DISCUSSION PAPER SERIES

DP15776
(v. 2)

A Stepping Stone Approach to Understanding Harmful Norms

Selim Gulesci, Sam Jindani, Eliana La Ferrara, David Smerdon, Munshi Sulaiman and H. Peyton Young

DEVELOPMENT ECONOMICS

ECONOMIC HISTORY

PUBLIC ECONOMICS

ISSN 0265-8003

A Stepping Stone Approach to Understanding Harmful Norms

Selim Gulesci, Sam Jindani, Eliana La Ferrara, David Smerdon, Munshi Sulaiman and H. Peyton Young

Discussion Paper DP15776
First Published 07 February 2021
This Revision 10 February 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

A Stepping Stone Approach to Understanding Harmful Norms

Abstract

Harmful social norms often persist despite legal and economic sanctions against them. Can the abandonment of a harmful norm be facilitated by the presence of a 'mildly harmful' alternative that may act as a stepping stone in the medium run? Or will this become a new absorbing norm? We propose a dynamic, game-theoretic model to analyze this question, focusing on interim dynamics. We derive necessary and sufficient conditions for a stepping stone transition, which involve as key parameters the social penalty factors and the difference in intrinsic utilities of the various actions. We explore the implications of the model using original data on female genital cutting in Somalia, where a transition is underway from an extremely invasive to a milder form of cutting. The framework is general and can be applied to other examples, both contemporary and historical, including footbinding, dueling, child marriage and smoking. Our analysis underlines the importance of considering intermediate alternatives when formulating policies against harmful norms.

JEL Classification: C73, O12, Z10

Keywords: Social norms, norm dynamics, female genital cutting, female genital mutilation, FGC, FGM

Selim Gulesci - gulescis@tcd.ie
Trinity College Dublin and CEPR

Sam Jindani - sj608@cam.ac.uk
Cambridge University

Eliana La Ferrara - eliana.laferrara@unibocconi.it
IGIER, Università Bocconi and CEPR

David Smerdon - d.smerdon@uq.edu.au
University of Queensland

Munshi Sulaiman - munshi.sulaiman@brac.net
BRAC

H. Peyton Young - peyton.young@economics.ox.ac.uk
Oxford University

Acknowledgements

We would like to thank seminar participants at Bocconi, CEU, Essex, Oxford, IFO Workshop on Gender-Based Violence, 2020 BREAD Conference on Behavioural Economics & Development and 2020 ThReD Conference for useful comments. Julien Manili provided outstanding research assistance. We also thank Beatrice Bonini, Mattia Chiapello, Matteo Courthoud, Enrico Guizzardi, Vrinda Kapoor, Jack Melbourne, Giovanni Pisauro and Armin Satzger for excellent work on the data. La Ferrara acknowledges financial support from MIUR FARE grant J42F17000280001.

A Stepping Stone Approach to Understanding Harmful Norms*

Selim Gulesci[1] Sam Jindani[2] Eliana La Ferrara[3]
David Smerdon[4] Munshi Sulaiman[5] H. Peyton Young[6]

10 February 2021

Abstract

Harmful social norms often persist despite legal and economic sanctions against them. Can the abandonment of a harmful norm be facilitated by the presence of a 'mildly harmful' alternative that may act as a stepping stone in the medium run? Or will this become a new absorbing norm? We propose a dynamic, game-theoretic model to analyze this question, focusing on interim dynamics. We derive necessary and sufficient conditions for a stepping stone transition, which involve as key parameters the social penalty factors and the difference in intrinsic utilities of the various actions. We explore the implications of the model using original data on female genital cutting in Somalia, where a transition is underway from an extremely invasive to a milder form of cutting. The framework is general and can be applied to other examples, both contemporary and historical, including footbinding, dueling, child marriage and smoking. Our analysis underlines the importance of considering intermediate alternatives when formulating policies against harmful norms.

* We would like to thank seminar participants at Bocconi, CEU, Essex, Oxford, IFO Workshop on Gender-Based Violence, 2020 BREAD Conference on Behavioural Economics & Development and 2020 ThReD Conference for useful comments. Julien Manili provided outstanding research assistance. We also thank Beatrice Bonini, Mattia Chiapello, Matteo Courthoud, Enrico Guizzardi, Vrinda Kapoor, Jack Melbourne, Giovanni Pisauro and Armin Satzger for excellent work on the data. La Ferrara acknowledges financial support from MIUR FARE grant J42F17000280001.

1 Trinity College Dublin; gulescis@tcd.ie.
2 University of Cambridge; sj608@cam.ac.uk.
3 Bocconi University and LEAP; eliana.laferrara@unibocconi.it.
4 University of Queensland; d.smerdon@uq.edu.au.
5 BRAC; munshi.sulaiman@brac.net.
6 University of Oxford; peyton.young@economics.ox.ac.uk.

1 Introduction

Harmful norms often persist despite the high costs they entail for individuals and society, and despite the presence of legislation against them. The conventional approach taken by governments and NGOs is to push for the outright abandonment of these norms, possibly on a matter of principle. In practice, such an approach is often ineffective. Currently, dowry and early child marriage persist in South Asia despite being outlawed (Anderson, 2007; Ambrus & Field, 2008; Corno *et al.*, 2017); female genital cutting is widespread in Africa despite the fact that many governments have passed laws against it.[7] Historically, footbinding in China and dueling in Europe persisted for centuries despite repeated attempts by governments to extirpate the practices (Mackie, 1996; Nye, 1993).

If the goal is the elimination of a harmful norm in the long run, can it be beneficial to introduce a 'mildly harmful' alternative in the short run?[8] Leaving aside moral principles (on which our analysis is silent), the answer to this question is not trivial. On the one hand, people who may be reluctant to abandon a costly practice or trait T in favor of the complete absence of the practice may be persuaded instead to go from T to a slightly less costly trait t. On the other hand, precisely because t is less costly, it may become an absorbing state in the long run, in the sense that the forces that may have led to the ultimate abandonment of T over time given its costs may no longer work if the costs associated with t are lower.

In this paper we propose a model of norm dynamics that allows us to analyze the above trade-off and characterize the equilibria and the conditions that govern norm transitions. In particular, we are interested in understanding the conditions under which the existence of an intermediate alternative leads to the elimination of a costly norm – a situation we describe as a 'stepping stone' transition.

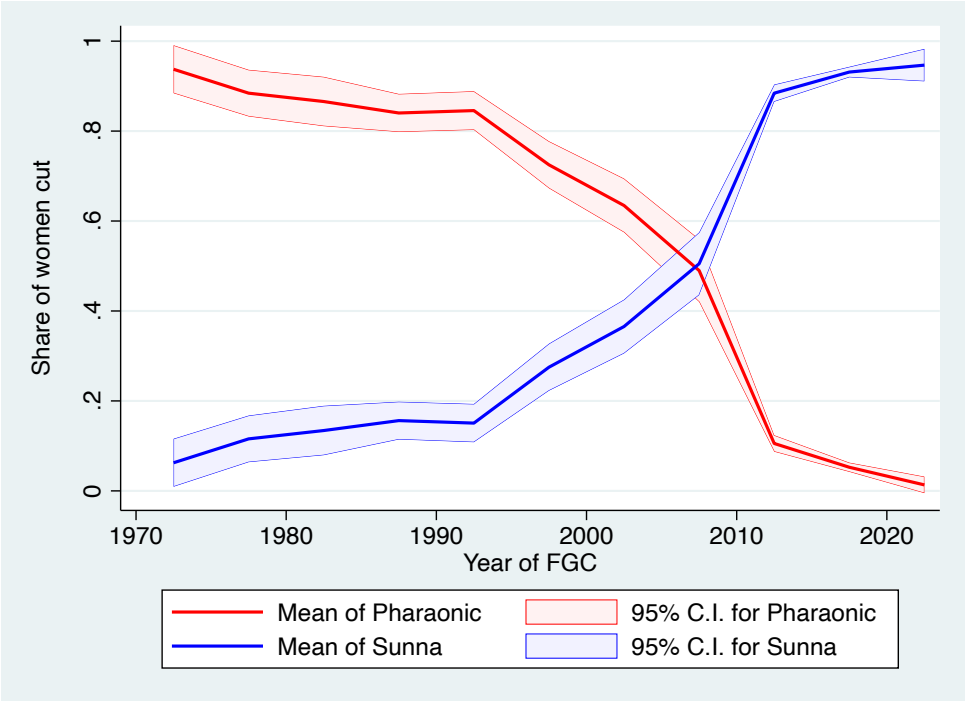
We then provide an empirical illustration of some key features of our model using original data on female genital cutting (FGC) from Somalia, a country where 98 percent of the women aged 15 to 49 are cut (UNICEF, 2016). Somalia is an interesting case study because in this country FGC can take two forms: an extremely invasive procedure called 'Pharaonic' circumcision and a less invasive one called 'Sunna'.

Figure 1 uses original data we collected in Somalia to show the type of FGC performed as a function of the year in which the girl or woman was cut. Until about twenty years ago, the vast majority of women who were cut received Pharaonic circumcision. Then a

7 A recent study reports that in 2018, among the 28 African countries with the highest prevalence of FGC, 22 had national legislation criminalizing it (28 Too Many, 2018).

8 For the examples mentioned above, the 'mildly harmful' alternatives might consist in setting a cap on dowry payments, setting a different minimum age for marriage, or allowing for less invasive forms of female genital cutting.

Figure 1: Type of FGC by year of cutting



Source: Authors' calculations on original data from Somalia. Sample includes women aged 17-82 at the time of interview and their daughters aged 12 or older, if cut.

striking transition occurred, reversing the trend, and in recent years the vast majority of girls who were cut received Sunna. Our model can be used to understand this transition from the more to the less harmful type of cutting and to study the conditions under which the system could remain stuck with Sunna, as opposed to transitioning to a no-cutting equilibrium in the future.

We propose a discrete choice model in which a finite population of players choose an action to maximize a utility that has two components: (i) the intrinsic utility of the action for the individual, and (ii) a 'social' utility that is decreasing in the shares of players who choose actions different from one's own. The latter term captures conformity and is commonly used in models with social interactions (e.g., Akerlof, 1997; Brock & Durlauf, 2001). We assume agents disapprove more of actions further from their own, as measured by the difference in intrinsic utilities. We start by solving the model in its simplest version – without heterogeneity and with three actions: L, M and H (for low, medium and high intrinsic utility). We then add heterogeneity and generalize to n actions.

We define an intermediate action M as a 'stepping stone' if it destabilizes the costly norm L but is not itself stable. In other words, starting from a situation where the costly norm L is stable in the sense that it would not be abandoned in favor of H, M is a stepping stone if it renders L unstable but at the same time it is not itself stable relative to H, so that the entire population will eventually converge to the 'good' norm H.

We derive necessary and sufficient conditions for the intermediate action to be a stepping stone, and we show that these ultimately require that, in the face of a relatively high social penalty for transitioning from L to H, the social penalties for going from L to M and then from M to H are small relative to the gains in intrinsic utility that they generate. In other words, M must be a good 'social substitute' for both L and H.

We also show that a necessary condition for M to be a stepping stone in the model without heterogeneity is the reverse triangle inequality: that the sum of the social penalties for the two 'intermediate' transitions (L to M and M to H) should not exceed the social penalty for the direct transition L to H. Failure to satisfy this condition implies that either the intermediate norm is not attractive enough and will fail to destabilize the costly norm L, or it is too attractive in which case it will become absorbing.

We then characterize the dynamics of stepping stone convergence and derive the expected waiting time, discussing welfare implications. When the costly norm L is stable, introducing an intermediate norm weakly increases welfare, since it may dislodge the costly norm. However, when the costly norm is unstable, introducing an intermediate norm typically decreases welfare. If it is absorbing, then the process may get stuck at the new intermediate norm. If it is not absorbing, the transition to the high-utility norm will tend to be slower.

