuous history and and reputation for aggression (Heald, 1989). These findings highlight that the cultural context of in and out-groups is relevant in determining which domains are likely to be affected: in the United States, political identity is a larger sub-set of an individual's wider identity than is typical in Uganda. Better understanding these subtleties is an area for further research.

The next contribution of this chapter is to investigate the mechanisms for belief, including those suggested by Bursztyn and Yang (2022). This is novel and provides some key insights for policy makers and potential future research. For both the in and out-groups, there is strong evidence that Motivated Reasoning and Cultural Identity are underlying mechanisms driving the beliefs of others. This is an important finding for two key reasons. Firstly, it shows that the effect identified in the conceptual replication is driven by both a cognitive bias and identity based errors. Secondly, specifically for Cultural Identity, there is variation between domains, revealing again that mechanisms need to be considered in conjunction with the relevant cultural context. This is equally true when considering the other in-group mechanisms, both Curse of Knowledge and Projection Bias have significant results, but these vary in terms

of the domains which are affected.

The second key finding from Bursztyn and Yang (2022) meta-analysis is that the accuracy of in-group perceptions tends to be higher when compared to out-groups. This chapter makes several contributions to this literature. Firstly, I replicate these findings, but across multiple domains and using more than one ethnic group. Inaccuracy does significantly differ across all domains between the in and out-group, however inaccuracy is actually highest for the in-group, which is the opposite relationship observed in United States based political studies. This result again highlights the importance of local context: the Gisu have had an unusual history, as well as a cultural identity which is low in trust with an out-group bias (Hargreaves Heap et al., 2012; Heald, 1989). Further, domains too play a part, as inaccuracy does not vary by in and out-group for Violence.

Further novel contributions to the literature are provided by investigating whether inaccuracy is significantly affected by either self-reported confidence levels or cultural distance. Self-reported confidence is shown to be significant across all domains, where an increase in confidence reduces inaccuracy. This relationship is true across the pooled data, but participants were most (erroneously) confident about the Gisu in-group, while actually being the least accurate for this group. Inaccuracy also differs significantly by cultural distance, across all domains, though an increase in cultural distance actually reduces inaccuracy. These are somewhat counterintuitive findings from a western stand point, but the Gisu were specifically selected to test previous findings in a more extreme context where cultural values, such as autonomy and independence, are the central pillars of identity (Heald, 1989). In-group bias is common and well documented, but it does not exist in all cultures. Clearly for the Gisu there is a degree of dissonance and a false belief that the in-group is better understood than it truly is. This is a common theme throughout the empirical chapters of this thesis and a valuable insight for the applicability of misperception corrections more generally in the developing world.

This chapter has a number of limitations which provide opportunities for further future research. The first limitation is that, while the Gisu were deliberately selected to 'stress test' the existing findings, this study still represents a small sample, and lacks the robustness of a cross sectional study. Thus, further data from a wider range of contexts is required before more significant inference can be draw from the findings, particularly given some of the more unusual Gisu specific findings. This is also true for the number of domains; while five domains is a substantial improvement on the typical focus on politics, future research should consider a broader range, with a specific focus on those which more directly support policy aims.

Chapter 7

Conclusion

With my PhD, I contribute to a better understanding of the determinants of human choice behaviour, and specifically how factors such as the language, context and misperceptions can have significant and large affects. By focusing on Uganda, I provide a valuable contribution to the broader non-Western, Educated, Industrialised, Rich and Demographic ('WEIRD') economic literature. I conduct multiple pre-registered (Hill, 2024; Clist and Hill, 2024, 2019) experiments between 2019 and 2024 across two geographies in Uganda (Kampala and Mbale), a country which is representative of how a significant proportion of the global population experience day to day life, as set out in Chapter 3 *'Uganda - Culture & Context'*. Social norms, values and preferences serve as the common thematic structure throughout, as shown in Chapter 2: *'Norms, Values & Preferences'*. This thesis contributes to the wider literature which shows that social norms can have a significant impact on individual decision making (Nosenzo and Görges, 2020; Gneezy et al., 2020; Abeler et al., 2019; Barr et al., 2018; Reuben and Riedl, 2013; Krupka and Weber, 2013; Gächter et al., 2013).

In my first empirical chapter, Chapter 4: *'Bilinguals in the Lab: Why do Norms and Expectations not Predict Contributions?'* I explored whether the language you speak affects the decisions that you make and, if so why. I replicated the study with the largest existing effect size and found that the total effect of my experiment (29.9%) was within 0.1% of the original Clist and Verschoor (2017) study. This was despite changing both the location (Mable to Kampala) and the sample demographic, demonstrating how robust such effects are. Secondly, this chapter explores the underlying mechanisms for the first time, ruling out both norms and expectations and concluding that cooperative preferences, at least in this instance, differ by language. The design of this experiment rules out a number of potential confounds, such as language structure. Rich anthropological evidence describes the Gisu as having a self-sufficient and non-cooperative culture. Speaking the associated language appears to activate this low-cooperation cultural frame, whereas speaking another leads to higher cooperation. These insights have potentially far reaching implications for policy makers. The majority of the world's population is both multi-lingual and use more than one language a day. This is particularly true for 'link' languages, such as Arabic, English, French, Hindi, Malay, Portuguese and Spanish, which connect populations of otherwise disconnected linguistic groups (Tucker, 1999). Understanding how choices changes when decisions are made in such languages is vitally important. A simple example is a nation's tax morale (Lago-Peñas and Lago-Peñas, 2010). If one language is associated with significantly higher levels of cooperation, this could be transformative in scenarios where simply changing the language of a tax return leads to higher tax revenues.

My second empirical contribution, Chapter 5: *'Hypothetical Norms & Cultural Exposure'* explored whether individuals house multiple sets of social norms and whether these can be elicited hypothetically across different contexts (in the spirit of Hoff and Stiglitz (2016)). I applied two contexts: geographical 'place' and 'frame'. I find that norms differ significantly by both place and frame, but that expectations do not. I also find a significant interaction between place and language for norms. These results are insightful, to both policy makers and future researchers alike. First, the methodological enhancements (extending Krupka and Weber (2013)) show that not only can hypothetical norms be measured out of context, but that these are accurate in signalling actual behaviours. This means that preferences

and norms for multiple locations or frames can potentially be determined from a single experiment. Equally, this significantly reduces the cost of policy makers collecting data to inform decisions, as well as assessing the likely outcomes of those decisions. Secondly, the interaction with language and norms reveals the complexity of real interactions between context and language, while also demonstrating the extended methodology is able to capture such nuances. This is an area ripe for further research and exploration. Lastly, this chapter also explored whether social norms change with cultural exposure within a domestic, rural to urban migration context. Despite norms being significantly different between Kampala and Mbale within-subject, exposure (being the amount of time a Gisu had spent in Kampala) was not a significant explanatory variable. Unlike international migration, where migrants tend to assimilate by learning new social norms over time (Berge et al., 2018; Cameron et al., 2015; Algan et al., 2012), this suggests rural to urban migrants are already aware of differences in norms before arriving. This may mean that rural to urban migrants, who maintain linkages with their original communities (Stites et al., 2014), may not integrate as readily, meaning sustained cultural fragmentation in large urban centres. My research on this was limited due to the availability of funds, but these open questions would benefit from further investigation.

My final experimental chapter, Chapter 6 *'Social Norms, Value & Misperceptions'* makes numerous contributions to our understanding of how misperceptions vary between in and out-groups. By expanding beyond the US political domain, I have investigated how the previously noted effects (Bursztyn and Yang, 2022) differ across multiple domains and ethnic groups (i.e. multiple out-groups), within a non-WEIRD context. This simple expansion of previous experiments has been insightful. Politics is significantly less important in Uganda than the US, and overall there is substantial variation between the degree an in-group misperceives an out-group by domain. Secondly, I have contributed to the literature by investigating four potential mechanisms driving the noted effect, that individuals believe the in-group is more like them than they really are, and out-groups are less like them than they really are. I find that there is strong evidence that Motivated Reasoning and Cultural Identity are underlying mechanisms driving both in and out-group effects, whereas Curse of knowledge is also significant for the in-group. This again is of value to policy makers, particularly those attempting to promote cultural integration and reduce social fragmentation. I also investigated the mechanisms underpinning the overall inaccuracies between groups. However, I also find that non-WEIRD groups, such as the Gisu, can hold beliefs which are not easily reconciled by Western standards and perceptions. For this reason, further research in this area would be valuable, specifically in a non-WEIRD setting.

To conclude, I hope that the findings of my PhD will be useful for both researchers and policy makers, and that they will inspire future research on the role of social norms, language and context on decision making, specifically in a non-WEIRD setting. Most importantly, I hope that the insights generated through this thesis can contribute in a some small way to the helping humanity find a path forward on some of the great challenges of our time.

References

Bibliography

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

Afrobarometer (2023). Merged Round 9 Data: (39 countries). `https://data.worldbank.org/indicator/SI.POV.DDAY`, Accessed: 17/10/2023.

Agarwal, S., Abuya, T., Kintu, R., Mwanga, D., Obadha, M., Pandya, S., and Warren, C. E. (2021). Understanding community health worker incentive preferences in uganda using a discrete choice experiment. *Journal of global health*, 11:07005.

Ahler, D. J. (2014). Self-fulfilling misperceptions of public polarization. *The Journal of Politics*, 76(3):607–620.

Ahler, D. J. and Sood, G. (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics*, 80(3):964–981.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Alekseev, A., Charness, G., and Gneezy, U. (2017). Experimental methods: When and why contextual instructions are important. *Journal of Economic Behavior & Organization*, 134:48–59.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8:155–194.

Alesina, A. and Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.

Alesina, A. and La Ferrara, E. (2000). Participation in heterogeneous communities. *The Quarterly Journal of Economics*, 115(3):847–904.

Alesina, A. and La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics*, 85(2):207–234.

Algan, Y., Bisin, A., and Verdier, T. (2012). Introduction: Perspectives on cultural integration of immigrants. *Cultural integration of immigrants in Europe*, pages 1–48.

Algan, Y. and Cahuc, P. (2013). Trust and growth. *Annual Review of Economics*, 5(1):521–549.

Allen, T. (2006). Trial justice: The international criminal court and the Lord's Resistance Army. *African Affairs*, 103(412):306–307.

Amelio, A. and Zimmermann, F. (2023). Motivated memory in economics—a review. *Games*, 14(1):15.

Asch, S. E. (2016). Effects of group pressure upon the modification and distortion of judgments. In *Organizational Influence Processes*, pages 295–303. Routledge. Originally published in New York: Russell and Russell, 1963.

Ashraf, Q. and Galor, O. (2013). The 'Out of Africa' hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, 103(1):1–46.

Atkinson, R. R. (1994). *The Roots of Ethnicity: The Origins of the Acholi of Uganda Before 1800*. University of Pennsylvania Press, Philadelphia.

Bahrami-Rad, D., Becker, A., and Henrich, J. (2021). Tabulated nonsense? Testing the validity of the Ethnographic Atlas. *Economics Letters*, 204:109880.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*.

Barr, A., Lane, T., and Nosenzo, D. (2018). On the social inappropriateness of discrimination. *Journal of Public Economics*, 164:153–164.

Barry, A. and Thrift, N. (2007). Gabriel tarde: Imitation, invention and economy. *Economy and Society*, 36(4):509–525.

Bayarri, M. and Mayoral, A. (2002). Bayesian design of "successful" replications. *The American Statistician*, 56(3):207–214.

Bayarri, M. J., Berger, J. O., Forte, A., and García-Donato, G. (2012). Criteria for bayesian model choice with application to variable selection. *The Annals of statistics*, 40(3):1550–1577.

Behrend, H. (2000). *Alice Lakwena and the Holy Spirits: War in Northern Uganda, 1985–1997*. Ohio University Press, Athens, Ohio.

Benjamin, D. J., Choi, J. J., and Strickland, A. J. (2010). Social identity and preferences. *American Economic Review*, 100(4):1913–28.

Bennett, N. J., Whitty, T. S., Finkbeiner, E., Pittman, J., Bassett, H., Gelcich, S., and Allison, E. H. (2018). Environmental stewardship: A conceptual review and analytical framework. *Environmental management*, 61:597–614.

Bennion, E. A. and Nickerson, D. W. (2011). The cost of convenience: An experiment showing e-mail outreach decreases voter registration. *Political Research Quarterly*, 64(4):858–869.

Berge, L. I. O., Bjorvatn, K., Galle, S., Miguel, E., Posner, D. N., Tungodden, B., and Zhang, K. (2018). Ethnically biased? Experimental evidence from Kenya. *Journal of the European Economic Association*.

Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, 47(2):241–262.

Berkman, L. F., Glass, T., Brissette, I., and Seeman, T. E. (2000). From social integration to health: Durkheim in the new millennium. *Social Science Medicine*, 51(6):843–857.

Bhandari, M. P. (2024). What is next for the Sustainable Development Goals, what are the challenges concerning SDG 10 — Reduced inequalities? *Sustainable Earth Reviews*, 7(1):23.

Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms - First Edition*. Cambridge University Press.

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms - Second Edition*. Cambridge University Press.

Bicchieri, C. (2010). Norms, preferences, and conditional behavior. *Politics, Philosophy & Economics*, 9(3):297–313.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bicchieri, C., Dimant, E., Gächter, S., and Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132:59–72.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026.

Binmore, K. (2010). Social norms or social preferences? *Mind & Society*, 9(2):139–157.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic theory*, 97(2):298–319.

Bonomi, G., Gennaioli, N., and Tabellini, G. (2021). Identity, beliefs, and political conflict. *The Quarterly Journal of Economics*, 136(4):2371–2411.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2017). Memory, attention, and choice. *The Quarterly Journal of Economics*.

Boroditsky, L. (2019). Language and the brain. *Science*, 366(6461):13–13.

Boyd, R. and Richerson, P. J. (1987). The evolution of ethnic markers. *Cultural Anthropology*, 2(1):65–79.

Branch, A. (2011). *Displacing Human Rights: War and Intervention in Northern Uganda*. Oxford University Press, Oxford.

Brigham, J. C. (1971). Ethnic stereotypes. *Psychological bulletin*, 76(1):15.

Brodeur, A., Cook, N., and Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, 110(11):3634–60.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review*, 110(10):2997–3029.

Bursztyn, L. and Yang, D. Y. (2022). Misperceptions about others. *Annual Review of Economics*, 14:425–452.

Camerer, C., Loewenstein, G., and Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of political Economy*, 97(5):1232–1254.

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280):1433–1436.

Camerer, C. F., Fehr, E., et al. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, 97:55–95.

Cameron, L., Erkal, N., Gangadharan, L., and Zhang, M. (2015). Cultural integration: Experimental evidence of convergence in immigrants' preferences. *Journal of Economic Behavior & Organization*, 111:38–58.

Campo, F., Nunziata, L., and Rocco, L. (2024). Business is tense: new evidence on how language affects economic activity. *Journal of Economic Growth*, 29(1):121–149.

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, 97(3):818–827.

Capraro, V., Halpern, J. Y., and Perc, M. (2024). From outcome-based to language-based preferences. *Journal of Economic Literature*, 62(1):115–154.

Cardenas, J. C. and Carpenter, J. (2008). Behavioural development economics: Lessons from field labs in the developing world. *The Journal of Development Studies*, 44(3):311–338.

Chang, D., Chen, R., and Krupka, E. (2019). Rhetoric matters: A social norms explanation for the anomaly of framing. *Games and Economic Behavior*, 116:158–178.

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review*, 103(2):690–731.

Chow, Y. S. and Teicher, H. (2003). *Probability theory: independence, interchangeability, martingales*. Springer Science & Business Media.

Chung, A. and Rimal, R. N. (2016). Social norms: A review. *Review of Communication Research*, 4:1–28.

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., and Reno, R. R. (2006). Managing social norms for persuasive impact. *Social Influence*, 1(1):3–15.

Cialdini, R. B., Kallgren, C. A., and Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in experimental social psychology*, volume 24, pages 201–234. Elsevier.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin*, 16(2):200–210.

Cialdini, R. B. and Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. *McGraw-Hill*.

Clarke, D. (2021). Rwolf2: Stata module to calculate romano-wolf stepdown p-values for multiple hypothesis testing. *Stata Journal*.

Clist, P. and Hill, J. (2019). Pre-analysis plan for a public goods game in urban Uganda. *AEA RCT Registry*, `https://doi.org/10.1257/rct.4616-1.0`.

Clist, P. and Hill, J. (2024). Pre-analysis plan for hypothetical social norms. *AEA RCT Registry*, `https://doi.org/10.1257/rct.13518-1.1`.

Clist, P. and Hong, Y.-Y. (2023). Do international students learn foreign preferences? The interplay of language, identity and assimilation. *Journal of Economic Psychology*, 98:102658.

Clist, P. and Verschoor, A. (2017). Multilingualism and public goods provision: An experiment in two languages in Uganda. *Journal of Development Economics*, 129:47–57.

Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., and Keysar, B. (2014). Your morals depend on language. *PloS one*, 9(4):e94842.

Crystal, D. (2012). *English as a global language*. Cambridge University Press.

Cullen, Z. and Perez-Truglia, R. (2022). How much does your boss make? The effects of salary comparisons. *Journal of Political Economy*, 130(3):766–822.

d'Adda, G., Drouvelis, M., and Nosenzo, D. (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, 62:1–7.

d'Andrade, R. G. (1995). *The development of cognitive anthropology*. Cambridge University Press.

Darwin, J. (1988). *Britain and Decolonisation: The Retreat from Empire in the Post-War World*. Macmillan, London.

de la Cuesta, B. and Wantchekon, L. (2016). Is language destiny? The origins and consequences of ethnolinguistic diversity in Sub-Saharan Africa. In *The Palgrave Handbook of Economics and Language*, pages 513–537. Springer.

De Saussure, F. (1916). *Course in General Linguistics*.

Desmet, K., Gomes, J. F., and Ortuño-Ortín, I. (2020). The geography of linguistic diversity and the provision of public goods. *Journal of Development Economics*, 143:102384.

Desmet, K., Ortuño-Ortín, I., and Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of Development Economics*, 97(2):322–338.

Devine, P. G. and Elliot, A. J. (1995). Are racial stereotypes really fading? The princeton trilogy revisited. *Personality and social psychology bulletin*, 21(11):1139–1150.

D'Exelle, B. and Verschoor, A. (2015). Investment behaviour, risk sharing and social distance. *The Economic Journal*, 125(584):777–802.

Diamond, J. M. (1998). *Guns, germs and steel: A short history of everybody for the last 13,000 years*. Random House.

DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology*, 23(1):263–287.

Durkheim, (1893). *The Division of Labor in Society*. Free Press. Reprinted in 1997.

Durkheim, (1895). *Rules of Sociological Method.* Free Press. Reprinted in 1982.

Eberhard, D. M., Simons, G. F., and Fennig, C. D. (2019). Ethnologue: languages of the world, Dallas. *Texas: SIL International*.

Eberhard, David M., G. F. S. and Fennig, C. D. (2022). *Ethnologue: Languages of the World. Twenty-fifth edition*. Dallas, Texas: SIL International.

Edgerton, R. B. (2010). *Sick societies*. Simon and Schuster.

Ellingsen, T., Johannesson, M., Mollerstrom, J., and Munkhammar, S. (2012). Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, 76(1):117–130.

Elster, J. (1989a). *The Cement of Society: A Study of Social Order*. Cambridge University Press.

Elster, J. (1989b). *Nuts and bolts for the social sciences*. Cambridge University Press.

Elster, J. (1989c). Social norms and economic theory. *Journal of Economic Perspectives*, 3(4):99–117.

Elster, J. (1991). Rationality and social norms. *European Journal of Sociology*, 32(1):109–129.

Enke, B. (2019). Kinship, cooperation, and the evolution of moral systems. *The Quarterly Journal of Economics*, 134(2):953–1019.

Epley, N. and Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic perspectives*, 30(3):133–40.

Erkut, H. (2020). Incentivized measurement of social norms using coordination games. *Analyse & Kritik*, 42(1):97–106.

Erkut, H., Nosenzo, D., and Sefton, M. (2015). Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters*, 130:28–31.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692.

Fallers, L. A. (1956). *Bantu Bureaucracy: A Study of Integration and Conflict in the Political Institutions of an East African People*. Cambridge University Press, Cambridge.

Fallucchi, F. and Nosenzo, D. (2022). The coordinating power of social norms. *Experimental Economics*, 25(1):1–25.

Fehr, E. and Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, 2(7):458.

Finnström, S. (2006). Wars of the past and war in the present: The Lord's Resistance movement/army in Uganda. *Africa: Journal of the International African Institute*, 76(2):200–220.

Fischbacher, U. and Gachter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1):541–56.

Fischhoff, B. (1975). Hindsight is not equal to foresight: the effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human perception and performance*, 1(3):288.

Flynn, D. J., Nyhan, B., and Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38:127–150.

Gächter, S., Gerhards, L., and Nosenzo, D. (2017). The importance of peers for compliance with norms of fair sharing. *European Economic Review*, 97:72–86.

Gächter, S., Nosenzo, D., and Sefton, M. (2013). Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association*, 11(3):548–573.

Galor, O., Özak, Ö., and Sarid, A. (2018). Geographical roots of the coevolution of cultural and linguistic traits. *National Bureau of Economic Research*.

Gatiso, T. T., Vollan, B., Vimal, R., and Kühl, H. S. (2018). If possible, incentivize individuals not groups: evidence from lab-in-the-field experiments on forest conservation in rural uganda. *Conservation Letters*, 11(1):e12387.

Gibson, J., McKenzie, D., Rohorua, H., and Stillman, S. (2020). Reprint of: The long-term impact of international migration on economic decision-making: Evidence from a migration lottery and lab-in-the-field experiments. *Journal of Development Economics*, 142:102391.

Gilovich, T., Savitsky, K., and Medvec, V. H. (1998). The illusion of transparency: biased assessments of others' ability to read one's emotional states. *Journal of personality and social psychology*, 75(2):332.

Girling, F. K. (1960). *The Acholi of Uganda*. Her Majesty's Stationery Office, London.

Giuliano, P. and Nunn, N. (2021). Understanding cultural persistence and change. *The Review of Economic Studies*, 88(4):1541–1581.

Glennerster, R., Miguel, E., and Rothenberg, A. D. (2013). Collective action in diverse Sierra Leone communities. *The Economic Journal*, 123(568):285–316.

Gneezy, U., Saccardo, S., Serra-Garcia, M., and van Veldhuizen, R. (2020). Bribing the self. *Games and Economic Behavior*, 120:311–324.

Golooba-Mutebi, F. and Hickey, S. (2008). The politics of chronic poverty: From the politics of exclusion to the politics of citizenship - insights from Uganda. *European Journal of Development Research*, 20(4):571–587.

Graham, J., Nosek, B. A., and Haidt, J. (2012). The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum. *PloS one*, 7(12):e50092.

Granholm, J. W. (1962). How to design a kludge. *Datamation*, 8:30–31.

Grosjean, F. (2021). *Life as a bilingual: Knowing and using two or more languages*. Cambridge University Press.

Habyarimana, J., Humphreys, M., Posner, D. N., and Weinstein, J. M. (2009). *Coethnicity: diversity and the dilemmas of collective action*. Russell Sage Foundation.

Haidt, J. (2013). Moral psychology for the twenty-first century. *Journal of Moral Education*, 42(3):281–297.

Hammarström, H., Forkel, R., Haspelmath, M., and Bank, S. (2021). glottolog/glottolog: Glottolog database 4.5.

Hansen, H. and Twaddle, M. (1988). *Uganda Now: Between Decay and Development*. James Currey, London.

Harari, Y. N. (2014). *Sapiens: A brief history of humankind*. Random House.

Hargreaves Heap, S., Verschoor, A., and Zizzo, D. J. (2012). A test of the experimental method in the spirit of Popper. *Journal of Economic Methodology*, 19(1):63–76.

Heald, S. (1989). *Controlling anger: the sociology of Gisu violence*. Manchester University Press.

Heald, S. (1995). *The Social Life of the Gisu: A Study of Gisu Culture in Eastern Uganda*. University of California Press.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). Most people are not weird. *Nature*, 466(7302):29–29.

Herbst, D. and Mas, A. (2015). Peer effects on worker output in the laboratory generalize to the field. *Science*, 350(6260):545–549.

Herz, H., Huber, M., Maillard-Bjedov, T., and Tyahlo, S. (2021). Time preferences across language groups: Evidence on intertemporal choices from the swiss language border. *The Economic Journal*, 131(639):2920–2954.

Hill, J. (2024). Pre-analysis plan for social norms, values and misperception. *AEA RCT Registry*, `https://doi.org/10.1257/rct.14167-1.0`.

Hodler, R., Srisuma, S., Vesperoni, A., and Zurlinden, N. (2020). Measuring ethnic stratification and its effect on trust in Africa. *Journal of Development Economics*, 146:102475.

Hoff, K. and Stiglitz, J. E. (2016). Striving for balance in economics: Towards a theory of the social determination of behavior. *Journal of Economic Behavior & Organization*, 126:25–57.

Hole, A. R. (2007). Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal*, 7(3):388–401.

Iliffe, J. (1998). *East African doctors: a history of the modern profession*. Cambridge University Press.

Iversen, V., Jackson, C., Kebede, B., Munro, A., and Verschoor, A. (2011). Do spouses realise cooperative gains? Experimental evidence from rural Uganda. *World Development*, 39(4):569–578.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22(1):129–146.