We then introduce heterogeneity between agents (which may be due to innate differences in preferences or to preference shocks). The main intuitions from the model without heterogeneity carry through, but we show that M can play the role of a stepping stone under a larger set of parameter values than previously. The reason is that heterogeneity entails a persistent non-zero proportion of M- and H-players, which reduces the social cost of deviating, first to the intermediate action M, and then to the good action H.

Finally, we extend the model to the case of n actions, under the assumption that social sanctions are a function of the distance between actions (in terms of intrinsic utility). We show that the shape of the sanctions function plays a key role in the dynamics of the model. In particular, we show that the existence of a stepping stone is not compatible with a monotone decreasing average sanction function $f(u)$, defined as the ratio of the sanction $s(u)$ to the intrinsic utility u. The intuition is that when $f(u)$ is decreasing, the best deviation from a norm is always the action with highest intrinsic utility, hence there is no scope for intermediate actions to serve as stepping stones.

In the last part of the paper we use original data that we collected in Somalia to validate the assumptions of the model and to test its predictions. Our data cover 4,130 households from 141 communities across six districts and include information on whether female respondents (and their daughters) are cut, the type of cut, their beliefs about the expected costs and benefits from FGC, and about community sanctions against those who deviate from the practice. To map these data to our model, we interpret Pharaonic circumcision as action L, Sunna as M, and Uncut as H.

We start by showing that people are aware of the differential health costs of different types of FGC, with a higher fraction of respondents having experienced health complications from Pharaonic circumcision compared to Sunna, and no complications from Uncut. The ranking of these costs thus aligns with our assumption regarding the intrinsic utilities of the different actions. Second, we show that respondents' beliefs about social sanctions also align with the model's assumptions. In particular, respondents are more likely to believe that parents who chose Pharaonic circumcision for their daughters would be upset if their son married an uncut girl, compared to the son marrying a girl with Sunna. Similarly, respondents believe that the scenario in which a son marries an uncut girl would be more upsetting for parents who chose Pharaonic for their daughter than for those who chose Sunna. This is in line with our model's assumption that agents disapprove more of actions that are further away from their own.

We then proceed to explore equilibrium dynamics and model predictions. Figure 1 discussed above shows that, while until the late 1980s virtually all girls who were cut received Pharaonic circumcision, in the early 1990s a sharp switch towards Sunna occurred. This shift roughly corresponds to a period where, following numerous human-rights campaigns against FGC, religious leaders started disavowing Pharaonic circumcision and supporting Sunna as consistent with Islam (Newell-Jones, 2016). In terms of the model, this represents the scenario in which action M successfully invades (and replaces) the existing norm L, where H had been unable to do so.

The question, then, is whether a stepping-stone transition will occur or whether Sunna will instead become the new long-run equilibrium. First, we note that although the proportion of uncut women and girls has been increasing over time, this does not necessarily imply that a stepping-stone transition will occur. Indeed, the model predicts that the proportion of uncut women and girls would increase even if Sunna were absorbing. However, we show that there is some evidence that a shift may be occurring in a subset of communities. To do so, we exploit variation across communities in the key parameters of our model – in particular, in perceived sanctions against people who choose Uncut over Sunna. We show that in communities with relatively low sanctions, the threshold property predicted by our model is consistent with the data, leaving room for a transition to Uncut. However, this property fails to hold for the subsample of communities with high sanctions. This pattern is consistent with our model, and suggests that in some communities Sunna may be a stepping stone, while in others it may be absorbing.

We conclude the paper by briefly discussing other potential applications of our model, such as the historical norm of dueling, as well as the norms of child marriage and smoking.

Our paper relates to multiple literatures. First, we contribute to the theoretical

literature on the evolution of social norms.[9] Our paper is particularly close to the approaches of Akerlof (1980, 1997), and Brock & Durlauf (2001), who explicitly model conformity motives and their impact on the existence of norms equilibria.[10] Our paper also relates to work in evolutionary game theory that analyzes how systems of interacting agents converge to normative behaviors from out-of-equilibrium conditions (Young, 1993, 1998; Kandori *et al.*, 1993; Blume, 1993, 1995; Bowles, 2006).[11] This framework has been used to study, for example, the evolution of property rights (Bowles & Choi, 2013, 2019), as well as Pareto inferior institutions and cultures (Belloc & Bowles, 2013).[12]

Whereas most of the literature on the evolution of social norms focuses on long-run dynamics, our main theoretical contribution is to study intermediate-run dynamics and the time it takes to transit between variants of a norm. This is quite relevant for applications where decision revisions are made infrequently (e.g., in the case of FGC, a new choice can be made only when another daughter reaches a certain age). In such cases, the analysis of the welfare implications along the transition paths becomes important: policymakers want not only to be able to predict what the equilibrium will be, but also to understand how long it will take and/or what route society will take to get there.

More broadly, we speak to a growing literature in economics that has studied the persistence and welfare effects of gender norms (e.g., Alesina *et al.*, 2013; Ashraf *et al.*, 2020), including norms around fertility and female labor force participation (Fernandez *et al.*, 2004; Fernandez & Fogli, 2009), and signalling norms such as veiling among Muslim women (Carvalho, 2013). Only recently has female genital cutting come to the forefront of the analysis. Among others, Bellemare *et al.* (2015) study the individual and household level correlates of attitudes towards FGC in West Africa, while Becker (2018) and Corno *et al.* (2020) trace the origins of FGC to historical characteristics of societies. Mackie (1996) was the first to interpret FGC as a social interactions equilibrium. Subsequent work seeks to extend and test this theory (Shell-Duncan *et al.*, 2011; Efferson *et al.*, 2015; Bicchieri & Marini, 2016; Kudo, 2019; Novak, 2020; Efferson *et al.*, 2020). However, this literature does not formalize the dynamics of the adjustment process and it is silent on the role that 'intermediate' actions may play.

Finally, our empirical results contribute to an emerging literature on the evaluation of policies aimed at eradicating FGC (e.g., Diop *et al.*, 2004; UNICEF, 2008; Camilotti,

9 Schelling (1978) was one of the first to introduce a game-theoretic explanation of norms. For overviews of the subject, see Young (1998, 2015); Bowles (2004); and Bicchieri (2005).

10 Related work on social interactions includes the literatures on networks (Jackson, 2008; Goyal, 2012) and identity (Akerlof & Kranton, 2000, 2010). There is also a literature on the empirical estimation of social interactions (Manski, 1993; Moffitt, 2001; Blume *et al.*, 2011).

11 For book-length treatments of evolutionary game theory and its applications see Weibull (1995); Samuelson (1997); Young (1998); Vega-Redondo (1996); Bowles (2004); and Sandholm (2010).

12 Ellison (2000) showed that intermediate actions can speed up long-run transitions in an evolutionary framework with vanishingly small noise. In contrast, we focus on how intermediate actions affect the intermediate-run dynamics with non-vanishing noise.

2016; Vogt *et al.*, 2016; Hombrados & Salgado, 2020). These studies consider a binary choice, whereas here we show that intermediate actions are potentially important for policy.

The remainder of the paper is organized as follows. Section 2 presents our theoretical model. Section 3 contains our empirical application and section 4 concludes.

2 Theory

We present our theoretical framework starting from the simplest version that conveys the main intuition for the dynamics, before generalizing it. In section 2.1 we set out the model with three actions. In section 2.2 we consider the case without heterogeneity and in section 2.3 we add heterogeneity. In section 2.4 we generalize the model to more than three actions.

2.1 Model setup

Consider a population of players who can choose from actions L, M, and H (for low, medium, and high intrinsic utility, respectively). In the case of FGC, L represents Pharaonic cutting, M represents Sunna, and H represents not cutting. Let $A = \{L, M, H\}$. Throughout we shall assume that the number of players m is large but finite. Let $\delta = 1/m$. Let $p_i \in [0, 1]$ denote the proportion of agents playing action i and $p = (p_L, p_M, p_H)$ denote the *state* of the process. Let $\Delta = \{p \in \mathbb{R}^3_+ : \sum_{i \in A} p_i = 1\}$ denote the three-dimensional simplex and $\tilde{\Delta}$ denote the set of feasible states; that is,

$$\tilde{\Delta} = \{p \in \Delta : \forall i \in A, p_i \in \{0, \delta, 2\delta, \ldots, 1\}\}. \tag{1}$$

The utility of an agent playing action i consists of an intrinsic and a social component. The *intrinsic utility* from playing action $i \in A$ is denoted by u_i. We assume actions are ranked in the following order:

$$u_L < u_M < u_H. \tag{2}$$

The social component of an agent's utility represents the pressure to conform to others' actions, as in Akerlof (1997) and Brock & Durlauf (2001). The *social utility* of an agent playing action i is:

$$-\sum_{j \in A} s_{ji} p_j,$$

where s_{ji} represents how much pressure an agent choosing j exerts on someone choosing i. As is standard in the literature, we assume $s_{ii} = 0$ for all i, $s_{ij} > 0$ for all $i \neq j$, and $s_{ij} = s_{ji}$ for all i, j. There are therefore three distinct parameters: s_{LM}, s_{LH}, and s_{MH}. We assume

$$s_{LH} > s_{LM} \text{ and } s_{LH} > s_{MH}. \tag{3}$$

This means that agents disapprove more of actions that are more different than theirs (as measured in terms of their intrinsic utilities). For example, in the case of FGC, agents choosing Pharaonic disapprove more of agents not cutting than of agents choosing Sunna.[13]

In summary, the utility of an agent choosing action i in state p is

$$v_i(p) = u_i - \sum_{j \in A} s_{ji} p_j. \tag{4}$$

In section 2.3 we relax the assumption that all agents have the same utility functions by introducing heterogeneity. We show that the key dynamics are unchanged.

The population size m, the action set A, and the payoff function $v_i : \tilde{\Delta} \to \mathbb{R}$ define a stage game, which we will call $\mathscr{G}$.

For $i \neq j$, define the unit switching vector $e^{ij} \in \mathbb{R}^3$ as follows: $e^{ij}_i = -\delta$, $e^{ij}_j = \delta$, and $e^{ij}_k = 0$ for $k \neq i, j$. Thus if the current state is p and an agent switches from i to j, the new state is $p + e^{ij}$. It will be convenient to let $e^{ii} = (0, 0, 0)$ for all i. Let $B_i(p)$ denote the set of best responses in state p for an agent playing i; that is,[14]

$$B_i(p) = \arg\max_{j \in A} v_j(p + e^{ij}). \tag{5}$$

Time is continuous and agents update their actions via independent Poisson arrival processes with unit expectation. Thus, every agent updates once per unit of time in expectation. In the version of the model without heterogeneity (section 2.2), agents always choose a best response to the current state when they update. In the version with heterogeneity (section 2.3), agents will usually choose a best response but sometimes deviate. Let $\sigma_{ij}(p)$ be the probability that an agent playing i switches to j when the current state is p. Let $\sigma(p) = (\sigma_{ij})_{i,j \in A}$ represent the process thus defined. The process is a *perturbed best-reply process* if $\sigma_{ij}(p)$ is positive and small for all non-best replies.

The analysis below will be simplified by observing that the game $\mathscr{G}$ is a potential game (Monderer & Shapley, 1996).

Claim 1. $\mathscr{G}$ is a potential game with potential function $\rho : \tilde{\Delta} \to \mathbb{R}$ given by

$$\rho(p) = m \sum_{i \in A} p_i u_i - \frac{m}{2} \sum_{i \in A} \sum_{j \in A} p_i p_j s_{ij}. \tag{6}$$

Proof. See appendix. □

13 As we discuss in section 3.3.2, there is evidence that this assumption holds in Somalia.

14 The unit switching vector is necessary because an agent who switches action does not sanction herself.

Let ρ^* be the maximum potential across all states; that is,

$$\rho^* = \max_{p \in \tilde{\Delta}} \rho(p). \tag{7}$$

Finally, let γ be the smallest nonzero increase in payoff a player can get from switching actions. To be precise, let γ be the solution to:

$$\begin{aligned} \min_{i,j \in A,\, p \in \tilde{\Delta}} \quad & v_j(p + e^{ij}) - v_i(p) \\ \text{s.t.} \quad & p_i > 0, \text{ and} \\ & v_j(p + e^{ij}) > v_i(p). \end{aligned} \tag{8}$$

2.2 Best-response dynamics

We start by considering the case without heterogeneity, so that agents always choose best responses to the current state. In case of a tie, they pick an action at random. The probability of a player switching from i to j in state p is therefore

$$\sigma_{ij}(p) = \frac{\mathbf{1}_{B_i(p)}(j)}{|B_i(p)|}. \tag{9}$$

Let $p^i \in \tilde{\Delta}$ denote the state in which all players choose i. We will refer to p^i as a *norm*. We first observe that any strict equilibrium must be a norm.