Jackson, M. O. (2011). An overview of social networks and economic applications. *Handbook of social economics*, 1:511–585.

Jakiela, P. and Ozier, O. (2018). *Gendered language*. The World Bank.

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Jonas, E. and Frey, D. (2003). Information search and presentation in advisor–client interactions. *Organizational Behavior and Human Decision Processes*, 91(2):154–168.

Jones, B. A. (2022). A review of social discounting: The impact of social distance on altruism. *The Psychological Record*, 72(3):511–515.

Kahan, D. M. (2015). The politically motivated reasoning paradigm, part 2: Unanswered questions. *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*, pages 1–15.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5):1449–1475.

Katz, D. and Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology*, 28(3):280.

Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R., et al. (2016). D-place: A global database of cultural, linguistic and environmental diversity. *PloS one*, 11(7):e0158391.

Kiwanuka, M. S. M. (1971). *A History of Buganda: From the Foundation of the Kingdom to 1900*. Longman, London.

Kölle, F. and Quercia, S. (2021). The influence of empirical and normative expectations on cooperation. *Journal of Economic Behavior & Organization*, 190:691–703.

Kollmuss, A. and Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, 8(3):239–260.

Kremer, M. and Levy, D. (2008). Peer effects and alcohol use among college students. *Journal of Economic perspectives*, 22(3):189–206.

Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6):1121.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3):480.

La Fontaine, J. (1967). Tribalism among the Gisu: an anthropological approach. *Tradition and Transition in East Africa: Studies of the Tribal Element in the Modern Era*, pages 177–92.

La Fontaine, J. S. (1969). Tribalism among the Gisu. In *Tradition and Transition in East Africa*, pages 177–192. University of California Press.

Lago-Peñas, I. and Lago-Peñas, S. (2010). The determinants of tax morale in comparative perspective: Evidence from European countries. *European Journal of Political Economy*, 26(4):441–453.

Lambarraa, F. and Riener, G. (2015). On the norms of charitable giving in Islam: Two field experiments in Morocco. *Journal of Economic Behavior & Organization*, 118:69–84.

Lapinski, M. K. and Rimal, R. N. (2005). An explication of social norms. *Communication Theory*, 15(2):127–147.

Legros, S. and Cislaghi, B. (2020). Mapping the social-norms literature: An overview of reviews. *Perspectives on Psychological Science*, 15(1):62–80.

Levendusky, M. S. and Malhotra, N. (2016). (Mis) perceptions of partisan polarization in the American public. *Public Opinion Quarterly*, 80(S1):378–391.

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives*, 21(2):153–174.

Lewandowsky, S., Ecker, U. K., and Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, 6(4):353–369.

Lewis, M. P. (2013). *Ethnologue: Languages of the world*. SIL international.

Li, K. K. (2017). How does language affect decision-making in social interactions and decision biases? *Journal of Economic Psychology*, 61:15–28.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Liberman, V., Samuels, S. M., and Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin*, 30(9):1175–1185.

Logan, C. (2022). Afrobarometer merged round 8 data (34 countries) (2022). https://www.afrobarometer.org/survey-resource/merged-round-8-data-34-countries-2022/.

Low, D. (1971). *The Mind of Buganda: Documents of the Modern History of an African Kingdom*. University of California Press, Berkeley.

Ly, A., Etz, A., Marsman, M., and Wagenmakers, E.-J. (2019). Replication bayes factors from evidence updating. *Behavior Research Methods*, 51(6):2498–2508.

Madarász, K. (2012). Information projection: Model and applications. *The Review of Economic Studies*, 79(3):961–985.

Mamdani, M. (1976). *Politics and Class Formation in Uganda*. Heinemann Educational Books, New York, NY.

Marsman, M., Schönbrodt, F. D., Morey, R. D., Yao, Y., Gelman, A., and Wagenmakers, E.-J. (2017). A bayesian bird's eye view of 'replications of important results in social psychology'. *Royal Society Open Science*, 4(1):160426.

Martin, C. L. (1987). A ratio measure of sex stereotyping. *Journal of Personality and Social Psychology*, 52(3):489.

Mehta, J., Starmer, C., and Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, 84(3):658–673.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. *American Economic Review*, 102(4):1508–1539.

Miguel, E. (2004). Tribe or nation? Nation building and public goods in Kenya versus Tanzania. *World politics*, 56(3):327–362.

Mildenberger, M. and Tingley, D. (2019). Beliefs about climate beliefs: the importance of second-order opinions for climate politics. *British Journal of Political Science*, 49(4):1279–1307.

Morris, M. W., Hong, Y.-y., Chiu, C.-y., and Liu, Z. (2015). Normology: Integrating insights about social norms to understand cultural dynamics. *Organizational Behavior and Human Decision Processes*, 129:1–13.

Mukwaya, P. and Bamutaze, Y. (2012). Rural-urban transformation in uganda. *International Food Policy Research Institute, Uganda*.

Murdock, G. P. (1967). Ethnographic Atlas: A summary. *Ethnology*, 6(2):109–236.

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49.

Newson, R. B. (2010). Frequentist q-values for multiple-test procedures. *The Stata Journal*, 10(4):568–584.

Nieuwenhuis, S., Forstmann, B. U., and Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature neuroscience*, 14(9):1105–1107.

Nosenzo, D. and Görges, L. (2020). Measuring social norms in economics: Why it is important and how it is done. *Analyse & Kritik*, 42(2):285–312.

Nunn, N. (2021). History as evolution. In *The handbook of historical economics*, pages 41–91. Elsevier.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*, 94(1):20–36.

Nyhan, B. and Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2):303–330.

Parsons, T. (1951). *The Social System*. Routledge. Reprinted in 2013.

Pepinsky, T. B. (2022). On whorfian socioeconomics. *Language*, 98(1):e44–e79.

Pérez, E. O. and Tavits, M. (2017). Language shapes people's time perspective and support for future-oriented policies. *American Journal of Political Science*, 61(3):715–727.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. Penguin UK.

Ray, B. C. (1991). *Myth, Ritual, and Kingship in Buganda*. Oxford University Press, Oxford.

Reid, R. (2017). *A History of Modern Uganda*. Cambridge University Press, Cambridge, United Kingdom.

Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1):122–137.

Richards, A. I. (1952). *Economic development and tribal change: A Study of Immigrant Labour in Buganda*. East African Institute of Social Research: Kampala, Uganda.

Roberts, S. G., Winters, J., and Chen, K. (2015). Future tense and economic decisions: Controlling for cultural evolution. *PloS one*, 10(7):e0132145.

Rogers, A. R. (1994). Evolution of time preference by natural selection. *The American Economic Review*, pages 460–481.

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., and Webb, M. D. (2019). Fast and wild: Bootstrap inference in stata using boottest. *The Stata Journal*, 19(1):4–60.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., and Zamir, S. (1991). Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. *The American Economic Review*, pages 1068–1095.

Rouder, J. N. and Morey, R. D. (2011). A bayes factor meta-analysis of bem's esp claim. *Psychonomic Bulletin & Review*, 18(4):682–689.

Royall, R. (2000). On the probability of observing misleading statistical evidence. *Journal of the American Statistical Association*, 95(451):760–768.

Rubinstein, A. (2000). *Economics and language: Five essays*. Cambridge University Press.

Rustichini, A. and Villeval, M. C. (2014). Moral hypocrisy, power and social preferences. *Journal of Economic Behavior & Organization*, 107:10–24.

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica*, 15(60):243–253.

Scheibehenne, B., Jamil, T., and Wagenmakers, E.-J. (2016). Bayesian evidence synthesis can reconcile seemingly inconsistent results: The case of hotel towel reuse. *Psychological Science*, 27(7):1043–1046.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18(5):429–434.

Schwartz, S. H. (1977). Normative influences on altruism. In *Advances in Experimental Social Psychology*, volume 10, pages 221–279. Academic Press.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In *Advances in Experimental Social Psychology*, volume 25, pages 1–65. Academic Press.

Schwartz, S. H. (2012). An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1):11.

Sharma, S. and Siddique, A. (2024). Non-weird preferences. *Available at SSRN 4755229*.

Simonsohn, U. (2013). Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological science*, 24(10):1875–1888.

Sseruyange, J. and Bulte, E. (2018). Do incentives matter for the diffusion of financial knowledge? experimental evidence from uganda. *Journal of African Economies*, 27(5):612–631.

Stephan, W. G., Ageyev, V., Stephan, C. W., Abalakina, M., Stefanenko, T., and Coates-Shrider, L. (1993). Measuring stereotypes: A comparison of methods using Russian and American samples. *Social Psychology Quarterly*, pages 54–64.

Stites, E., Burns, J., and Akabwai, D. (2014). It's better to sweat than to die: Rural-to-urban migration, Northern Karamoja, Uganda. *Somerville: Feinstein International Center Tufts University*.

Sunstein, C. R. (1996). Social norms and social roles. *Colum. L. Rev.*, 96:903.

Sutter, M., Angerer, S., Glätzle-Rützler, D., and Lergetporer, P. (2018). Language group differences in time preferences: Evidence from primary school children in a bilingual city. *European Economic Review*, 106:21–34.

Swidler, A. (2013). *Talk of love: How culture matters*. University of Chicago Press.

Tabellini, G. (2008). Institutions and culture. *Journal of the European Economic Association*, 6(2-3):255–294.

Tajfel, H., Turner, J. C., Austin, W. G., and Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader*, 56(65):9780203505984–16.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press, New Haven.

Tindigarukayo, J. K. (1988). Uganda, 1979–85: Leadership in transition. *The Journal of Modern African Studies*, 26(4):607–622.

Tripp, A. M. (2010). *Museveni's Uganda: Paradoxes of Power in a Hybrid Regime*. Lynne Rienner Publishers, Boulder, Colorado.

Tucker, G. R. (1999). A global perspective on bilingualism and bilingual education. *Washington, DC*.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., and Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory.* Basil Blackwell.

Twaddle, M. (1993). *Kakungulu and the Creation of Uganda, 1868-1928*. Ohio University Press, Athens, Ohio.

UBoS (2016). *The National Population and Housing Census 2014*. Kampala, Uganda.

UBoS (2024). National population and housing census 2024 preliminary tables, Uganda Bureau of Statistics. https://www.ubos.org/explore-statistics/20/, Accessed: 08/11/2024.

UN (2024). United Nations: Pact for the Future. https://www.un.org/en/summit-of-the-future/pact-for-the-future.

UNDESA (2024). World population prospects 2024. https://population.un.org/wpp/.

UNDP (2021). Gender Inequality Index - Uganda. https://hdr.undp.org/en/indicators/67106, Accessed: 08/11/2024.

UNDP (2023). Urbanisation in Uganda. https://population.un.org/wpp/, Accessed: 08/11/2024.

UNESCO (2021). Uganda literacy rate data. http://uis.unesco.org/en/country/ug, Accessed: 08/11/2024.

UNFPA (2022). United Nations Population Fund. https://www.unfpa.org/press/world-set-reach-8-billion-people-15-november-2022.

Verhagen, J. and Wagenmakers, E.-J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General*, 143(4):1457.

Warren, C., McGraw, A. P., and Van Boven, L. (2011). Values and preferences: Defining preference construction. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(2):193–205.

Weber, M. (1921). *Economy and Society: An Outline of Interpretive Sociology*. University of California Press. First printed in English 1968 (1921 in German), Reprinted in 1978.

Weber, T. O., Schulz, J. F., Beranek, B., Lambarraa-Lehnhardt, F., and Gächter, S. (2023). The behavioral mechanisms of voluntary cooperation across culturally diverse societies: Evidence from the US, the UK, Morocco, and Turkey. *Journal of Economic Behavior & Organization*, 215:134–152.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT press.

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*.

Wood, T. and Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, 41:135–163.

World Bank (2021). Access to electricity - Uganda. https://data.worldbank.org/indicator/EG.USE.PCAP.KG.OE?locations=UG, Accessed: 08/11/2024.

World Bank (2022). Uganda's political overview. https://data.worldbank.org/indicator/SI.POV.DDAY, Accessed: 08/11/2024.

World Bank (2023a). Health data - Uganda. https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=UG, Accessed: 08/11/2024.

World Bank (2023b). Population, total - Uganda. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=UG, Accessed: 08/11/2024.

World Bank (2024). Uganda's economic overview. https://www.worldbank.org/en/country/uganda/overview, Accessed: 08/11/2024.

Wrigley, C. C. (1996). *Kingship and State: The Buganda Dynasty*. Cambridge University Press, Cambridge.

Young, H. P. (1998). *Individual strategy and social structure*. Princeton University Press.

Young, H. P. (2015). The evolution of social norms. *Annual Review of Economics*, 7(1):359–387.

Young, R. (2001). *Postcolonialism: An Historical Introduction*. Blackwell, Malden, MA.

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, 6(3):299–310.