Claim 2. If $p \in \tilde{\Delta}$ is a strict Nash equilibrium of $\mathscr{G}$, then p is a norm. Moreover, p^H is always a strict Nash equilibrium of $\mathscr{G}$.

Proof. For the first part, suppose to the contrary that p is a strict Nash equilibrium of $\mathscr{G}$ but is not a norm. Then there exist two actions i and j such that $p_i, p_j > 0$, $v_i(p) > v_j(p + e^{ij})$, and $v_j(p) > v_i(p + e^{ji})$. Since $v_i(p + e^{ji}) > v_i(p)$, we have $v_j(p) > v_i(p)$. But similarly $v_j(p + e^{ij}) > v_j(p)$, so $v_i(p) > v_j(p)$, which contradicts the previous statement.

For the second part, note that $v_H(p^H) = u_H$ and that for all $p \in \tilde{\Delta}$, $v_L(p) < u_H$ and $v_M(p) < u_H$. □

We say that a norm is *stable* if it is a Nash equilibrium of $\mathscr{G}$; it is *strictly stable* if it is a strict Nash equilibrium. As a shorthand, we will refer to p^i as norm i and we will say i is stable if p^i is stable. Claim 2 implies that there exists at least one strictly stable norm.

Next, we establish that the process converges to a strictly stable norm from any initial state. Recall that γ is the smallest nonzero increase in payoff from switching actions.

Theorem 1. From any initial state $p \in \tilde{\Delta}$, the process converges to a strictly stable norm in finite time with probability one, and the expected waiting time is at most

$$4\frac{\rho^* - \rho(p)}{\gamma}. \tag{10}$$

Proof. The theorem is a special case of theorem 3 below. □

Note that theorem 1 rules out the possibility of converging to a weak Nash equilibrium; intuitively, this is because from any weak Nash equilibrium there is a positive probability of entering the basin of attraction of a strictly stable norm.

We can now define the concept of stepping stone. For any two actions i, j, we will say i is *j-stable* if $v_i(p^i) \geq v_j(p^i + e^{ij})$. Clearly, i is stable if and only if it is j-stable for all $j \neq i$. It is straightforward to check that H is always stable and M is always L-stable. Action i is *strictly j-stable* if $v_i(p^i) > v_j(p^i + e^{ij})$.

We will say that action M is a stepping stone if it destabilizes L and is not strictly stable. Intuitively, this means that players at p^L will deviate to M but not H, after which players at M will deviate to H.

Definition 1. M is a *stepping stone* in the best-response model if L is H-stable and is not strictly M-stable, and M is not strictly stable.

If M is a stepping stone, then theorem 1 immediately implies that from any starting state, the process will converge to p^H. This is because when M is a stepping stone, H is the only strictly stable norm.

Proposition 1 establishes necessary and sufficient conditions for M to be a stepping stone. Recall that $\delta = 1/m$.

Proposition 1. M is a stepping stone if and only if

$$\begin{aligned} &\frac{s_{LH}}{u_H - u_L} \geq \frac{1}{1-\delta}, \\ &\frac{s_{LM}}{u_M - u_L} \leq \frac{1}{1-\delta}, \text{ and} \\ &\frac{s_{MH}}{u_H - u_M} \leq \frac{1}{1-\delta}. \end{aligned} \tag{11}$$

Proof. i is j-stable if and only if

$$v_i(p^i) \geq v_j(p^i + e^{ij}) \tag{12}$$

$$\iff u_i - \sum_{k \in A} s_{ik} p_k^i \geq u_j - \sum_{k \in A} s_{jk}(p_k^i + e_k^{ij}) \tag{13}$$

$$\iff u_i - u_j \geq \sum_{k \in A} (s_{ik} - s_{jk}) p_k^i - \sum_{k \in A} s_{jk} e_k^{ij} \tag{14}$$

$$\iff u_i - u_j \geq -s_{ij} + \delta s_{ij} \tag{15}$$

$$\iff u_i - u_j \geq -(1-\delta) s_{ij}. \tag{16}$$

Thus if $u_i \geq u_j$, i is always j-stable. If $u_i < u_j$, i is j-stable if and only if

$$\frac{s_{ij}}{u_j - u_i} \geq \frac{1}{1-\delta}. \tag{17}$$

Then the inequalities in expression (11) follow. □

Intuitively, the first inequality in expression (11) means that the social sanction for going from L to H should be high relative to the corresponding gains in intrinsic utility. The last two inequalities mean that the social sanctions for going from L to M and then from M to H should be small relative to the corresponding gains in intrinsic utility. In other words, M must be a good 'social substitute' for both L and H. A way to see this directly is to note that a necessary condition for M to be a stepping stone is the *reverse triangle inequality*:

$$s_{LH} \geq s_{LM} + s_{MH}. \tag{18}$$

We now ask how a stepping-stone transition happens. To avoid degeneracy, we consider the case where M is a strict stepping stone, in the sense that L is strictly H-stable. Let

$$q^* = \frac{u_M - u_H + (1-\delta)(s_{LH} - s_{LM})}{s_{LH} - s_{LM} - s_{MH}}. \tag{19}$$

We will see that q^* is the proportion of M-players at which agents start to switch to H.

Proposition 2. When M is a strict stepping stone, starting at p^L, agents will first deviate to M. When at least q^* of agents are playing M, L-players will start to switch to H, after which M-players also switch to H.

Proof. Let q be the proportion of agents at M and $1-q$ the proportion at L, so there are there are none at H. Let $v_{ij}(q)$ be the payoff a player at i gets from playing j given q. Since M is a strict stepping stone, we know that $v_{LM}(0) \geq v_{LL}(0) > v_{LH}(0)$. That is, at p^L, L-players switch to M with positive probability.

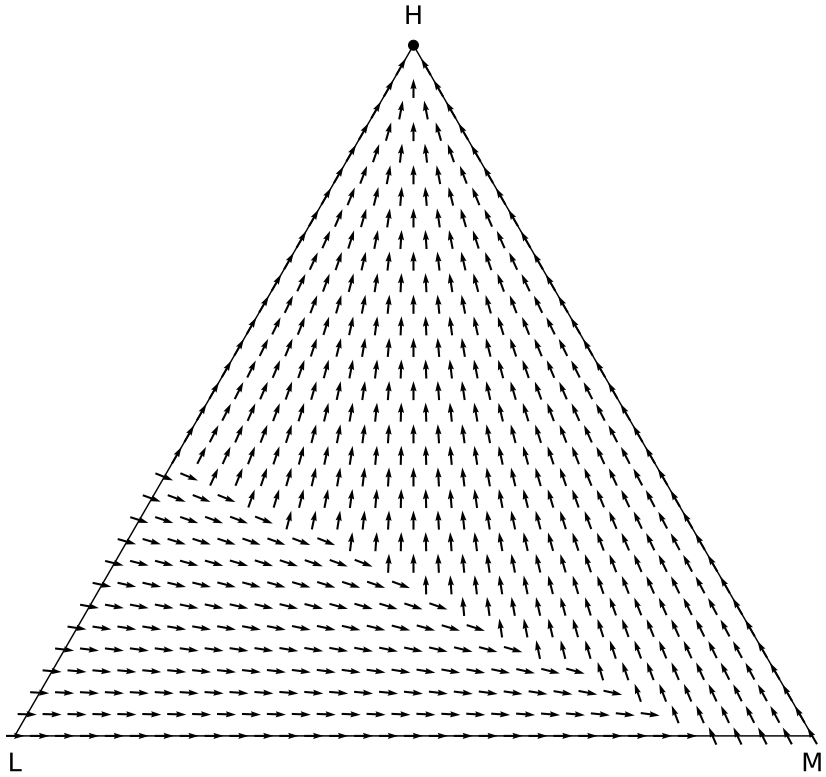
L-players start to switch to H when

$$\begin{aligned}
& v_{LH}(q) \geq v_{LM}(q) && (20)\\
\iff & u_H - (1-q-\delta)s_{LH} - qs_{MH} \geq u_M - (1-q-\delta)s_{LM} && (21)\\
\iff & q \geq \frac{u_M - u_H + (1-\delta)(s_{LH} - s_{LM})}{s_{LH} - s_{LM} - s_{MH}} && (22)\\
\iff & q \geq q^*. && (23)
\end{aligned}$$

Similarly, M-players start to switch to H when

$$\begin{aligned}
& v_{MH}(q) \geq v_{MM}(q) && (24)\\
\iff & u_H - (1-q)s_{LH} - (q-\delta)s_{MH} \geq u_M - (1-q)s_{LM} && (25)\\
\iff & q \geq \frac{u_M - u_H + s_{LH} - s_{LM} - \delta s_{MH}}{s_{LH} - s_{LM} - s_{MH}} && (26)\\
\iff & q \geq q^* + \delta. && (27)
\end{aligned}$$

Figure 2: M is a stepping stone.



So L-players will start to switch to H when least q^* of agents are playing M, after which M-players will also switch to H. □

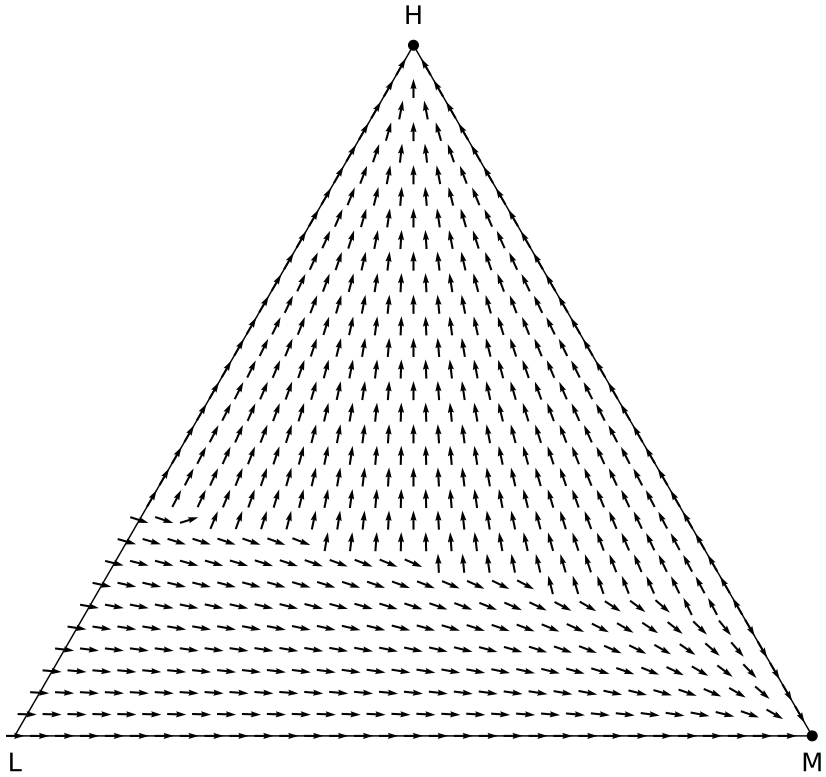
Proposition 2, and the behavior of the process in general, can be illustrated using phase diagrams. These become increasingly exact for large m.

Example 1 ($\boldsymbol{u_L = 0}$, $\boldsymbol{u_M = 0.5}$, $\boldsymbol{u_H = 1}$, $\boldsymbol{s_{LM} = 0.4}$, $\boldsymbol{s_{LH} = 1.5}$, $\boldsymbol{s_{MH} = 0.4}$)**.** Figure 2 shows an example where M is a stepping stone. The arrows show the direction of travel and the filled circle shows the unique fixed point of the process. Starting at p^L, the norm is destabilized as agents switch to M. As the process gets closer to p^M, action H becomes a best response and agents start to switch to H. The process converges to p^H.

Example 2 ($\boldsymbol{u_L = 0}$, $\boldsymbol{u_M = 0.5}$, $\boldsymbol{u_H = 1}$, $\boldsymbol{s_{LM} = 0.4}$, $\boldsymbol{s_{LH} = 1.5}$, $\boldsymbol{s_{MH} = 0.75}$)**.** Figure 3 shows an example where M is not a stepping stone. The only difference in the parameters is that s_{MH} is higher than in example 1. As before, from p^L, agents start to switch to M. However H never becomes a best response because M is stable. Instead the process converges to p^M.

Proposition 2 suggests a possible drawback of stepping stones: because the dynamics must spend time travelling towards the intermediate norm before converging to the

Figure 3: M is not a stepping stone.



high-utility norm, the introduction of a stepping stone does not unambiguously improve welfare. If it is possible to convince players to switch to the high-utility action directly this could take less time overall and lead to higher total welfare. In proposition 3, we provide estimates of the waiting time for direct and indirect transitions. In contrast to the result in theorem 1, which was an upper bound on convergence time, here we provide an estimate that becomes exact as m becomes large.

Recall that q^* is the fraction of M-players needed for agents to start switching to H when M is a stepping stone.

Proposition 3. Given $q \in (0,1)$, starting from p^L, let T be the expected waiting time to reach a state in which at least q of agents play H.

1. If H is the unique best response at p^L, then

$$T \approx \ln\left(\frac{1}{1-q}\right) \tag{28}$$

   when m is large.

2. If M is a strict stepping stone, then

$$T \approx \ln\left(\frac{1}{1-q^*-\delta}\right) + \ln\left(\frac{1}{1-q+\delta}\right) \tag{29}$$

when m is large.

Proof. First, consider the case in which H is the unique best response at p^L. Starting at p^L, each time a revision opportunity arises, if an L-player is selected she switches to H, and if an H-player is selected she stays at H. Recall that $\delta = 1/m$. In expectation, it takes 1 revision opportunity for the first switch to occur, $\frac{1}{1-\delta}$ for the second to occur, $\frac{1}{1-2\delta}$ for the third, and so on. In expectation, m revision opportunities occur per unit of time. Therefore, when m is large, the expected waiting time to reach a state in which q of players have switched to H is approximately

$$\delta + \frac{\delta}{1-\delta} + \frac{\delta}{1-2\delta} + \cdots + \frac{\delta}{1-q} \tag{30}$$

$$\approx \int_{1-q}^{1} \frac{1}{x} \mathrm{d}x \tag{31}$$

$$= \ln\left(\frac{1}{1-q}\right). \tag{32}$$

Next, consider the case in which M is a stepping stone. Again, we start at p^L. Initially, each time a revision opportunity arises, if an L-player is selected she switches to M, and if an M-player is selected she stays at M. By proposition 2, this continues until the fraction of players having switched to M is q^* or more. Thereafter, any L-player who revises switches to H, and as soon as this happens at least once, M-players also switch to H. The first step of the transition takes approximately a proportion $q^* + \delta$ of agents switching; the second step takes approximately a proportion q of agents. Thus by an analogous argument to the one in the preceding paragraph, when m is large, the expected waiting time to reach a state in which at least q of players have switched to H is approximately

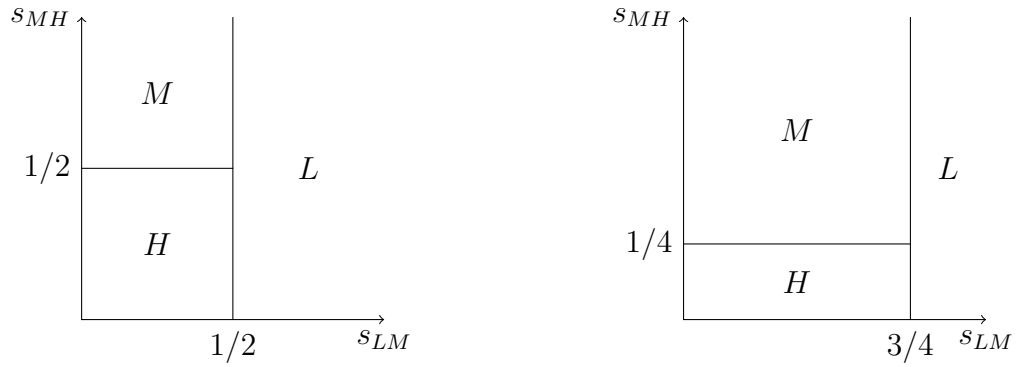
$$\ln\left(\frac{1}{1-q^*-\delta}\right) + \ln\left(\frac{1}{1-q+\delta}\right). \tag{33}$$

□

Thus, it is always faster to go from L to H directly (when that is possible) than via the intermediate action M. The time it takes to converge to H depends on the frequency of agents' revision opportunities. In the context of FGC we may think of the parents as taking the decision, and the frequency of decisions will depend on the spacing between daughters. Thus the rate of convergence to a higher-welfare equilibrium could be quite slow with or without a stepping stone.

To conclude our analysis, we look at convergence under different parameter values, starting at p^L. First, consider the case in which L is H-stable. If M is a stepping stone, the process will converge to H. Otherwise, it will either stay at L or converge to M. Figure 4 shows which norm the process converges to for different values of the parameters. In figure 4a we take $u_L = 0$, $u_M = 1/2$, $u_H = 1$, and $s_{LH} = 3/2$, and

Figure 4: Convergence under different values of s_{LM} and s_{MH} when $s_{LH} = 3/2$.



(a) $u_M = 1/2$.

(b) $u_M = 3/4$.

we study the process as s_{LM} and s_{MH} vary. (We also take the population size m to be arbitrarily large.) When s_{LM} is too large, the process never leaves L as it is stable. When s_{LM} is small but s_{MH} is large, the process stays at M. Only when both s_{LM} and s_{MH} are small enough – that is, when M is a close enough 'social substitute' for both L and H – does the process converge to H. This corresponds to the conditions for M to be a stepping stone in expression (11).

How is the process affected by changes in the intrinsic utility of different actions? In figure 4b, we increase u_M to 3/4. Changing u_M has two effects: on the one hand, L becomes less stable, in the sense that there are fewer parameter values for which the process stays at L; on the other hand, M becomes more stable and the process is more likely to stay stuck at M.

Next, consider the case in which L is not H-stable. Then the process will either converge to H or M. In figure 5a we keep $u_L = 0$, $u_M = 1/2$, and $u_H = 1$, and reduce s_{LH} to 3/4. The figure shows that the process will converge to M if s_{LM} is small enough and s_{MH} is large enough. If s_{LM} is too large, then H is more attractive than M starting from L, and the process converges directly to H. If s_{MH} is too small, then M is not stable so the process doesn't stay there. In figure 5b, we increase u_M to 3/4. This unambiguously increases the set of values of s_{LM} and s_{MH} for which the process converges to M.

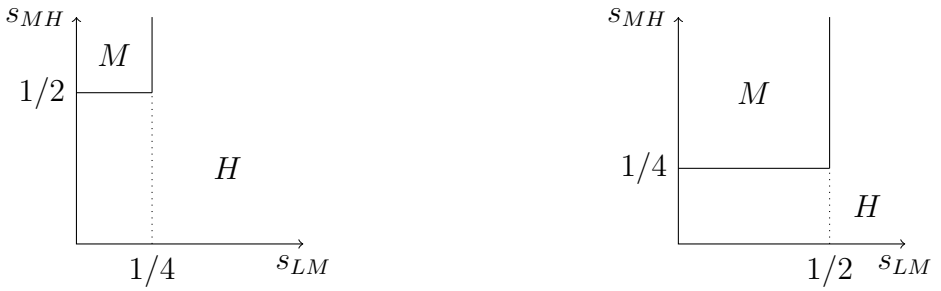
Finally, we summarise the welfare implications of the introduction of an intermediate norm. The analysis above allows us to distinguish two main cases:

Case 1: L is H-stable. In this case, introducing the intermediate norm M weakly improves welfare. The process will either stay at L, converge to M, or converge to H. But since it would otherwise stay at L, welfare is at least as high as before.

Case 2: L is not H-stable. In this case, the welfare impact of introducing M is ambiguous. In particular:

(i) if H remains the unique best response when M is introduced, then welfare is unaffected;

Figure 5: Convergence under different values of s_{LM} and s_{MH} when $s_{LH} = 3/4$.



(a) $u_M = 1/2$. (b) $u_M = 3/4$.

(ii) if M becomes a best response and is absorbing, then welfare is reduced;

(iii) if M becomes a best response and is not absorbing, then the process still converges to H but at a slower pace, so that welfare is reduced.

The main takeaway is therefore that intermediate norms have the potential to increase welfare when starting from a situation in which society is 'locked' into a bad equilibrium (in the sense that L would not be abandoned in favor of H). However, if the prevailing norm is L but society would spontaneously transition to H, then the introduction of M can only reduce welfare –or at best leave it unchanged.

2.3 Heterogeneity

We now consider the situation where agents' decisions are subject to heterogeneity. heterogeneity can arise from shocks to agents' preferences on the one hand, and from random deviations due to inattention, misperceptions, experiments, and the like on the other hand. A standard model for heterogeneity arising from random deviations is the *uniform error process* (Young, 1993; Kandori *et al.*, 1993; Jackson, 2008) defined by

$$\sigma_{ij}(p) = \frac{\varepsilon}{3} + (1-\varepsilon)\frac{\mathbf{1}_{B_i(p)}(j)}{|B_i(p)|}. \tag{34}$$

A standard model for heterogeneity arising from preference shocks is the logit distribution. This results when an agent's utility for different actions is subjected to i.i.d. shocks that are extreme-value distributed (Blume, 1993; McKelvey & Palfrey, 1995; Brock & Durlauf, 2001). In this case

$$\sigma_{ij}(p) \approx \frac{e^{\beta v_j(p)}}{\sum_{k\in A} e^{\beta v_k(p)}}, \tag{35}$$

with the approximation becoming increasingly exact as the population size increases.[15]

The distribution governing the agents' updating process determines the amount of

15 When the population size is infinite, fixed points of the process are guaranteed to exist. These correspond to McKelvey and Palfrey's concept of quantal response equilibrium (McKelvey & Palfrey, 1995).

heterogeneity in the population. Heterogeneity is increasing in ε (in the uniform model) and decreasing in β (in the logit model). It is important to note, however, that the *realized* degree of heterogeneity is constantly in flux due to the random nature of the updating process. In particular, some states are more likely to be observed than others and these states can be characterized theoretically using stochastic stability theory (Foster & Young, 1990; Kandori *et al.*, 1993; Young, 1993, 1998; Blume, 1993).[16]

As we shall see, heterogeneity can retard the transition time to the high welfare norm. It also means that at any given point in time, the prevailing norm will be followed by most people but not by everyone, which is typically what we see in practice.

Let $\sigma(p, \alpha)$ be a *perturbed best-response dynamic*, parameterized by $\alpha \geq 0$, where σ converges to a pure best-response process as $\alpha \to 0$. Namely, for every $\eta \in [0, 1)$ there is a value $\alpha(\eta)$ such that for all $\alpha \in [0, \alpha(\eta)]$ and all $p \in \Delta$, $\sigma(p, \alpha)$ puts probability at least $1 - \eta$ on all best replies at p. Examples include the logit with $\alpha = 1/\beta$ and the uniform error model with $\alpha = \varepsilon$. Let $p^t(\alpha, m)$ be the associated stochastic process when the population size is m.