Appendices

Appendix 1 - 2019 Experimental Script

Experimental script (English)

Script: Kampala

Preparation of the experiment

Material needed:

- Pre-regristration sheets
- ID numbers
- A basket
- Two envelopes
- Money
- Paper and pen, data entry sheets, randomisation sheets
- Visual aids (manila paper)
- Box to collect tokens

Brief Explanation For Experimenters Only *We will play a public goods game in Luganda or Lugisu (randomly determined in advance of each session). Each subject will play it only once in one language and it is important that we don't draw extra attention to the fact that two different languages are used. This means all communication in a session must take place in the language of that session (even between experimenters). We need an even number of players in each session as players are paired. If there is an odd number players, choose the last person to arrive and give them a show-up fee of 9,000 Ugandan Shillings and send them away. It is important that subjects do not know who their partner is.*

Pre-screening *In advance of the experimental sessions, pre-screening sessions will have been completed in order to ensure that all participants have a Bagisu background and that they can speak Lugisu fluently (proficiency in Luganda is assumed, based on local information). During the pre-screening, candidates were asked questions to ensure fluency.*

As subjects arrive *When people enter the meeting room, they are asked for their name and student number as recorded during pre-screening. These should match exactly. If participants are not pre-registered, they will not be able to participate.*

Their name should then be recorded on the attendance register, alongside their game ID, which we give them on a card. We randomly match subjects, so these ID cards are important. At the end of the experiment, they hand in their card in exchange for their payment. The ID card allows us to identify them during the exercise while guaranteeing complete confidentiality. This is important, as they are able to earn real money in the exercise. Further instructions are given once sufficient people [10 - 30] have shown up.

Formal Introduction

Welcome. Thank you for taking the time to come today. [Introduce Experimenters and Assistants.] Later, you can ask any of us questions during today's programme. For this raise your hand so that we can come and answer your question in private.

We are from the Field Lab in Mbale, and are doing research in Kampala. We have invited you here today, because we want to learn about how people that are Bagishu, or from Bugishu region, make decisions. You are going to be asked to make decisions about money. The money that results from your decisions will be yours to keep.

What you need to do will be explained fully in a few minutes. But first we want to make a couple of things clear. First of all, this is not our money. We belong to a university in the UK, and this money has been given to us for research. Second, participation is voluntary. You may still choose not to participate in the exercise. Third, this is research about your decisions. Therefore you cannot talk with others. This is very important. I'm afraid that if we find you talking with others, we will politely ask you to leave, and you will not be able to earn any money here today. Of course, if you have questions, you can ask one of us. We also ask you to switch off your mobile phones.

Make sure that you listen carefully to us. You will be able to make some money here today, and it is important that you follow our instructions. During today's programme, you will be asked to make several decisions, which will be explained to you very clearly. Now, before we explain what you need to do, it is really important to bear one more thing in mind. The first decision that you will make is not a matter of getting it right or wrong. It is about what you prefer. It is important to think seriously about all your choices because they may affect how much money you can take home.

Explaining the Game

Today, you have randomly been paired with someone else in this room. You will not find out the identity of your partner, and they will not find out any information about you. All decisions are anonymous. However, we can tell you that your partner is (or was) also a student here at Makerere [*or other university*] and is either a Mugishu or spent time in the Bugishu region. I will explain all of the decisions slowly, and ask you to write down your answers the paper in front of you. You cannot change your answers after they've been written down, so think carefully before you write anything. Any questions that you have can be answered privately.

You will be given 12,000 Ugandan shillings, and you can decide what to do with it. First, we will demonstrate the decision using real money. You will make your choice on paper in front of you. You have two possible options: you can place money in either a private envelope (show) or a common basket (show). You can choose to put some of the money in the common basket, and the rest in the private envelope, but only in intervals of 3,000 Shillings. You can choose to keep this money for yourself, by placing it in the private envelope (show). This is your money to take home with you. There is also a common basket, which both you and you partner can put money in (show). We will add half of the money in the common basket (show). It will then be shared equally between the two players (show).

Recap, together [short, direct answers only]:

- What happens with any money you decide to put in your private envelope? [*You take it home*]
- How much is added to any money you and your partner put in the common basket? [*Half*]
- And after half is added, how do we split the money in the common basket between you and your partner? [*Equally*]

1. Control questions

We will now check for your understanding, using 3 examples. Imagine two people are paired: person A and person B. They would not know who they are paired with. [*Demonstrate with real money, and using the visual aid. Read question number, pause with each instruction to write down, indicating where to write down.*]

[Two of the following control questions should be solved collectively in the room with discussion. The third question is to be answered individually.]

1. Imagine that person A chooses to put nothing in the common basket, and everything in the private envelope. And imagine that person B chooses to put nothing in the common basket, and everything in the private envelope. Write on your paper, in the appropriate boxes, how much is in the common basket. How much is there after we have added half. And how much each player goes home with.
2. Imagine that person A chooses to put everything in the common basket, and nothing in the private envelope. And imagine that person B chooses to put everything in the common basket, and nothing in the private envelope. Write on your paper, in the appropriate boxes, how much is in the common basket. How much is there after we have added half. And how much each player goes home with.
3. Imagine that person A chooses to put everything in the common basket, and nothing in the private envelope. But, imagine that person B chooses to put nothing in the common basket, and everything in the private envelope.

Write on your paper, in the appropriate boxes, how much is in the common basket. How much is there after we have added half. And how much each player goes home with.

Thank you. These are just examples, you can decide what you prefer. When you make the decision, you can choose an amount, between 0 and 12,000 shillings, to put in the private envelope. You can choose an amount, between 0 and 12,000 shillings, to put in the common basket. Remember that we will pay you real money at the end of the experiment, depending on what you and your partner decide. Please make your choice now, by ticking once for question 4. [*Indicate where on visual aid*]

2. Expectations

We will now ask you questions 5-16 about behaviour in this game. Once you have made all of these decisions on this page, we will randomly pick one question. If you get the answer correct in the question we pick, we will give you another 4,000 shillings as a bonus. Let us remind you that it is very important that you do not talk during the experiment, and that you only mark one box per question. If you mark more than one box you will not be able to receive the bonus.

We will now ask you to make 4 guesses on what people decided in this game. If this question is chosen you could earn another 4,000 shillings on top of the money from the first section of the experiment. You would win the bonus if you are within 10% of the real answer. So, if your guess is good but not perfect, you will still get the bonus.

- For question 5, there are five boxes, each showing difference scenarios.
- The left hand box shows that the entire 12,000 shillings have been placed in the private envelope.
- The right hand box shows that the entire 12,000 shillings has been places in the common basket.
- In each of the boxes, in intervals of 10%, how many people in this room do you think contributed the amounts shown? For example, 0, 10, 20, 30, 40%?

We previously played this game with over one hundred, randomly selected people in Nakaloke sub-county, near Mbale.

- Again, for question 6, there are five boxes, each showing difference scenarios.
- The left hand box shows that the entire 12,000 shillings have been placed in the private envelope.
- The right hand box shows that the entire 12,000 shillings has been places in the common basket.
- In each of the boxes, in intervals of 10%, how many people in Mbale do you think contributed the amounts shown?

3. Norms

Now we will give a series of situations where someone made a decision. I will ask you to consider the different possible choices available and to decide, for each of the possible actions, whether taking that action would be "socially acceptable" and "consistent with moral or proper social behaviour" or "socially unacceptable" and "inconsistent with moral or proper social behaviour."

By socially acceptable, we mean behaviour that most people agree is the "correct" or "ethical" thing to do. Another way to think about what we mean is that if someone were to select a socially unacceptable choice, then someone else might be angry at them for doing so.

If this set of questions is chosen, you could earn another 4,000 shillings on top of what you earned in the first section of the experiment. You would earn that money if you give the same answer as the *most popular choice*. For these questions, we are not interested in your preferences. Rather, we are interested in what you think the *most popular* choice would be.

We will now go through an example. Imagine someone is at a local coffee shop near campus. While there, they notice that someone has left a wallet at one of the tables. Someone sees, and must decide what to do. They have four possible choices, and you need to rate how socially acceptable, "correct" or "ethical" that action is.

[*Read each choice out, ask 'how would you rate that action?' give the 4 possible ratings, and get experimenter 2 to answer using the below scale. Use visual aid throughout*]

	Very socially unacceptable	Socially unacceptable	Socially acceptable	Very socially acceptable
	- -	-	+	++
Take the wallet	X			
Ask others nearby if the wallet belongs to them			X	
Leave the wallet where it is		X		
Give the wallet to the shop manager				X

[Experimenter 2:] I think most people in this room would say 'action' is 'rating'. So I would tick *here*.

Now that we've gone through an example, we will turn to our questions. Remember that if you give the same answer as the most popular option, and if that question is randomly chosen, you could earn extra money.

For questions 7-11 in Kampala, for the people in this room, imagine someone put nothing in the private envelope

and everything in the common basket. Please rate this as either very socially unacceptable (–), somewhat socially unacceptable (-), somewhat socially acceptable (+) or very socially acceptable (++) by ticking once in that row.

You will see another four possible choices, where someone put either 3,000, 6,000, 9,000 or 12,000 in the private envelope. Please rate each choice as either very socially unacceptable (–), somewhat socially unacceptable (-), somewhat socially acceptable (+) or very socially acceptable (++). Remember, you can only get a bonus if you tick once per row.

For questions 12-16, imagine that instead of playing in Kampala, we are playing in Mbale. Everyone in this room is also playing the game, and that everything else is the same. The only difference is that we are imagining playing the game in Mbale. Please rate the same 5 choices as either very socially unacceptable (–), somewhat socially unacceptable (-), somewhat socially acceptable (+) or very socially acceptable (++). Remember, you can only get a bonus if you tick once per row.

4. Choice of bonus question

Now collect each the participants' answer sheets, with assistants writing ID number on answer sheets as they are collected. At the same time hand out the survey questions, again adding ID numbers. It is important that participants keep their ID cards as they will require this to collect their earnings at the end of the session.

Before participants begin to complete the survey questions, select someone at random to choose the bonus question blind from a cup. This cup should include the numbers from 5 - 16, denoting the different questions. Note questions 5 and 6 will need five intervals. There will be 20 questions to select from.

5. Survey

We will give you a new sheet. While we calculate your earnings, we'd like to ask a few general questions, to understand more about you. All information is anonymous, will not affect your earnings and is given voluntarily. If you wish not to answer a question, you are allowed to skip it.

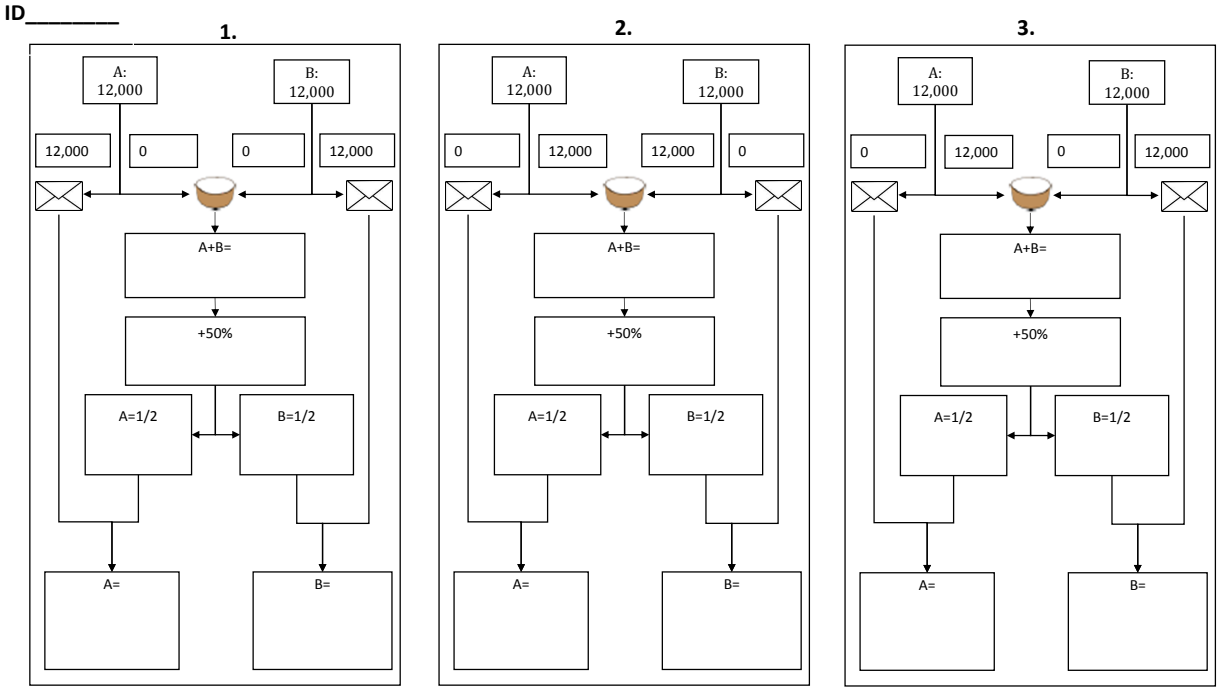
1. How old are you? [In years]
2. What is your gender? [Male/Female]
3. How many years, in total, have you spent in Kampala? [Answer in whole years]
4. How many people in this room do you know by name? [Please don't include the experimenters].
5. How many days in the last year have you spent in Bugishu region?
6. Which language did you learn first?