Theorem 2 bounds the expected waiting time to reach the vicinity of p^H from p^L when noise is small. Recall that the notation '$O(1/\nu)$' means 'of the same order as $1/\nu$'.

Theorem 2. Suppose that M is a strict stepping stone in the best-response model. For all sufficiently small $\nu > 0$ there is a finite time $T_\nu \approx O(1/\nu)$ and threshold $\alpha_\nu > 0$ such that, for every $\alpha \in [0, \alpha_\nu]$ and all $m > 1/\nu$, the expected waiting time for the perturbed process to transit from p^L to a state p satisfying $p_H \geq 1 - \nu$ is bounded above by T_ν.

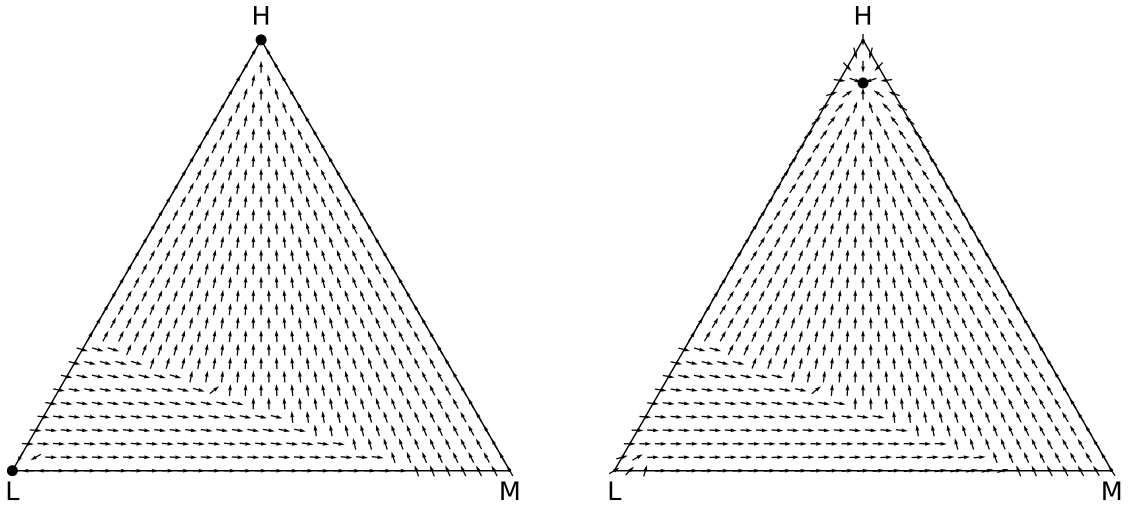
Proof. See appendix. □

In contrast, if M is not a stepping stone, the process can become trapped in the vicinity of p^L for a very long period of time. In particular, if there is a locally stable fixed point near p^L (in the associated mean-field dynamic), then the expected waiting time to transit from p^L to a neighborhood of p^H becomes exponentially large as the size m of the population grows (Benaim & Weibull, 2003).

When heterogeneity is present the introduction of an intermediate action M *can play the role of a stepping stone even though it is not actually a stepping stone in the pure best-response model*. To see why, suppose first that the agents are choosing between just L and H according to a perturbed best-response process $\sigma(p, \alpha)$. When the heterogeneity are sufficiently small, there will typically exist a locally stable fixed point near p^L, say $p^* = p^*(\alpha)$, that reflects a stable degree of heterogeneity in the two-action case. Now suppose that M is introduced as an intermediate alternative between L and H. We say that M is a stepping stone in the perturbed best response model if in state p^*, M is a

16 Under the logit updating process the long-run probability of every state p is proportional to $e^{-\beta\rho(p)}$, where $\rho(p)$ is the potential of state p. This holds for all finite population sizes.

Figure 6: M becomes a stepping stone with enough heterogeneity.



(a) M is not a stepping stone.

(b) M is a stepping stone.

best response and M is not itself stable. Note that this situation can arise even when M is not a best response to p^L itself. The intuition is that, in the presence of heterogeneity, there is always a strictly positive fraction of the population that is not playing L, hence it is easier to destabilize the norm p^L.

Example 3. Figure 6 shows an example in the uniform error case where M is not a stepping stone for $\varepsilon = 0$, but becomes one once ε is large enough. In left-hand diagram, when $\varepsilon = 0$, L is stable, so M is not a stepping stone. In the right-hand diagram, when $\varepsilon = 0.15$, there is no fixed point near p^L that is M-stable and M is itself not stable, so M is a stepping stone. Note that even a small degree of heterogeneity can lead to substantial changes in the dynamics.

2.4 More than three actions

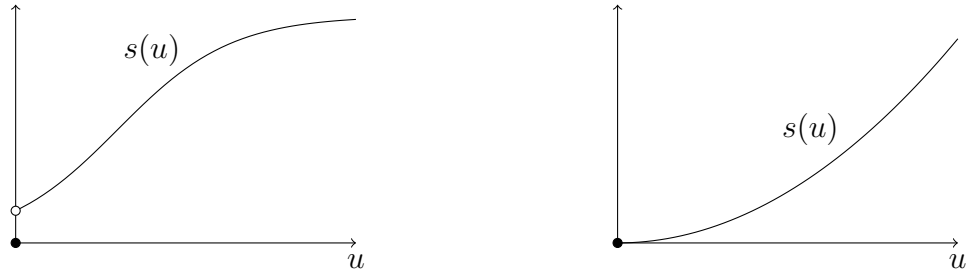
Finally we consider the general case with more than three actions. This will allow us to show that stepping stones are related to *increasing average sanctions*. For simplicity, we consider the case without heterogeneity, as in section 2.2.

There is a set of n actions $A = \{1, 2, \ldots, n\}$. Let u_i be the intrinsic utility of action i. We assume that utilities are distinct and indexed such that

$$u_1 > u_2 > \cdots > u_n. \tag{36}$$

We assume that the degree of social pressure that an agent imposes on another depends on the perceived distance between their two actions, as measured by the difference in their intrinsic utilities. Specifically, we will assume that there is a weakly increasing function

Figure 7: Two sanctions functions.



(a) $f(u)$ strictly decreasing. (b) $f(u)$ strictly increasing.

$s : \mathbb{R} \to \mathbb{R}_+$ such that $s(0) = 0$ and for all actions $i \neq j$, the sanction is equal to

$$s(|u_i - u_j|) > 0. \tag{37}$$

As before, there are m agents and p_i represents the proportion of agents playing action i. Let $p = (p_1, p_2, \ldots, p_n)$ be the *state* of the game and let $\tilde{\Delta}$ be the set of possible states.

The utility of an agent choosing action i in state p is

$$v_i(p) = u_i - \sum_{j \in A} p_j \cdot s(|u_i - u_j|). \tag{38}$$

Note that the n-action game remains a potential game with the potential function defined in equation (6).

As before, time is continuous and agents update their actions via independent Poisson arrival processes with unit expectation. We consider the best-response dynamics defined as in section 2.2.

Let A^* be the set of stable norms. Action 1 is always stable, and action $i > 1$ is stable if and only if for all $j < i$,

$$\frac{s(u_j - u_i)}{u_j - u_i} \geq \frac{1}{1 - \delta}. \tag{39}$$

The shape of the sanctions function plays a key role in the dynamics of the model. Define the *average sanctions function* as

$$f(u) = \frac{s(u)}{u} \tag{40}$$

for all $u > 0$. Of particular interest will be the case where f is (weakly) monotone, either increasing or decreasing. When f is decreasing, small differences in utility have a relatively greater impact than do large differences in utility. This condition is implied by concavity of s but is more general. When f is increasing, small differences in utility have a relatively smaller impact than do large differences in utility. This condition is implied by convexity of s but is more general.

The definitions of stability and j-stability carry over from section 2.2. Proposition 4 characterizes the set of stable norms A^* when f is monotone.

Proposition 4. If f is weakly decreasing, then there exists i such that $A^* = \{j \in A : j \leq i\}$. If f is weakly increasing, then $i > 1$ is stable if and only if

$$\frac{s(u_{i-1} - u_i)}{u_{i-1} - u_i} \geq \frac{1}{1-\delta}. \tag{41}$$

Proof. For the first part, suppose f is monotone decreasing. Suppose i is a norm. We show that any $j < i$ is also a norm. If $j = 1$ this holds straightforwardly, so suppose $j > 1$. Since i is a norm, we have for all $k < j$,

$$f(u_k - u_i) \geq \frac{1}{1-\delta}.$$

But since f is monotone decreasing, we have for all $k < j$,

$$f(u_k - u_j) \geq f(u_k - u_i) \geq \frac{1}{1-\delta}.$$

So j is a norm.

For the second part, suppose f is monotone increasing. Action i is a norm if and only if for all $j < i$,

$$f(u_j - u_i) \geq \frac{1}{1-\delta}.$$

But since f is monotone increasing, $f(u_j - u_i)$ is minimised for $j = i - 1$. Therefore i is a norm if and only if

$$f(u_{i-1} - u_i) \geq \frac{1}{1-\delta}, \tag{42}$$

as required. □

Proposition 4 follows from the fact that when f is decreasing, the best action for a player to deviate to (starting from a norm) is always action 1. This is because social pressure is less important relative to individual utility for actions that are further away. In contrast, when f is increasing, the best action for a player to deviate to is always the next closest action, because actions that are further away have a higher relative social pressure. This dynamic is the core of the n-action model.

We now show that the process will tend to converge to a norm in finite time; this result is a generalization of theorem 1. Recall that γ is the smallest nonzero increase in payoff from an agent switching actions.

Theorem 3. From any initial state $p \in \tilde{\Delta}$, the process converges to a strictly stable norm in finite time with probability one, and the expected waiting time to reach a norm

is at most

$$4\frac{\rho^* - \rho(p)}{\gamma}. \tag{43}$$

Proof. See appendix. □

We define the concept of stepping stone in the n-action model as follows:

Definition 2. Given actions $i < j < k$, action j is a stepping stone from k to i if k is i-stable but not j-stable, and j is not i-stable.

Proposition 5. Suppose there exist actions $i < j < k$ such that j is a stepping stone from k to i. Then f is not monotone decreasing.

Proof. Since j is a stepping stone from k to i, we have

$$f(u_i - u_k) \geq \frac{1}{1-\delta} > f(u_j - u_k). \tag{44}$$

But $u_i - u_k > u_j - u_k$, so f is not monotone decreasing. □

Proposition 6. If f is monotone decreasing, then from any unstable state p^i, the process converges to p^1.

Proof. Let $i > 1$ be unstable. Let $v_{ii} = u_i$ be the utility agents get from playing i at p^i. Let $v_{ij} = u_j - (1-\delta)s(u_j - u_i)$ be the utility agents get from playing $j < i$. Note that

$$v_{ij} - v_{ii} = u_j - u_i - (1-\delta)s(u_j - u_i), \tag{45}$$

$$= (u_j - u_i)\left(1 - (1-\delta)f(u_j - u_i)\right). \tag{46}$$

If f is monotone decreasing, then $v_{ij} - v_{ii}$ is uniquely maximised when $u_j - u_i$ is maximised. So $j = 1$ is the unique best response. Hence the process converges to p^1. □

2.5 Discussion

Propositions 5 and 6 generalize the results of section 2.2 to the case with n actions. They highlight the key role played by the shape of f, that is, the strength of social sanctions in relation to intrinsic utility differences. In particular, they establish that stepping stones are related to increasing average sanctions. Estimating the relationship between social sanctions and intrinsic utility differences empirically poses an empirical challenge; in section 3 we bring descriptive survey data from Somalia to bear on this problem.

In the interest of analytical tractability the preceding model makes several simplifying assumptions that are fairly standard in the social interactions literature, but that may or may not be empirically valid. One is the assumption that everyone knows the exact prevalence of each action in society, and therefore can precisely anticipate social sanctions.

This assumption is most plausible in small communities; in larger communities it may be more reasonable to assume agents only have partial information, say based on the actions of a subset of the community.