7. Where were you born?
8. Where is your father from?
9. Where is your mother from?
10. Which language do you prefer to speak at home?
11. How important is it to follow social rules, even if there is a cost? Please answer [1] not at all important, [2] not very important, [3] somewhat important or [4] very important.

Thank you very much for your answers. You have now all completed all of the tasks. We now invite you to come forward, one by one, to collect up your earnings. Thank you for coming today, your participation has been greatly appreciated.

Decision Sheets

Below is the first page of the decision sheet, showing the control questions.



On the following page, the answer sheet shows the contribution decision (4), the two sets of expectations (5 and 6) and norms (7 and 16).

ID_____

4.	12,000	9,000	6,000	3,000	0	
	☐	☐	☐	☐	☐	
	0	3,000	6,000	9,000	12,000	

		12,000 0	9,000 3,000	6,000 6,000	3,000 9,000	0 12,000
5.	Kampala	%	%	%	%	%
6.	Mbale	%	%	%	%	%

	Kampala	**--**	**-**	**+**	**++**
7.	: 12,000 : 0				
8.	: 9,000 : 3,000				
9.	: 6,000 : 6,000				
10.	: 3,000 : 9,000				
11.	: 0 : 12,000				

	Mbale	**--**	**-**	**+**	**++**
12.	: 12,000 : 0				
13.	: 9,000 : 3,000				
14.	: 6,000 : 6,000				
15.	: 3,000 : 9,000				
16.	: 0 : 12,000				

The final sheet shows where survey questions are answered. Note no language is used at any point.

ID:_____

1. _____

2. _____

3. _____

4. _____

5. _____

6. _____

7. _____

8. _____

9. _____

10. _____

	1	2	3	4
11.	☐	☐	☐	☐

Appendix 2 - Replication

Appendix 2 - Replication

This appendix sets out a number of tests which assess the incremental value of a replication to the existing body of work. This relevant as the initial contribution of this chapter is a replication. A growing body of complementary literature has emerged to determine whether a replication study can be deemed a statistical success or failure. In classical statistics the success of a replication is typically determined by comparing the p values of the replication and original study. However, the differences in experimental power (among other factors) can lead to counter-intuitive conclusions (Verhagen and Wagenmakers, 2014; Nieuwenhuis et al., 2011). While steps have been suggested to overcome these issues (small effects - see Simonsohn (2013); fixed-effects models - see Rouder and Morey (2011)), a consistent shortcoming of the classical approach is that it cannot quantify support in favor of the null hypothesis (being that of the sceptic). This appendix presents a range of statistical tests designed to assess the value of a replication. Each of these measures support that this study, either in isolation or when data is pooled with Clist and Verschoor (2017), presents very strong evidence of a language effect.

Bayesian approach to replication testing

Bayesian statistics offers a solution to the limitations of classical statistics. To illustrate this, consider a standard one-sample t test. Two dependent measures, Y_1 and Y_2, are assumed to be normally distributed and both the means and standard deviations are unknown. Where $\bar{y_i}$ is the mean score on measure Y_i, s is the standard deviation of the difference between Y_1 and Y_2 and N is the number of observations, the quantity $t = (\bar{y_1} - \bar{y_2})/(s/\sqrt{N})$ has a Student-t distribution with $df = N - 1$ degrees of freedom. Hence, the effect size $\delta = (\bar{y_1} - \bar{y_2})/s = t/\sqrt{N}$.

The null hypothesis is $\mathcal{H}_0 : \delta = 0$ and the alternative hypothesis is $\mathcal{H}_1 : \delta = \delta_1$. As the models are each defined with a single specific effect size, it is straightforward to compute (for each model) the probability of the observed data under the hypothesised effect size. The ratio of these probabilities is a likelihood ratio (Royall, 2000). The extent to which the observed data Y are more likely to have occurred under $\mathcal{H}_1$ relative to $\mathcal{H}_0$ is determined by the following:

$$LR = \frac{p(Y \mid \mathcal{H}_1)}{p(Y \mid \mathcal{H}_0)} = \frac{t_{df,\delta 1}\sqrt{N}(t_{obs})}{t_{df,}(t_{obs})} \tag{7.1}$$

However, the the exact alternative effect size δ_1 is never known beforehand. In Bayesian statistics, the uncertainty about δ_1 is addressed by assigning it a prior distribution. After assigning δ_1 a prior distribution, the immediate computation of the likelihood ratio (equation 1) is complicated by the fact that $\mathcal{H}_1$ is now a composite hypothesis, reflecting an entire distribution of effect sizes instead of a single value. Mathematically, this complication is overcome by computing a weighted average of likelihood ratios across all possible values of effect size, where

$p(Y \mid \mathcal{H}_1)$ serves as the weight function. To compute this weighted average, an integral is taken with respect to effect size. The resultant average likelihood ratio is better known as the Bayes factor (Verhagen and Wagenmakers, 2014):

$$
\begin{aligned}
B_{10} &= \frac{p(Y \mid \mathcal{H}_1)}{p(Y \mid \mathcal{H}_0)} \\
&= \frac{\int p(Y \mid \delta, \mathcal{H}_1) p(\delta \mid \mathcal{H}_1) d\delta}{p(Y \mid \mathcal{H}_0)} \\
&= \int \frac{t_{df,\delta} \sqrt{N}(t_{obs}) p(\delta \mid \mathcal{H}_1) d\delta}{t_{df,}(t_{obs})} \\
&= \int LR p(\delta \mid \mathcal{H}_1) d\delta
\end{aligned}
\tag{7.2}
$$

The result quantifies the strength of evidence that the data provides for $\mathcal{H}_1$ as opposed to $\mathcal{H}_0$, where values less than 1 support $\mathcal{H}_0$ (i.e. the view of a sceptic considering a replication has not successfully reproduced the original study). In terms of interpretation, a B_{10} of 10 suggests that it is 10 times more likely that the data occurred under $\mathcal{H}_1$ than $\mathcal{H}_0$. It should be noted that the outcome of the Bayes factor test is not a decision, but rather a grade of decisiveness. As a rule of thumb, factors between 1/3 and 3 are not considered to be conclusive evidence for either hypothesis (Jeffreys, 1998).

Bayes Factor Hypothesis Tests

Within the Bayesian statistics literature, there are multiple defensible ways in which to compare the results from both the original and replication study. Ly et al. (2019) summarise four common questions that may arise when evaluating replication studies. What evidence is there that the effect is present/absent in the replication study if:

1. The data of the original study is fully taken into account?
2. The data of the original study is completely ignored?
3. The data from the original study and replication are compared?
4. The data from the original study and replication are pooled?

Each approach makes use of a modified Bayes Factor to achieve a specific goal, as discussed in each section below. In order to assess the value and validity of our replication (being *whether session language affects contribution choice*), we present a range of Bayes factors in table 1. Table 1 uses this study and Clist and Verschoor's (2017) study,

which are similar in many respects, despite some changes to context and sample demographic[1].

Table 7.1: Analysis of replication

	Previous Evidence†	This Study‡
Effect size (δ)	0.23	0.22
T-Statistic	3.33	3.17
Sample size (N)	218	349
Bayes Factor Hypothesis Tests		
1. Original data entirely considered (B_{r0})	71.95	
2. Original data entirely ignored (JZS B_{10})	11.72	5.95
3. Original and replication data compared (Equality B_{01})	15.41	
4. Original and replication data pooled (Meta B_{10})	854.12	

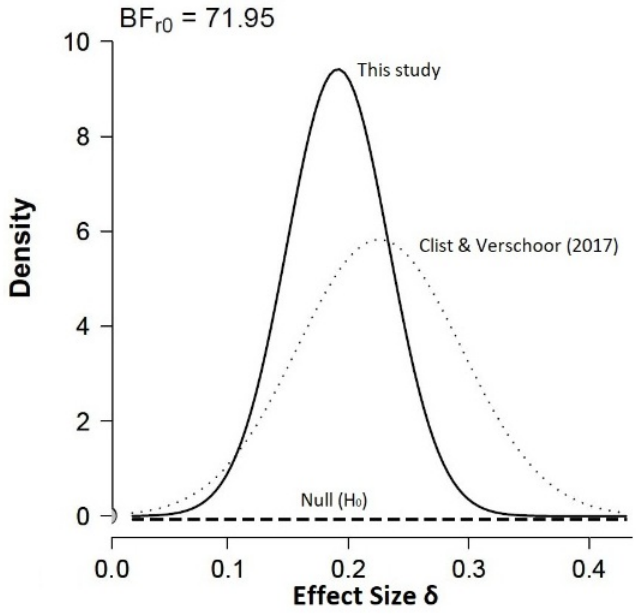
Note: † Data taken from Clist and Verschoor (2017). ‡ Data taken from this study. *Computed in `JASP` as per procedure set out in Ly et al. (2019). All stat computations supported by Verhagen and Wagenmakers (2014) R scripts, available at `http://www.josineverhagen.com`.

1. The data of the original study is entirely considered

Verhagen and Wagenmakers (2014) introduced an approach which pits against each other two hypotheses for the data of the replication. The first hypothesis (H_0) is the view of a sceptic, that the effect is spurious such that the effect $\delta = 0$. H_r is the proponent's replication hypothesis, an idealised belief that the effect is consistent with that found in the original study. Hence H_r is that $\delta \backsim$ the posterior distribution from original study. Figure 7.1 shows this graphically, with the distribution from the original study shown with the dotted line and the replication shown by the solid line. As there is substantial overlap (i.e. the area underneath the curves), there is strong evidence that the replication supports the findings of the original study, hence a large Bayes factor (72), meaning that the replication result is 72 times more likely to be consistent with the original Clist and Verschoor (2017) than the null H_0.

[1]Clist and Verschoor's (2017) study was conducted in Nakaloke sub-county of Mbale, a rural and predominately agricultural region in Eastern Uganda. Subjects were adults randomly selected based on village and parish. This study was conducted in Kampala, the capital of Uganda and an urban region. Adult subjects were recruited based on ethnic group (Gisu), supported by student social group membership at Makerere and Kyambogo Universities.

Figure 7.1: Analysis of Replication



Note: Dotted line represents distribution from original Clist and Verschoor (2017) study, solid line represents our replication. BF_{r0} represents Verhagen and Wagenmakers (2014) Bayes factor, which takes into account the original study entirely.

2. The data of the original study is completely ignored

If the original study is ignored, is there evidence that the effect is present or absent in the replication attempt? The most popular model for this question is the Jeffreys-Zellner-Siow (JZS) prior (Marsman et al., 2017; Liang et al., 2008; Bayarri et al., 2012). The test compares the null hypothesis (H_0) that effect size δ is zero against an alternative hypothesis (H_1) that effect size is not zero. In my analysis, I calculate the JZS for both the original Clist and Verschoor (2017) study and this study, hence the two values in table 1 providing support against the null effect of 0. In table 1, values of B_{10} greater than 1 supports the alternative hypothesis, were greater than 3 constitutes strong support. Thus a factor of 6 means that the effect is 6 times more likely to be non-zero than zero.

3. The data from the original study and replication are compared

When comparing data from the original study and the replication attempt, what is the evidence that the effect sizes are similar or dissimilar? Bayarri and Mayoral (2002) propose a Bayes factor test, where H_0 states that the effect sizes of the original and replication are equal against the alternate hypothesis (H_1) that they are not. In table 1, Equality B_{01} shows the Bayes factor for this test, with a value substantially above 3, suggesting strong support for the idea the effect

sizes are equal (Jeffreys, 1998). Evidence in favour of the null indicates a successful replication. Thus a factor of 15 means that the effect is 15 times more likely to be non-zero than zero.

4. The data from the original study and replication are pooled

Pooling the data from the original study and the replication attempt, what is the evidence that the effect is present or absent? This test assumes that there is a single correct underlying effect size and that the replication and original experiment are exchangeable (Scheibehenne et al., 2016; Rouder and Morey, 2011). Essentially, this approach takes the products of JZS approach (which computes each study as an individual result), over the number of studies. The resulting meta-analytic Bayes factor (Meta B_{10} in table 1), when high, indicates evidence against the null (H_0), where H_0 is that the true effect size is absent, versus H_1, that it is present. Given the size of Meta B_{10}, there is very strong evidence against the null. Thus a factor of 854 means that the effect is 854 times more likely to be non-zero than zero.

Conclusion

When assessing the overall value of a replication, the Bayesian statistical tests set out in this appendix provide a number of approaches which can avoid the shortcomings of classical testing approaches. Consistent across each of the four tests, the effect sizes are similar, with the effect seen in this chapter very similar to previous work. Thus, this provides confidence in the value of the replication, in addition to providing further support that the observed language effects are robust.

Appendix 3 - Experimental Script

Stage 1 - Regional Survey

Instructions for enumerators in italics. Instructions to be read in normal font. Note: The pre-registered script also included two additional ethnic groups, the Teso and Basoga. As set out in the PAP, these groups were dropped to reduce experimental complexity following piloting.

Locations: Mbale (Gisu; N=100), Kampala (Ganda, N=100) and Gulu (Acholi, N=100).

BRIEF EXPLANATION FOR EXPERIMENTERS ONLY:

This stage is a simple survey of three distinct 'groups': Gisu, Ganda and Acholi. The survey will need to be conducted in the local language of each tribe and physically carried out in the appropriate geographical region. Local language scripts need to be back translated to ensure consistency between each survey. Recruitment should be random and restricted to consenting adults (over 18 years old). Participants should be made aware that all responses will be anonymised.

Pre-screening: For each 'group' I am aiming to collect 100 sets of responses, 300 sets in total. To ensure participants have the appropriate group membership, the following pre-screening questions should be asked (verbally) in the local language:

- *Which tribe/ethnic group do you most identify with?*
- *Where were you born?*
- *Are you fluent in [local language]?*

Enumerator(s) should assess this evidence, coupled with spoken language proficiency, to determine whether the individual is an ethnic Gisu, Ganda, or Acholi, as appropriate. Otherwise, we should thank the applicant for their interest and inform them that they do not meet the criteria.

PREPARATION OF THE SURVEY

Material needed:

- *Question Sheet (see below)*

- *Answer Sheet (See below)*
- *Money*
- *Anonymous submission box*

FORMAL INTRODUCTION

Welcome. Thank you for taking the time to come today. *[Introduce Enumerators/Assistants]*. We are from The Field Lab, Mbale and are here on behalf of researchers from a UK University to measure and better understand different groups in Uganda. We will ask you five brief questions, which should take about 5 minutes. For your time we will give you UGX 5,000. All questions are optional, you may choose not to answer the questions or withdraw at any time. This will not cost you anything, but you will only receive the UGX 5,000 if you answer all of the questions. All your responses will be anonymised. We will not observe you completing you form, and when you are finished, we ask that you place your responses in the box here *[indicate]* so that they remain anonymous. There is no way for us to identify which answers belong to you, so please answer truthfully and what you really think.

THE SURVEY

Questions should be asked verbally, with the enumerator making responses on an answer sheet. Add a sequential ID number to the sheet. Please mark all responses on the provided answer sheet. Once the survey has been completed, pay the participant UGX 5,000.

Questions:

Question	Range
QS.1 Do you think politics is important?	[1-10]
QS.2 Do you think religion is the most important aspect of identity?	[1-10]
QS.3 Do you think it is acceptable to deceive others?	[1-10]
QS.4 Do you think the use of violence is acceptable?	[1-10]
QS.5 Do you think cheating on your spouse is acceptable?	[1-10]

Answer Sheet:

Question	✗	1	2	3	4	5	6	7	8	9	10	✓
QS.1		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
QS.2		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
QS.3		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
QS.4		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
QS.5		☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	

Stage 2 - Mbale Lab-in-the-Field

Locations: Mbale Region (Gisu, **Target** N=300)

Pre-screening: The sampling frame set out in section 2.2.1 of the "Pre-Analysis Plan for Social Norms, Values and Misperception" should be followed to select a sample, as discussed in planning meetings with The Field Lab.

Phase One - Survey

PREPARATION OF THE SURVEY

Material needed:

- *ID numbers for tables*
- *Question Sheet*
- *Answer Sheets 1.0, 1.1 and 1.2*
- *Handout 1.0*
- *Payout computation & data entry sheet (excel based)*
- *Bag with 30 pieces of paper with question numbers [1-5.2 and 1-5.4, across five cultures] for random participant to select bonus question*

BRIEF EXPLANATION FOR EXPERIMENTERS ONLY:

This stage is an incentivised survey which builds on Stage 1. Some of the questions ask the participant's own opinions, while others ask the participant to record their belief of 'others'. The 'other' (X) refers to the Gisu, Ganda and Acholi (i.e. this will include their own group (Gisu), plus the other two). The answer sheets have been designed to make clearer which questions require a response relating to these groups.

All subjects will have been screened for language proficiency during sampling to ensure they are ethnic Bagisu – see questions in Stage 1 and use sampling frame as set out in consultation with TFL. As recruitment occurred a different point to the experiment, participants will need to confirm their name and telephone number when arriving for the experiment. If there is any doubt that the participant is not the selected individual, they should not be admitted.

Prior to participant admittance, place ID numbers and answer sheet 1.0 and 1.1 on tables. Participants can select any seat randomly.

FORMAL INTRODUCTION

Welcome. Thank you for taking the time to come today [introduce Enumerators and Assistants]. Later, you can ask any of us questions during today's programme. For this raise your hand so that we can come and answer your question in private.

We are from The Field Lab, based here in Mbale. We have invited you here today, because we want to learn about the beliefs and decision making of different Ugandan groups. You are going to be asked to answer some questions, most of which could earn you money. This money will be yours to keep.

What you need to do will be explained fully in a few minutes. But first we want to make a couple of things clear. First of all, this is not our money. We are associated with a university in the UK, and this money has been given to us for research. Second, participation is voluntary. You may still choose not to participate in the exercise or withdraw at any time by simply leaving the room. If you choose to remain in the room, you are signalling to us your consent to participate in this experiment. Please raise your hand to confirm that you have understood this and that you wish to signal your consent.

If any participant does not raise their hand, an enumerator should privately discuss with the participant whether they understand or have concerns. During screening, written consent forms should be completed by the participant in Lugisu (as this is the language of the session).

If you choose not to answer a question, you will still receive a flat participation fee of UGX 1,500, but you will not be eligible for the additional payouts associated with specific questions. This will become clearer as we explain the tasks. Third, this is research about your decisions and individual beliefs. Therefore, you cannot talk with others. This is very important. I'm afraid that if we find you talking with others, we will politely ask you to leave, and you will not be able to earn any money here today.

Of course, if you have questions, you can ask one of us. We also ask you to switch off your mobile phones. Make sure that you listen carefully to us. You will be able to make some money here today, and it is important that you follow our instructions and carefully consider your responses, as it will impact how much you will earn. During today's programme, you will be asked to answer some questions and make several decisions, which will be explained to you very clearly.

THE SURVEY

First, we are going to ask you some questions about your opinions and your perception of other ethnic groups. For ten of the questions *(QE1-5.2 and QE1-5.4)*, we will ask you to try and guess what members of other groups answered. We recently interviewed 100 Gisu, Baganda and Acholi and asked them to answer these questions. These interviews were completed on the streets of either Mbale, Kampala or Gulu, where passers-by were randomly selected. All responses were collected blind and placed in a sealed box so that responses could not be traced back to individual interviewees. For these questions, there is an opportunity to earn money. At the end of the survey, we will ask one of you to randomly select which one of these questions will be eligible for a 5,000-shilling bonus *[use QE1.2 as an example of how this works]*. If you have correctly guessed within 10 persons, either way, you will receive the payout. For example, if you guess that 60 people answered six or higher, so long as the real answer is between 50 and 70, then you will get the bonus payout if that question is selected. This range is always ten people higher and lower than your guess. It is important that you use the correct range shown, otherwise you will not be eligible for the bonus payout.

Remember all of your answers are completely confidential. For each question, please use the answer sheet to mark your responses. Before you start to answer, please ensure you have written your ID number at the top of your answer sheet. We will not be able to calculate your payout without this.

We will now read each question in turn: *Enumerators verbally ask each question, making clear the response range. Refer to answer sheets which also contain this information and the number of responses required. Where shown, a response is required for all three (X) cultural groups. For Q.E.1-5.2, ask the participants to mark their answers on both Answer Sheets 1.0 and 1.1 (they will retain 1.1 after 1.0 has been collected).*

Question	Range
QE.1.1 Do you think politics is important?	[1-10]
QE.1.2 We asked 100 `X' "do you think politics is important?" on a scale 1-10. How many do you think answered 6 or higher? We will give you 5,000 shillings if you are within 10 people of the correct answer.	[0-100]
QE.1.3 Where 1 = not at all confident and 10 = completely confident, how confident do you feel about your answer to QE.1.2?	[1-10]
QE.2.1 Do you think religion is the most important aspect of identity?	[1-10]
QE.2.2 We asked 100 `X' "do you think religion is the most important aspect of identity?" on a scale 1-10. How many do you think answered 6 or higher? We will give you 5,000 shillings if you are within 10 people of the correct answer.	[0-100]
QE.2.3 Where 1 = not at all confident and 10 = completely confident, how confident do you feel about your answer to QE.2.2?	[1-10]
QE.3.1 Do you think it is acceptable to deceive others?	[1-10]
QE.3.2 We asked 100 `X' "do you think it is acceptable to deceive others?" on a scale 1-10. How many do you think answered 6 or higher? We will give you 5,000 shillings if you are within 10 people of the correct answer.	[0-100]
QE.3.3 Where 1 = not at all confident and 10 = completely confident, how confident do you feel about your answer to QE.3.2?	[1-10]

Question	Range
QE.4.1 Do you think the use of violence is acceptable?	[1-10]
QE.4.2 We asked 100 `X' "do you think the use of violence is acceptable?" on a scale 1-10. How many do you think answered 6 or higher? We will give you 5,000 shillings if you are within 10 people of the correct answer.	[0-100]
QE.4.3 Where 1 = not at all confident and 10 = completely confident, how confident do you feel about your answer to QE.4.2?	[1-10]
QE.5.1 Do you think cheating on your spouse is acceptable?	[1-10]
QE.5.2 We asked 100 `X' "do you think cheating on your spouse is acceptable?" on a scale 1-10. How many do you think answered 6 or higher? We will give you 5,000 shillings if you are within 10 people of the correct answer.	[0-100]
QE.5.3 Where 1 = not at all confident and 10 = completely confident, how confident do you feel about your answer to QE.5.2?	[1-10]
Where 1 = completely disagree and 10 = completely agree, as a Gisu how much do you agree the following statements reflect stereotypes of `X' (ask for each of the 5 cultures):	
QE.6.1 People from `X' are faithful?	[1-10]
QE.6.2 People from `X' are intelligent?	[1-10]
QE.6.3 People from `X' are stubborn?	[1-10]
QE.6.4 People from `X' are aggressive?	[1-10]
QE.6.5 People from `X' are efficient?	[1-10]
QE.6.6 People from `X' are very religious?	[1-10]
QE.6.7 People from `X' are arrogant?	[1-10]
QE.6.8 People from `X' are deceitful?	[1-10]
QE.6.9 People from `X' are straight forward?	[1-10]
QE.6.10 People from `X' are political?	[1-10]
QE.7 How well do you feel other Gisu understand you and your beliefs?	[1-10]

Confirm, by show of a raise of hands, whether participants would like more time or any or the questions repeating.

Now that you have completed your responses, please ensure that you have written your answers to Q1-5.2 on both Answer Sheets 1.0 and 1.1. Please also ensure that your ID number is written at the top of both pieces of paper. We are about to ask you the second set of incentivised questions. For these, you will need to refer to your answers to QE1-5.2 in order to win any potential bonus payout.

We will now collect your first answer sheet and give you the second response sheet and some additional information.

Collect Answer Sheet 1.0, ensuring responses are complete and consistent with Answer Sheet 1.1 for QE.1-5.2. Check that the participant's ID is written on both answer sheets correctly. Hand out Answer Sheet 1.2 and Handout 1.0 (this must be after Answer Sheet 1.0 has been collected).

We have now provided you with a further answer sheet and a handout which contains some additional

information. Like QE1-5.2, these questions are eligible for the bonus payout, if you are within 10 of the actual answer. For these questions we are providing some additional information – the actual responses of 10% of those surveyed, shown in Handout 1.0.

Go back and re-read the initial questions and ask the participants to re-answer the question now that they have the updated information.

QE.1.4 Go back to your answer to QE1.2. The handout provided shows the actual responses for a randomly selected 10% of all `X' who answered 6 or higher in QE.1.2. Given this additional information, what would your answer to QE.1.2 be now, for each group?	[0-100]
QE.2.4 Go back to your answer to QE2.2. The handout provided shows the actual responses for a randomly selected 10% of all `X' who answered 6 or higher in QE.2.2. Given this additional information, what would your answer to QE.2.2 be now, for each group?	[0-100]
QE.3.4 Go back to your answer to QE3.2. The handout provided shows the actual responses for a randomly selected 10% of all `X' who answered 6 or higher in QE.3.2. Given this additional information, what would your answer to QE.3.2 be now, for each group?	[0-100]
QE.4.4 Go back to your answer to QE4.2. The handout provided shows the actual responses for a randomly selected 10% of all `X' who answered 6 or higher in QE.4.2. Given this additional information, what would your answer to QE.4.2 be now, for each group?	[0-100]
QE.5.4 Go back to your answer to QE5.2. The handout provided shows the actual responses for a randomly selected 10% of all `X' who answered 6 or higher in QE.5.2. Given this additional information, what would your answer to QE.5.2 be now, for each group?	[0-100]

Would anyone like us to repeat any of the questions?