Two other assumptions are that social sanctions depend on the intrinsic difference in payoffs between two actions i and j – an assumption we introduce in the n-actions version of the model – and that they are symmetric ($s_{ij} = s_{ji}$). These assumptions are made largely for analytical convenience, but may be reasonable to a first approximation. The analysis of both the long-run and intermediate-run dynamics can be conducted without these assumptions using the theory of perturbed Markov processes (Young, 1993, 1998; Kandori *et al.*, 1993).

3 Empirical Application: FGC in Somalia

In this section we use originally collected data from Somalia to test some of the model assumptions and illustrate the transition dynamics for a particularly harmful norm: female genital cutting (FGC).

3.1 Context

We start by giving some background on FGC in general and on FGC in Somalia in particular. FGC is the practice of cutting or removing part of the external female genitalia for non-medical reasons. An estimated 200 million women are cut worldwide (UNICEF, 2016) and every year, 3 million female infants and children are at risk of undergoing FGC (Spisma *et al.*, 2012). This practice is also quite geographically dispersed, being present in 29 African and Middle Eastern countries (Camilotti, 2016). In some of these countries FGC is almost universal: the share of cut women is 98% in Somalia, 96% in Guinea, 93% in Djibouti, and 91% in Egypt and Sierra Leone (Yoder *et al.*, 2013). In Somalia, the context of our study, the number of women who have undergone FGC is approximately 6 million (UNICEF, 2013).

FGC is a harmful practice, as it leads to serious health consequences both at the time of cutting (e.g., excessive bleeding and increased mortality) and in the long run (e.g., birth-related complications). Adam *et al.* (2010) estimate that across six African countries and within a cohort of 15-year-old women, a loss of 130,000 life years is expected due to FGC's association with obstetric hemorrhage. Also, given that FGC is generally performed on young girls without their consent, it is considered a human rights violation (Mackie, 1996).

WHO distinguishes 3 main types of female circumcision: *Type I* is the partial or total removal of the clitoris and/or the prepuce (clitoridectomy); Type II is the partial or total removal of the clitoris and the labia minora, with or without excision of the labia majora (excision); *Type III*, also known as infibulation, is the narrowing of the vaginal orifice

with the creation of a covering seal by cutting and appositioning the labia minora and/or the labia majora, sometimes through stitching.[17]

We will focus on this difference in Somalia, where FGC is divided into two broad categories: 'Sunna' (types I-II) and 'Pharaonic' (type III). Historically, Pharaonic circumcision was the dominant type in Somalia, practiced almost universally (Abdalla, 1982). Sunna was introduced in Somalia much later than the Pharaonic type.[18] In 1984 the Inter-African Committee on harmful traditional practices affecting the health of women and children was established and in the 1990s pushes for abandonment gained momentum. Anecdotal evidence suggests that there was "a shift from infibulation to Sunna [...] as a result of FGC campaigns that have been emphasizing health effects of infibulation" (MOLSA, 2009). This shift was facilitated by religious leaders who opposed Pharaonic as harmful and non-Islamic, but supported Sunna as in line with Islam (Newell-Jones, 2016). More recently, a religious Fatwa issued in 2018 in Somaliland clearly banned the practice of Pharaonic circumcision, but not of the Sunna type.[19]

3.2 Data Description

We use data from a household survey that we conducted in 141 communities in Somalia between January and May 2020. The communities are located in Somaliland and Puntland regions of Somalia.[20] Appendix Figure B.1 shows the locations of these communities. The survey collected information on a sample of $4,130$ individuals – on average 29 per community – roughly equally split by gender ($2,040$ men and $2,090$ women). The respondents were sampled from the list of participants in community meetings that were conducted as part of an ongoing project (Gulesci *et al.*, 2020).[21]

The survey collected information on household demographics, socioeconomic status, and detailed questions on gender norms. In particular, we elicited information on respondents' own attitudes as well as their perceptions of community members' attitudes towards different types of FGC. We also collected the history of FGC types in the respondents' family, a list of those among their daughters that were cut (or were intended

17 A fourth type includes "all other harmful procedures to the female genitalia for non-medical purposes, e.g. pricking, piercing, incising, scraping and cauterizing the genital area," but traditionally the main types considered are types I to III.

18 Historical accounts suggest that the intermediate type of cutting was first introduced in Sudan, as a result of the prohibition of infibulation by the British colonial authorities in 1945 (El Dareer, 1982; Slack, 1988).

19 The text of the Fatwa is the following: "It's forbidden to perform any circumcision that is contrary to the religion which involves cutting and sewing up, like the pharaoh circumcision. *(...)* Any one proven to be performing the practice will receive punishment depending on the extent of the violation." (Ahmed *et al.*, 2018)

20 More specifically, the communities in our sample are located in the districts of Badhan, Buraou, Erigabo, Galdagob, Galkayo and Hargeisa.

21 These meetings were organized in collaboration with the NGO *Save the Children* and were facilitated by local personnel trained in interpersonal communication on sensitive topics.

Table 1: Summary statistics

	(1) *Pharaonic*		(2) *Sunna*	
	Mean	Std. dev.	Mean	Std. dev.
Age of FGC	9.090	(2.088)	8.543	(1.745)
Decision to cut by:				
Mother	0.779	(0.417)	0.895	(0.307)
Father	0.326	(0.471)	0.408	(0.492)
Grandmother	0.137	(0.346)	0.041	(0.199)

Notes: Sample is restricted to women (female respondents and their daughters) who have been cut. "Age of FGC" is the age at which the individual was cut. "Decision to cut by" is the fraction of respondents reporting that the decision to cut was taken by, respectively, their mother, father, or grandmother.

to be cut) and those that were uncut (and intended to remain so). For all cut daughters, we asked what type of FGC was performed on them. Finally, we elicited expectations about community reactions to deviations from FGC, to understand the sanctions that are commonly used to sustain the norm.

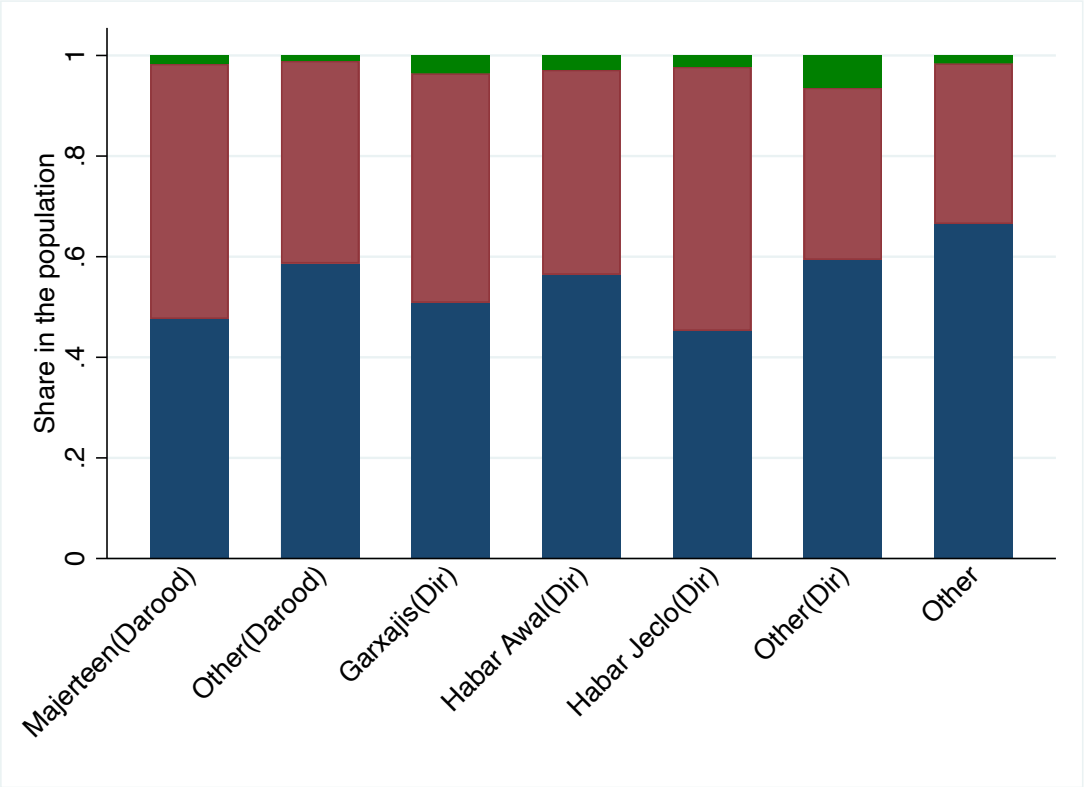
Table 1 presents summary statistics on the types of FGC. We see that on average, girls are cut at the age of nine and this is true for both Sunna and Pharaonic FGC. The decision to cut is most often taken by the mother. Mothers are reported as making the decision about the circumcision for 78 percent of girls who are Pharaonic cut and 90 percent of the girls who are Sunna cut. Fathers are reported to have played a role in the decision for 33 percent of the girls who are Pharaonic cut and 41 percent of those Sunna cut. In a minority of cases (14 and 4 percent, respectively) the girl's grandmother is also reported to have played a role in the decision.

Figure 8a shows the prevalence of FGC by type for girls and women aged 12 and above, across sub-clans. Our sample includes two main clans, Darood and Dir, which can be further subdivided into two and four sub-clans, respectively. Two facts emerge from the figure: first, the fraction of uncut girls and women is very small, on average, across all sub-clans. Second, neither clans nor sub-clans specialize in a given type of FGC: while not exactly a 50:50 split, all sub-clans have sizeable fractions of women who are Pharaonic as well as Sunna-cut.

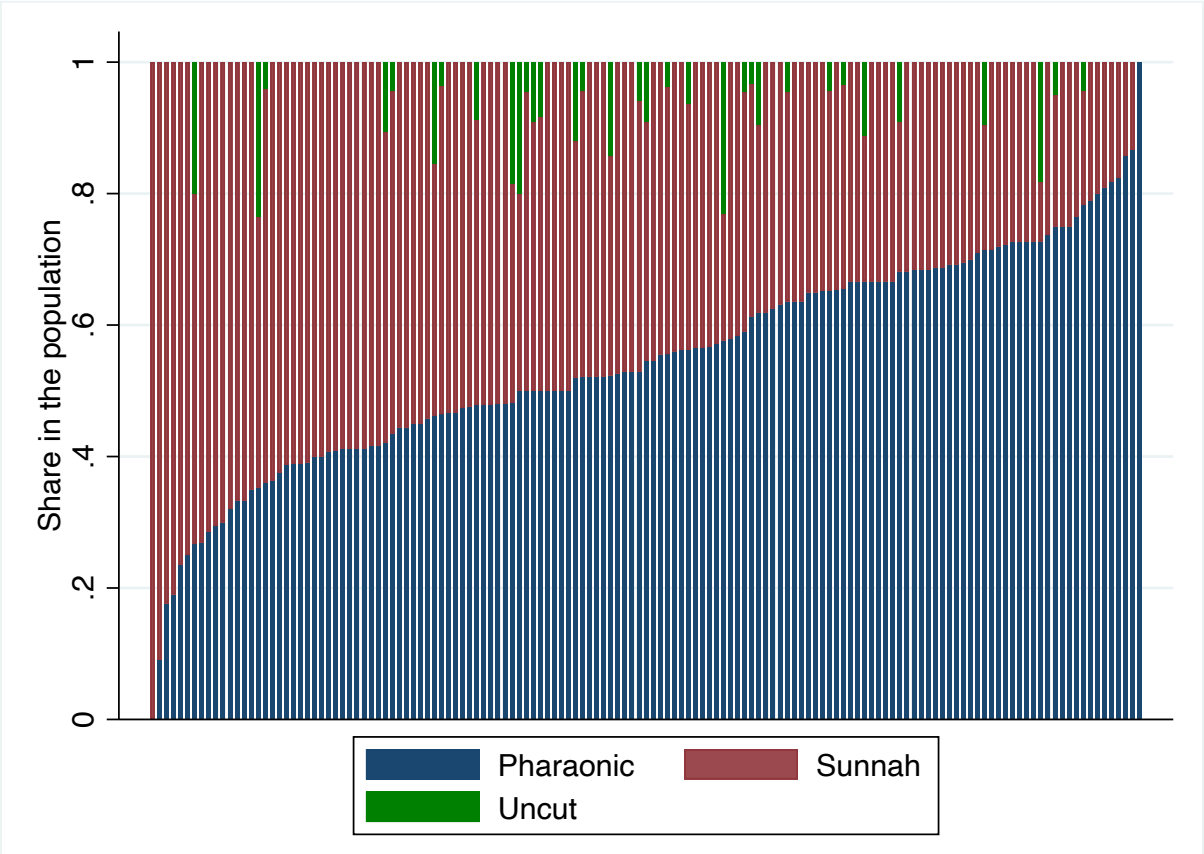
In Figure 8b we show the prevalence of FGC types at the community level (each bar corresponds to one of the 141 communities in our sample). At this level we find considerably more variation, both in the share of uncut women and, especially, in the share of Pharaonic vs. Sunna-cut. While in many communities the two types of FGC are almost equally prevalent, there is a considerable number of communities in which one form dominates. Below we relate this heterogeneity to the predictions of our model.

Figure 8: FGC Types

(a) By sub-clan



(b) By community



Notes: Sample is restricted to respondents and their daughters aged 12 or older. Sample shares: 17% Majerteen(Darood), 28% Other(Darood), 16% Garxajis(Dir), 14% Habar Awal(Dir), 13% Habar Jeclo(Dir), 7% Other(Dir), 5% other.

Table 2: Evidence on assumption: $u_L < u_M < u_H$

	(1) *Pharaonic*		(2) *Sunna*	
	Mean	Std. dev.	Mean	Std. dev.
Panel A: Complications experienced by respondents				
Any health complication (Yes=1)	0.626	(0.484)	0.115	(0.319)
N	1,437		406	
Panel B: Complications experienced by daughters				
Any health complication (Yes=1)	0.586	(0.493)	0.030	(0.170)
N	1,641		3,161	
Panel C: Perceived health complications				
Any perceived complication	0.749	(0.434)	0.062	(0.241)
Infection	0.402	(0.491)	0.043	(0.202)
Bleeding	0.566	(0.496)	0.030	(0.170)
Difficulty in delivery	0.606	(0.489)	0.036	(0.186)
Reduction in sexual feeling	0.463	(0.499)	0.032	(0.176)
Difficulty in penetration	0.326	(0.469)	0.004	(0.063)
Other	0.013	(0.115)	0.004	(0.063)
Number of complications	2.376	(1.732)	0.165	(0.731)
N	975		3,779	

Notes: In Panel A, the sample in column 1 (2) includes female respondents who reported that they were cut Pharaonic (Sunna). In Panel B, the sample in column 1 (2) includes respondents who reported that their daughter/s was/were cut Pharaonic (Sunna). In Panel C, the sample in column 1 (2) includes respondents who reported that Pharaonic (Sunna) cut was practiced in their community.

3.3 Evidence on model assumptions

We start by using our data to assess the validity of two key assumptions of our model. We focus on the three-action version of the model, as this is the relevant one for the empirical application to the Somali context.

3.3.1 Ranking of intrinsic utilities

Our model assumes that the intrinsic utilities satisfy $u_L < u_M < u_H$. To assess the validity of this assumption, we use as a proxy for u_i the health costs associated to the various alternatives, as reported by respondents in our survey. In other words, we test whether respondents are aware that Pharaonic circumcision is more likely to lead to health complications, compared to Sunna (the 'Uncut' option obviously has no health complications).

Table 2 shows that women in our sample are more likely to have experienced health complications if they had Pharaonic circumcision as opposed to Sunna. In particular, about 63 percent of Paraonic-cut respondents report having experienced health complic-

ations due to FGC, while the corresponding rate is 12 percent for Sunna-cut women. When asked about their daughters, 59 percent of respondents with daughters who have had Pharaonic circumcision report that their daughter(s) experienced health complications, while the corresponding rate for Sunna is only 3 percent. This supports the assumption that $u_L < u_M < u_H$.

In line with this, the lower panel of Table 2 shows that, when we break down responses by type of health complication expected from the different procedures, each and every category is perceived as more likely under Pharaonic circumcision. For example, 40 percent of respondents believe infection is a likely consequence of Pharaonic circumcision, ten times more than for Sunna; 57 associate severe bleeding with Pharaonic circumcision may, and only 3 percent with Sunna. Similarly, 61 percent believe Pharaonic circumcision may cause difficulties in giving birth, while only 4 percent think that Sunna would do so. Reductions in sexual feeling and difficulty in penetration are also much more associated with Pharaonic than with Sunna.

Overall, to the extent that health implications are a good proxy for intrinsic utility in our model, the pattern in Table 2 is in line with the assumption that $u_L < u_M < u_H$. However, one important aspect to notice – to which we return below – is that the perceived health costs of Sunna are rather low: when asked a separate question about expected (not experienced) health complications, only 6 percent of respondents thought that Sunna circumcision would lead to any health complications.

3.3.2 Ranking of social costs

Another important assumption in our model concerns the ranking of the social parameters s_{ji}.To measure these parameters, we need a proxy for how much pressure individuals who choose type j exert on individuals choosing type i. For this purpose, we asked respondents a series of questions specifically designed to elicit bilateral comparisons between Pharaonic, Sunna, and Uncut.

We presented each respondent with different situations where hypothetical parents have cut their daughter with a certain type of FGC, but they may get a daughter-in-law with a different type of FGC.[22] For example, we asked: "Suppose a mother and father in your community chose Pharaonic circumcision for their daughter, but their son wants to marry a girl with Sunna. How would these parents feel?". The possible answers were: happy, indifferent, or unhappy. We repeated the same question for Sunna vs. Uncut and for Pharaonic vs. Uncut. Appendix Figure B.2 provides a visual summary of the responses. The rationale underlying these questions is not to ask respondents how

22 We chose to frame this in the context of marriage choices because most of the literature on FGC highlights consequences in the marriage market as a potential cost for deviating from prevailing norms (see Wagner, 2015, for cross-country evidence).

Table 3: Evidence on assumption: $s_{LH} > s_{LM}$ and $s_{LH} > s_{MH}$

s_{LH}	0.594
	(0.312)
s_{LM}	0.415
	(0.311)
s_{MH}	0.459
	(0.267)
Hypothesis	p-value
$H_0 : s_{LH} \leq s_{LM}$	0.000
$H_0 : s_{LH} \leq s_{MH}$	0.000

Notes: s_{ji} is based on the question "Suppose a mother and father in your community chose type j circumcision for their daughter, but their son wants to marry a girl with type i circumcision. How would the parents feel?". The responses are "Happy", "Indifferent" or "Unhappy". We code them as **Happy=0, Indifferent=0, Unhappy=1** so that s_{ji} is increasing in the degree of disapproval of the parents (as in the model) and take the community average.

they themselves would feel, but to elicit second-order beliefs about the attitudes of other community members. In fact it is other people's views that matter if we want to measure expected sanctions for noncompliance with local norms.[23]

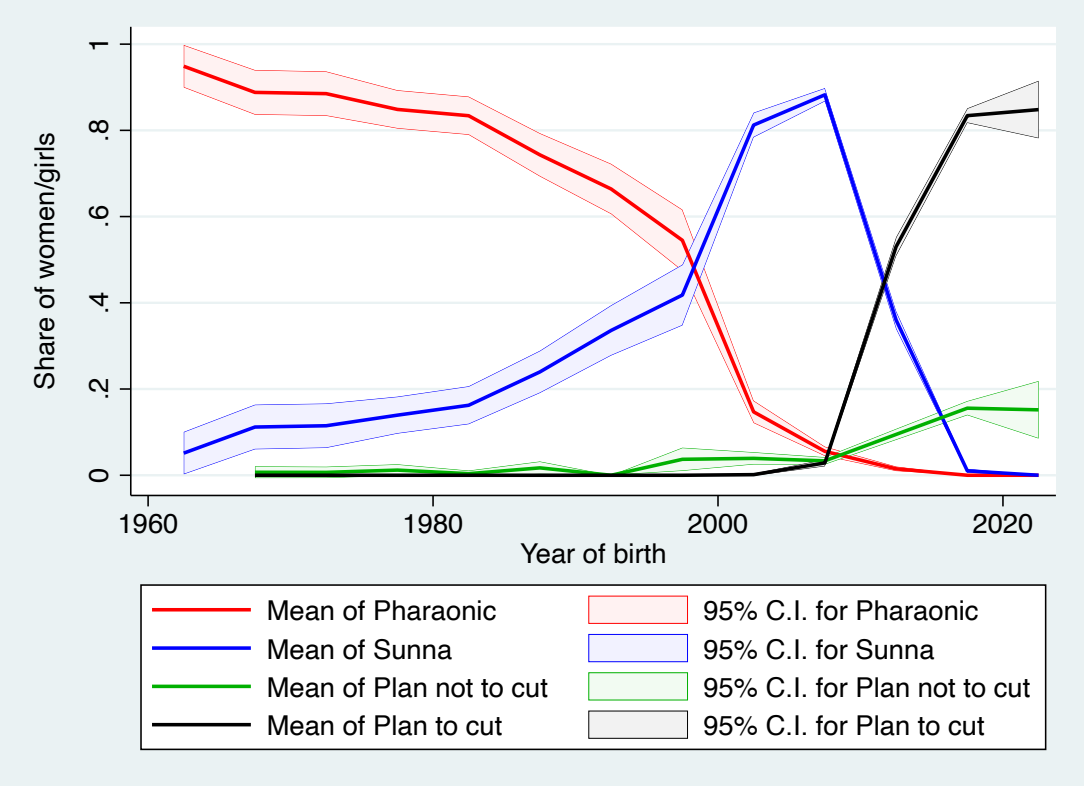
Using these data, we test the assumption about the ranking of social costs provided in expression (3) – namely that $s_{LH} > s_{LM}$ and $s_{LH} > s_{MH}$. To do so, we calculate the fraction of female respondents who said that parents who chose type j for their daughter would be unhappy if their son married a girl of type i.[24]

Table 3 shows that 59.4 percent of respondents in the average community think that parents who chose Pharaonic for their daughter would be unhappy if their son married an uncut girl. Thus, s_{LH} can be approximated by 0.594 for the average community in our sample. The corresponding figures for s_{LM} and s_{MH} are 0.415 and 0.459, respectively, clearly implying that $s_{LH} > s_{LM}$ and $s_{LH} > s_{MH}$. Statistically, we can reject the null hypotheses that $s_{LH} \leq s_{LM}$ or that $s_{LH} \leq s_{MH}$ at 99 percent confidence level. Hence, we conclude that the assumption on the ranking of social costs s_{ji} seems to hold for the average community in our sample.

23 Bicchieri (2005, p. 15 and ff.) emphasises that this is a key feature of social norms.

24 We pool 'happy' and 'indifferent' into one category because the model parameter $s_{ji} > 0$ captures a utility cost, hence we need a measure of disapproval – embedded in the response 'unhappy'. We focus on female respondents since the pattern in Table 1 shows that matriarchs play a greater role in the FGC decision. Therefore, their perceptions of the social costs are likely to matter more. The results are qualitatively similar if we use both male and female respondents' responses.

Figure 9: Type of FGC by year of birth



Source: Authors' calculations on original data from Somalia. Sample includes women aged 17–82 at the time of interview and their daughters aged 12 or older (including uncut ones).

3.4 Evidence on transition dynamics

Next, we explore the equilibrium dynamics generated by our model and assess the validity of the model's predictions.