We will now collect your second answer sheet and give you the next response sheet.

Collect Answer sheets 1.1 and 1.2 (ensuring that the student ID is clearly written at the top) as well as handout 1.0. The participant is not allowed to retain any of the sheets. Hand out Answer Sheet 2.1, as shown in Appendix 2.

CHOICE OF BONUS QUESTION

While sheets are being collected and handed out, ask a random participant to select the bonus question from bag while the answer sheets/handout are being collected. Ensure this is observable by the group for transparency.

Phase Two - Experiment

PREPARATION OF THE SURVEY

Material needed:

- *Explanation wall charts*
- *Answer Sheets 2.1, 2.2 and 2.3*
- *Payout computation & data entry sheet (excel)*
- *Props for explanation: basket and envelope*
- *Post-it notes*
- *Bag with 20 pieces of paper with question numbers for random participant to select bonus question [5.1-5, 6-10, 11.1-5, 12-17]*

BRIEF EXPLANATION FOR EXPERIMENTERS ONLY:

We will play a public goods game using either a 'Community Game' (CG) or a 'Wall Street Game' (WSG) (randomly determined) treatment. Each participant will play either the CG or WSG first, and then hypothetically play the WSG or CG second (in that order). We need an even number of players in each session as players are paired. It is important that subjects do not know who their partner is.

Participants will be told that they are playing a Community Game [Wall Street Game]. After the rules of the game have been explained, participants will be able to contribute either UGX 0, UGX 2,000, UGX 4,000 or UGX 8,000 to the common pot. This decision will attract a payout for the participant.

After this initial decision, we will ask the participants what percentage of participants in the session they think contributed to the common pot at each level. Next, we will ask them, on a four-point scale, how socially acceptable it is to contribute at each of the four levels. We will randomly select one of the eight questions here for a bonus UGX3,000 payout if participants are with +/- 10% of the actual answer for expectations and in the right category (one of four) for the norm questions. Need to be clear that this is different bonus and that this part of the experiment is separate to the previous survey. Lastly, we will ask participants to imagine they had never played the Community Game [Wall Street Game]. "Imagine you are seeing this for the first time, but it is instead called the Wall Street Game [Community Game]. What answers would you have given?" The expectations/norms questions here again attract a potential UGX3,000 bonus.

EXPLAINING THE GAME

We are now going to play a simple game about money called the Community Game [Wall Street Game]. Today, you have randomly been paired with someone else in this room. You will not find out the identity of your partner, and they will not find out any information about you. All decisions are anonymous. However, we can tell you that your partner is either a Gisu or spent time in the Bugishu region.

I will explain all of the decisions required for the Community Game [Wall Street Game] slowly, and ask you to write down your answers on the paper in front of you. You cannot change your answers after they've been written down, so think carefully before you write anything. Any questions that you have can be answered privately.

Here use a physical basket and envelope to explain the game.

At the start of the Community Game [Wall Street Game] you will be given 8,000 Ugandan shillings, and you can decide what to do with it. First, we will demonstrate the decision using real money. You will make your choice on paper in front of you. You have two possible options: you can place money in either a private envelope (*show*) or a common basket (*show*). You can choose to put some of the money in the common basket, and the rest in the private envelope, but only in intervals of 2,000 Shillings. You can choose to keep this money for yourself, by placing it in the private envelope (*show*). This is your money to take home with you. There is also a common basket, which both you and you partner can put money in (*show*). We will add half of the money in the common basket (*show*). It will then be shared equally between the two players (*show*). Let's recap the Community Game [Wall Street Game] rules together: [*short, direct answers only*]:

- What happens with any money you decide to put in your private envelope? [*You take it home*]
- How much is added to any money you and your partner put in the common basket? [*Half*]
- And after half is added, how do we split the money in the common basket between you and your partner? [*Equally*]

CONTROL QUESTIONS

We will now check for your understanding of the Community Game [Wall Street Game], using 3 examples. Firstly, please put your ID numbers at the top of your answer sheet.

[Use an illustrative wall chart here to explain the game, using post it notes to show what would be written in each box in order to calculate the overall payout. Ensure actual wall chart being used has an UGX 8,000 endowment.]

Imagine two people are paired: person A and person B. They would not know who they are paired with. *[Demonstrate with real money and using the visual aid. Jointly solve 1 and 2 below, but 3 must be solved by the participant directly without aid. Read question number, pause with each instruction to write down, indicating where to write down.]*

1. Imagine that person A chooses to put nothing in the common basket, and everything in the private envelope. And imagine that person B chooses to put nothing in the common basket, and everything in the private envelope. Write on your paper, in the appropriate boxes, how much is in the common basket. How much is there after we have added half. And how much each player goes home with. *[Jointly solve this, asking for answers from the room]*

2. Imagine that person A chooses to put everything in the common basket, and nothing in the private envelope. And imagine that person B chooses to put everything in the common basket, and nothing in the private envelope. Write on your paper, in the appropriate boxes, how much is in the common basket. How much is there after we have added half. And how much each player goes home with. *[Jointly solve this, asking for answers from the room]*

3. Imagine that person A chooses to put everything in the common basket, and nothing in the private envelope. But, imagine that person B chooses to put nothing in the common basket, and everything in the private envelope. Write on your paper, in the appropriate boxes, how much is in the common basket. How much is there after we have added half. And how much each player goes home with. Please write your answers on your answer sheets. *[This must be completed by the participant without aid].*

Collect answer sheets 2.1 ensuring that the participant's ID number is populated. Hand out answer sheet 2.2.

PGG DECISION – CG [WSG]

Thank you. These are just examples, and for the real Community Game [Wall Street Game] you can decide what you prefer. When you make the decision, you can choose an amount, between 0 and 8,000 shillings, to put in the private envelope. You can choose an amount, between 0 and 8,000 shillings, to put in the common basket. Remember that we will pay you real money at the end of the experiment, depending on what you and your partner decide. Please make your choice now for the Community Game [Wall Street Game], by ticking once for question 8. *[Indicate where on an illustrative wall chart].*

EXPECTATIONS – CG [WSG]

We will now ask you questions 9-14 about behaviour in the Community Game [Wall Street Game]. Once you have made all of these decisions on this page, we will randomly pick one question. If you get the answer correct in the question we pick, we will give you another 3,000 shillings as a bonus. If one of the responses to question 9 is selected you will need to be within +/-10% of the actual answer to get the payout.

Let us remind you that it is very important that you do not talk during the experiment, and that you only mark one box per question. If you mark more than one box you will not be able to receive the bonus.

For question 9, there are five boxes, each showing difference scenarios.

[Use an illustrative wall chart to explain the question]:

- The left-hand box shows that the entire 8,000 shillings has been placed in the private envelope.
- The right-hand box shows that the entire 8,000 shillings has been placed in the common basket.
- In each of the boxes how many people in this room do you think contributed the amounts shown?

NORMS– CG [WSG]

Now we will give a series of situations where someone made a decision. I will ask you to consider the different possible choices available and to decide, for each of the possible actions, whether taking that action would be "socially acceptable" and "consistent with moral or proper social behaviour" or "socially unacceptable" and "inconsistent with moral or proper social behaviour."

By socially acceptable, we mean behaviour that most people agree is the "correct" or "ethical" thing to do. Another way to think about what we mean is that if someone were to select a socially unacceptable choice, then someone else might be angry at them for doing so.

Remember, if this set of questions is chosen, you could earn another 3,000 shillings on top of what you earned in the first section of the experiment. You would earn that money if you give the same answer as the most popular choice. For these questions, we are not interested in your preferences. Rather, we are interested in what you think the most popular choice would be.

We will now go through an example. *[use an illustrative wall chart with post it notes to illustrate during explanation].*

Imagine someone is at a local coffee shop. While there, they notice that someone has left a wallet at one of the tables. Someone sees and must decide what to do. They have four possible choices, and you need to rate how socially acceptable, "correct" or "ethical" that action is. *[Read each choice out, ask 'how would you rate that action?' give the 4 possible ratings, and get an assistant experimenter to answer using the below scale. Use visual aid throughout].*

	Very socially unacceptable	Socially unacceptable	Socially acceptable	Very socially acceptable
	--	-	+	++
Take the wallet	X			
Ask others nearby if the wallet belongs to them			X	
Leave the wallet where it is		X		
Give the wallet to shop manager				X

[Assistant Experimenter:] I think most people in this room would say 'action' is 'rating'. So I would tick here *[mark on the visual aid].*

Now that we've gone through an example, we will turn to our questions. Remember that if you give the same answer as the most popular option, and if that question is randomly chosen, you could earn extra money.

We are now going to answer questions 10-14. Remember this is how socially acceptable or unacceptable people in this room playing the Community Game [Wall Street Game] believe, on average, each contribution level to the common pot is. Imagine someone put nothing in the private envelope and everything in the common basket *[use an illustrative wall chart to support explanation].*

Please rate this as either very socially unacceptable (–), somewhat socially unacceptable (-), somewhat socially acceptable (+) or very socially acceptable (++) by ticking once in that row. You will see another four possible choices, where someone put either 2,000, 4,000, 6,000 or 8,000 in the private envelope. Please rate each choice as either very socially unacceptable (–), somewhat socially unacceptable (-), somewhat socially acceptable (+) or very socially acceptable (++). Remember, you can only get a bonus if you tick once per row.

REPEAT PGG, EXPECTATIONS & NORMS – WSG [CG]

Now imagine that you have just walked into this room. Instead of being told that you are going to be playing the Community Game [Wall Street Game] we told you that you will be playing the Wall Street Game [Community Game]. All of the conditions and rules are exactly as before. Take a moment to visualise this. Please mark what you would have contributed to the common pot in the Wall Street [Community] Game in question 15.

Now for question 16, in each of the boxes, how many people in this room do you think would have contributed

the amounts shown when hypothetically playing the Wall Street Game [Community Game]? As before there is a bonus 3,000 shillings if question 12 is selected at the end of session and you are within +/-10% of the actual answer.

Lastly, for questions 17 - 21, please mark how socially acceptable or unacceptable people in this room playing the Wall Street Game [Community Game] believe, on average, each contribution level to the common pot is. The 3,000 shilling bonus is also available for these questions. Remember, you can only get a bonus if you tick once per row.

CHOICE OF BONUS QUESTIONS

Collect participants Answer Sheets 2.2 ensuring their ID numbers are correctly written at the top of the page. Hand out Answer Sheet 2.3. As with stage 1, get a random participant to select one bonus questions blind from the bag.

Phase 3 - Survey

Please enter your ID number at the top of the new answer sheet. While we calculate your earnings, we'd like to ask a few general questions, to understand more about you. All information is anonymous, will not affect your earnings and is given voluntarily. If you wish not to answer a question, you are allowed to skip it.

- How old are you? [In years]
- What is your gender? [Male/Female]
- What is your highest educational qualification?
- How many people in this room do you know by name? [Please don't include the experimenter(s)]
- What is your religion?
- Do you have membership to any political parties? [Yes/No]
- If so, which parties?

FINAL INSTRUCTIONS

Thank you, you have now all completed all of the tasks. We now invite you to come forward, one by one, to collect up your earnings. Thank you for coming today, your participation has been greatly appreciated.

Stage 2: Experimental Answer Sheets & Handouts

For answer sheets, please use a random variant where the order of cultures has been randomised. Please record to variant used, as supplied separately to this script.

ANSWER SHEET 1.0

ID No: _____

Question	*Range*	Participant	Gisu	Baganda	Acholi
QE.1.1	*1-10*				
QE.1.2	*0-100*				
QE.1.3	*1-10*				
QE.2.1	*1-10*				
QE.2.2	*0-100*				
QE.2.3	*1-10*				
QE.3.1	*1-10*				
QE.3.2	*0-100*				
QE.3.3	*1-10*				
QE.4.1	*1-10*				
QE.4.2	*0-100*				
QE.4.3	*1-10*				
QE.5.1	*1-10*				
QE.5.2	*0-100*				
QE.5.3	*1-10*				
QE.6.1	*1-10*				
QE.6.2	*1-10*				
QE.6.3	*1-10*				
QE.6.4	*1-10*				
QE.6.5	*1-10*				
QE.6.6	*1-10*				
QE.6.7	*1-10*				
QE.6.8	*1-10*				
QE.6.9	*1-10*				
QE.6.10	*1-10*				
QE.7	*1-10*				

ANSWER SHEET 1.1

Question	*Range*	Gisu	Baganda	Acholi
QE.1.2	*1-100*			
QE.2.2	*1-100*			
QE.3.2	*1-100*			
QE.4.2	*1-100*			
QE.5.2	*1-100*			

ANSWER SHEET 1.2

Question	*Range*	Gisu	Baganda	Acholi
QE.1.4	*1-100*			
QE.2.4	*1-100*			
QE.3.4	*1-100*			
QE.4.4	*1-100*			
QE.5.4	*1-100*			

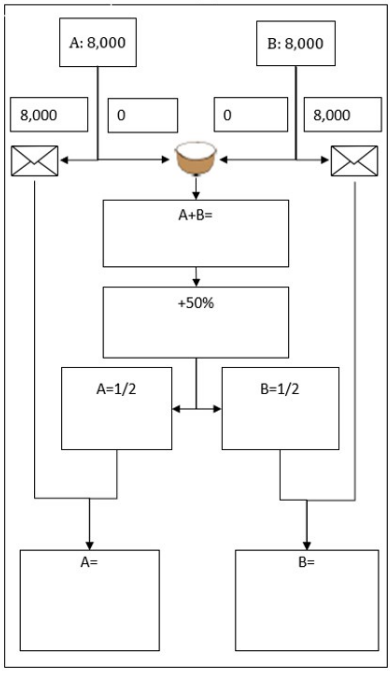
HANDOUT 1.0

Question	Gisu	Baganda	Acholi
QE.1.2	9	7	8
QE.2.2	8	9	5
QE.3.2	2	2	2
QE.4.2	1	1	5
QE.5.2	2	3	3

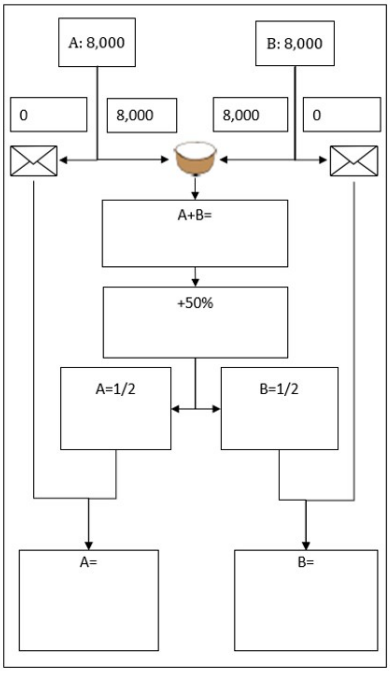
ANSWER SHEET 2.1

ID_____

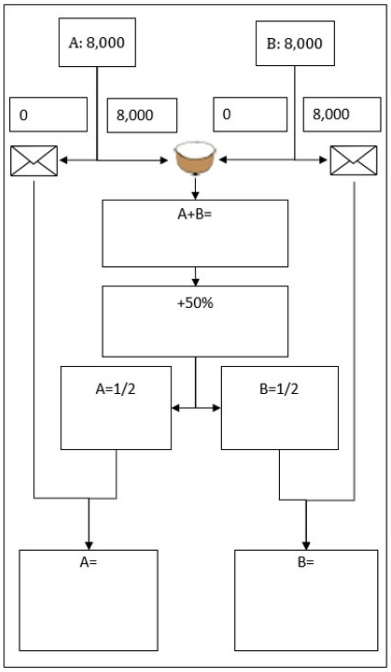
1.



2.



3.



ANSWER SHEET 2.2

ID_____

4.

8,000	6,000	4,000	2,000	0	
☐	☐	☐	☐	☐	
0	2,000	4,000	6,000	8,000	

5.

 8,000  0	 6,000  2,000	 4,000  4,000	 2,000  6,000	 0  8,000
%	%	%	%	%

	--	-	+	++
6.  : 8,000  : 0				
7.  : 6,000  : 2,000				
8.  : 4,000  : 4,000				
9.  : 2,000  : 8,000				
10.  : 0  : 8,000				

11.

8,000	6,000	4,000	2,000	0	
☐	☐	☐	☐	☐	
0	2,000	4,000	6,000	8,000	

12.

 8,000  0	 6,000  2,000	 4,000  4,000	 2,000  6,000	 0  8,000
%	%	%	%	%

	--	-	+	++
13.  : 8,000  : 0				
14.  : 6,000  : 2,000				
15.  : 4,000  : 4,000				
16. : 2,000 : 8,000				
17. : 0 : 8,000				

ANSWER SHEET 2.3

ID:_____

1. _____

2. _____

3. _____

4. _____

5. _____

6. _____

7. _____

Appendix 4 - Ethical Consent Forms

LIRA UNIVERSITY

P.O. Box 1035
Lira, Uganda
Tel: +256 414-694716
Web: www.lirauni.ac.ug
Email: lirarec@lirauni.ac.ug.

RESEARCH ETHICS COMMITTEE

SHIWANDIHO SHEHUFUKIRISA KHUKHWITUBASA MUMUSOMO

Lugisu

Sheyusemu sha ndwela kharonyese noti (1.0)

21st May 2024

Ukhwambasilisa kyimisiro nitsisambo tsikwanile mubandu-shisintsa shehabiri

Iwee Maayii/Papaa,

Intsanjula

Khwakhulangile khukhwitubasa mumusomo kwehuyentselesa. Lisina lyekumusomo ku bakulangille bali "Ukh wambasilisa kyimisiro nitsisambo tsikwanile mubandu". Bayentselesi khukhwama mu Univasite ya East Anglia bali khukana khumanya nabi lwashi, nanjeri shina bikanda byebandu benjawulo mu Uganda bakhalawo khubindu bihali bitwela taa. Barere kamapesi mumusomo kuu nisho shitongole shesi baranga bari South East Network for Social Sciences, Ibulayaa.

Lwashina kumusomo kuu kuli khukholewa?

Kumusomo kuu kuli khukholewa niyo shinyalise bayentselesi nibahola kamakambila kahambagana nintsowatsowana khutegela bulayi lwashina babandu bahalawo munjeri yesi bahalilewo, nitsisonga tsinyala tsanyonyola oba tsaloba khunyonyola ingalawo yabwe.

Babandu benga babitubasile mumusomo kuu?

Ambi babandu lunana (800) batsa khwitubasa mumusomo kuu, mubisintsa bibiri. Mushisintsa shehabiri munabamo

babandu ambi bisatu (300). Khwahulangile khukwitubasa mumusomo kuu lwekhuba olimutwela khubandu bo bekyimisiro nibulombolombo bwesi bayentselesi bali khukana khumanya.

Kyimitendela kyekumusomo

Kumusomo kulimo birebo nibibindu byekhuhalawo byesi banawandikha khutsimbapula tsesi babawele. Mbawo intsilamo ifuu oba intufu taa. Shama khushesi bibyambasobyo bikana. Unitubasa mumusomo khumala ambi isawaa ndwela.

Buwangafu bunyala bwangwaho

Mbawo buwangafu bwekamani bunyala bwakhukwaho lwekhwitubasa mumusomo kuu taa, kasita ubanga wiwulila bulayi ukhwitubasamo. Birebo bibyangu, ela unyala waloba khukhwilamo shirebo shosi kasita ubona uri shahuhusanyusa taa.

Khufunamo

Ahali khasimo hesi khunakhuwa, mbawo injeri yosi kyindi yesi unafunamo mumusomo ku iwee inga umundu taa. Nenga, aliwo tsinjeri tsindi tsesi kumusomo kunyala kwaba kwekumugaso, khekhulome nga khuyetaho khumanyisa babandu bahola kamakambile kahambagana nintsowatsowana.

Garama ili mumusomo

Mbawo garama yosi, oba shishindu shosi shesi unasasula niyo witubase mumusomo kuu taa.

Ukhuntisilawo/Ukhusasula

Mbawo garama yosi, oba shishindu shosi shesi unasasula niyo witubase mumusomo kuu taa. Nenga, unawebwa silingi akari waa kasanvu bitano, ni kyimitwalo kyibili kakaga (7,500-26,000)., nga khukhukoboselawo inyenda, nimbuka yowo. Silingi tsesi unafuna tsinasinzila khunjeri yesi wililemu birebo.

Lidembe nindabilila yebabandu babitubasa mumusomo

Khukhwitubasa mumusomo kuu shama mukhukana khwo. Uri niledembe khuloba khwitubasa mumusomo kuu, oba khukururamo imbuka yosi. Bibindu byosi byesi wahubolela nga ushili khurura mumusomo binyala biyetesebwa, oba binyala byaba nga babirambisa hale khurusa bitabo.

Nga wabeleho nibirebo byosi biwambagana nikumusomo,oba khukhwimulugunya khwosi khidembelyo nga mutwela khubandu babitubasile mumusomo kuu,unyala watabulila **Dr. Odur Andrew umukosi umuhulu uwe khashiho khahambanisa tsisambo tsebabandu bahola byehuyentselesa kha Univasite iye Lira khu namba iyee siimu ino**

0772 714386 oba khu andyodur55@gmail.com.

Kumusomo kuu kwafukirisiwa ni khashiho khahambanisa tsisambo tsebabandu bahola byehuyentselesa kha Univasite iye Lira; mubwimbi hesi hulanga huri (LUREC).

Nga abee nalobele khukhwitubasa mumusomo

Nga walobele khukhwitubasa mumusomo, khunafukirisa shesi unaba uhalilewo, era mbawo injeri yosi yesi unakhosebwamo taa.

Khubiha byesi ukhubolele nga byeshishama

Byosi byesi unakhubolela binabihibbwa nga byeshishama. Ukhwitaga khukhubolela shishindu shosi shesi umundu anyala iyetesa khukhumanyilaho taa, khulome nga lisina lyoo, oba indakiriro yosi. Banakhuwa inamba yesi unarambisa mushifo sherisina lyoo,nashiryo mbawo umundu yesi unakhumanya taa. Byosi byesi unakhubolela mumusomo kuu binabihibwa nga byeshishama. Mbawo shishindu shosi shinyala shayetaa umundu kundi khukhumanya shesi khunarusa mubindu byesi khunawandiha.

Birebo biyambagana nikumusomo

Urangilemo kumusomo kuu/ umuyentseresi anakhunyonyola biyambagana nikumusomo. Nga uriho niberebo unyala wabireba imbuka yino, oba imbuka yosi nga kumusomo kusa iburangise. Oba unyala wamutabulila, khundakililo ili khamalo:

Joshua Hill

Postgraduate Researcher

University of East Anglia,

Research Park,

Norwich NR4 7TJ

Mobile : +44 7757 057 225

Khukhwatula khwomundu uwitubasile mumusomo

Nga nihisa kumuhono khushiwandiho shehufukilisa khwitubasa mumusomo, imba nafukilisile khukhwitubasa mumusomo kuu. Kumusomo kwesi bali khukanikhaho bakulanga bari Ukhwambasilisa kyimisiro nitsisambo tsikwanile mubandu-shisintsa shehabiri uwitubasile mumusomo kuu ifukilisana nibili khamwalo:

- Byesi bawandihile mushiwandiho shehufukilisa khukhwitubasa mumusoma bansinyonyole byosi bulayi, ela nabitegele bulayi.
- Birebo byose byosi babiriremo bulayi ela namatile.
- Nafukirisile khukhwitubasa mumusomo kwesi balanga Ukhwambasilisa kyimisiro nitsisambo tsikwanile mubandushisintsa shehabiri musimbela tsesi bambolele mushiwandiho shehufukilisa khukhwitubasa mumusomo.

Nafunile ikopi iyeshiwandikho shinyonyola byekhufukilisa khukhwitubasa mumusomo.

- Nashimanyile indi khukhwitubasa mumusomo kuu shama mukhukana khwase, ela indi nyala naloba oba narula mumusomo imbuuka yosi.
- Khuloba khukhwitubasa mumusomo nga nanyowele khufukilisa ikhunakosaho ingeri yesi mbilisiwamo taa.

Kamasina khwomwitubasi mumusomo.....

Ukhwisaho kumuhono khwomwitubasi

Khupa shingumu muhabokisi hali yasii

Tsinakhu tsekumwesi:

Tsinakhu Kumwesi Kumwakha

*Nga abe umwitubasi shanyala khusoma ni/oba khuwandikha, omundu ukhihupila tsimbeka ali nikhubawo nga umujulizi nga bamunyonyola biwambagana nikhufikulisa khukhwitubasa mumusomo. Nga bamalile khusoma nikhunyonyola shiwandihko, ela nga bamalile khufukilisa khukhwitubasa mumusomo, nga besaho kumuhomo oba shingumu, umujulixi ahile naye khukhwisaho kumuhono. Nga isileho kumuhono, aba akhakasa ari biri mushiwandikho shekhwituasa mumusomo babinyonyole bulayi umwitubasi, ela afukirisire khukhwitubasa mumusomo khulwewe.