3.4.1 Evolution across cohorts

We start by exploring the 'big picture' of how adherence to different cutting norms has evolved over time. Figure 1, briefly discussed in the Introduction, plots the share receiving Pharaonic circumcision (red line) or Sunna (blue line) against the year in which cutting occurred. The sample includes female respondents and those among their daughters (aged 12 and above) who have been cut. Figure 9 includes *all* female respondents and daughters aged 12 and above – whether they were cut or not – and plots cutting shares against the individual's year of birth.

Figure 1 shows that in the 1970s and 1980s virtually all women who were cut were Pharaonic-cut. Figure 9 shows, for the same time period, that virtually every woman born in those years was cut, as the share uncut (green and black lines) is flat at zero. Based on these data, it is realistic to study the transition dynamics as starting from an equilibrium where Pharaonic circumcision is the dominant trait.

Figure 1 also shows that, starting from the early 1990s, the rate of Sunna-cut women has been steadily rising and girls circumcised in the mid-2000s or after are more likely to be Sunna-cut as opposed to Pharaonic. Nowadays, Sunna has clearly replaced Pharaonic circumcision as the dominant trait for younger generations in the average community.

The timing of this transition roughly coincides with the period in which human rights campaigns made a strong push against Pharaonic circumcision and religious leaders started emphasizing that this type of FGC was not a requirement of Islam – as we discussed in Section 3.1. In terms of our model, this can be represented in two, non-mutually exclusive ways. The first is a reduction in the value of s_{LM}, that is the social sanction imposed by people who support Pharaonic on those who choose Sunna. In a religious country like Somalia, the Imams' endorsement of the latter clearly makes it more socially acceptable to abandon the prevailing Pharaonic norm and shift to the less prevalent Sunna. The second interpretation is that the human rights campaigns and the position taken by religious leaders may have conveyed new information on the intrinsic (dis)utility of the various alternatives, decreasing u_L and/or increasing u_M. In both cases, these parameter changes make it more likely that M destabilizes L starting from equilibrium p^L (see the discussion about figures 4 and 5 in section 2.2).

The next question is whether the now dominant Sunna trait is a 'stepping stone' and will ultimately disappear – leaving 'Uncut' as the norm in the long run – or if it has become the new norm and will remain so in the future, in the absence of new exogenous shocks. To address this question we take several steps.

First, to understand if the trait 'Uncut' may begin to penetrate, one needs to study the full sample, including girls who are not cut. In Figure 9 we plot the shares of cut and uncut girls/women by birth cohort. The red and blue lines represent the shares of Pharaonic and Sunna-cut, respectively. For uncut girls, we asked their mothers whether they intended to cut them in the future or not. The black line represents the share of girls/women who are currently uncut, but intended to be cut in the future. For this group, 'Uncut' may be interpreted as a temporary condition, most likely due to the (young) age of the girl. The green line represents girls who are uncut and whose mothers state that they will not be cut in the future: these are the families that have chosen trait H.

Figure 9 shows that from the early 2000s we start seeing a slight increase in the fraction of girls who are not cut (black and green lines). Given that the average age of cutting in our sample is 9, cohorts born after 2011 are still at risk of being cut (recall that the data was collected in 2020). For the youngest cohorts in our sample, approximately 15 to 20 percent of the *mothers* report that they intend not to cut them (green line). Given that this share was virtually zero for most of the period, this may be seen as encouraging. However, more than 80 percent of the respondents whose daughters are young and uncut report that they intend to cut their daughters, suggesting that Sunna may as well remain

the dominant trait.[25] [26]

3.4.2 Transitions and model parameters

The two conditions embedded in proposition 1 for transitioning from L to M and from M to H are that social costs of moving from Pharaonic to Sunna (s_{LM}) and from Sunna to Uncut (s_{MH}) are low relative to the respective gains in intrinsic utility ($u_M - u_L$ and $u_H - u_M$). A rigorous test of the conditions in proposition 1 would require a metric that allows us to compare s_{ji} and $u_i - u_j$ cardinally, possibly entailing strong assumptions. We therefore prefer to explore our empirical proxies for these parameters in a descriptive way, bearing in mind that any result we present should be seen more as suggestive of a certain direction than as a proof that a specific equation or inequality is satisfied.

We start by discussing the average values of these parameters as reported in tables 2 and 3. Recall that our proxy for s_{LM} is the share of respondents who think that parents choosing Pharaonic for their daughters would be unhappy if their son married a Sunna-cut girl. This share is .41 in Table 3. Our proxies for u_L and u_M are the shares reporting complications from Pharaonic and Sunna, respectively. Table 2 showed that, among mothers, 63 percent had complications from Pharaonic and 11 percent from Sunna, while among daughters the corresponding figures were 59 and 3 percent. The difference $u_M - u_L$ is thus .51 for mothers and .56 for daughters, in both cases greater than .41. While we already stressed that we prefer not to give a strict cardinal interpretation to these differences, the order of magnitude suggests that qualitatively it seems plausible that in the aggregate the conditions for transitioning from Pharaonic to Sunna hold in the aggregate, and indeed we saw that transition taking place in the cohort analysis.

A similar statement cannot be made for the relation between s_{MH} and $u_H - u_M$. Table 3 shows that on average 46 percent of respondents thought that parents who chose Sunna for their daughter would be unhappy if their son married an uncut girl, indicating a relatively high value for $s_{MH.}$ As discussed above, however, Table 2 shows that the share of health complications from Sunna is quite low. Furthermore, when asked about 'perceived' complications from Sunna (in a separate question), only 6 percent of respondents associate any complications with Sunna, and the share perceiving a risk of specific problems (those in the bottom panel of Table 2) is 4 percent or less for all items. This suggests that $u_H - u_M$ may be quite small and that the condition for transitioning from M to H may not hold in the average community. Thus, if one looks at the aggregate picture, Sunna

25 To have a definitive answer one should of course wait many years and see whether these intentions materialize.

26 Another reason why the increase in the share mothers who plan not to cut their daughters does not allow us to infer that the process will eventually transition to Uncut is that in the presence of heterogeneity, our model predicts that we would see an increase in the share of Uncut even if Sunna were absorbing (e.g., assuming logit noise).

appears less as a stepping stone and more like a new equilibrium norm.

Inferences made on the basis of average values of the parameters in the full sample may be misleading, however, if the relevant reference group for individual decision makers is smaller than the whole region, for example if it is their community. In other words, the fact that the conditions for transitioning from Sunna to Uncut do not appear likely to be satisfied in the aggregate does not mean that this holds for each and every community. Recall that Figure 8b displayed a significant degree of variation across communities both in the type of FGC and in the share of women who were not cut. Appendix Figure B.3 shows that the social factors s_{ij} also vary considerably across communities. We thus proceed to assess whether the conditions of our model for a successful stepping stone transition may hold when we exploit variation across communities.

First, we assess whether communities with lower s_{LM} are more likely to have made the first step (that is, to have transitioned from Pharaonic to Sunna).[27] In Panel (a) of Figure 10 we plot the actual share of Sunna-cut women in a community against the estimated s_{LM} for that community. Panel (b) reports a similar plot, but the vertical axis shows the residual from a regression of the share Sunna-cut on u_L and u_M, in order to isolate the correlation with s_{LM} after conditioning on intrinsic utilities – in line with the inequality in proposition 1. In both cases we see that communities with higher sanctions (higher s_{LM} or higher residual) have a smaller share of Sunna-cut girls, as expected. In the top panel the correlation is -0.16 (p-value 0.000), which implies that a 10 percentage point increase in s_{LM} is associated with a 1.6 percentage point reduction in the share of Sunna-cut. The magnitude and significance are very similar in the 'conditional' version (bottom panel): the estimated correlation is -0.13 with a p-value of 0.001. This is due to the fact that in our data u_L and u_M exhibit very little variation across communities. Therefore, in line with proposition 1, communities with lower s_{LM} are more likely to have made the first step of transitioning from Pharaonic to Sunna.

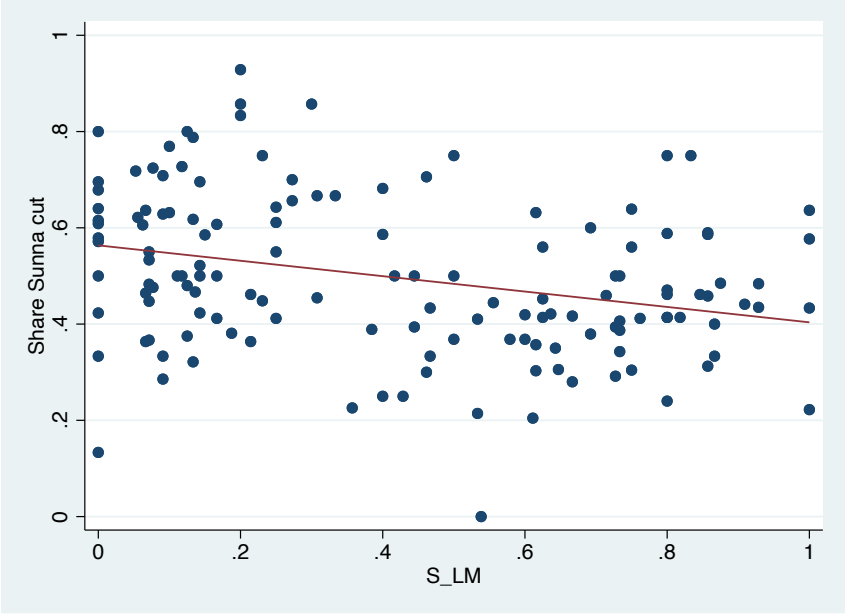
Next, we want to explore whether Sunna has the potential to be a stepping stone, i.e., to lead to an Uncut norm in the coming years, in a certain set of communities. Recall that our model predicts that, if Sunna is a stepping stone, people will switch from Sunna to Uncut when s_{MH} is small relative to $u_H - u_M$ and that the switch will be triggered by a threshold q^* of people having already chosen Sunna (proposition 2). To test this, we separate communities into those with relatively low sanctions s_{MH} and those with relatively high sanctions. It is only in the former set of communities that we expect a stepping stone transition might take place.

In Figure 11, we plot the relationship between the likelihood that a girl is currently 'uncut and planned not to be cut' (on the vertical axis), against the share of Sunna-cut

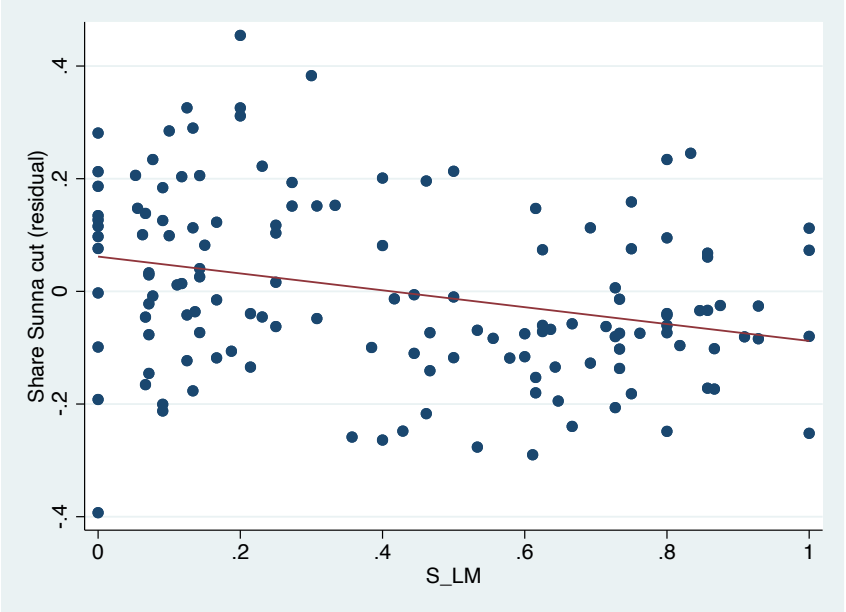
27 The proxies for s_{ji} and u_i are the same as above, only calculated at the community level for each one of our 141 communities.

Figure 10: Sunna cut and social sanctions

(a) Unconditional



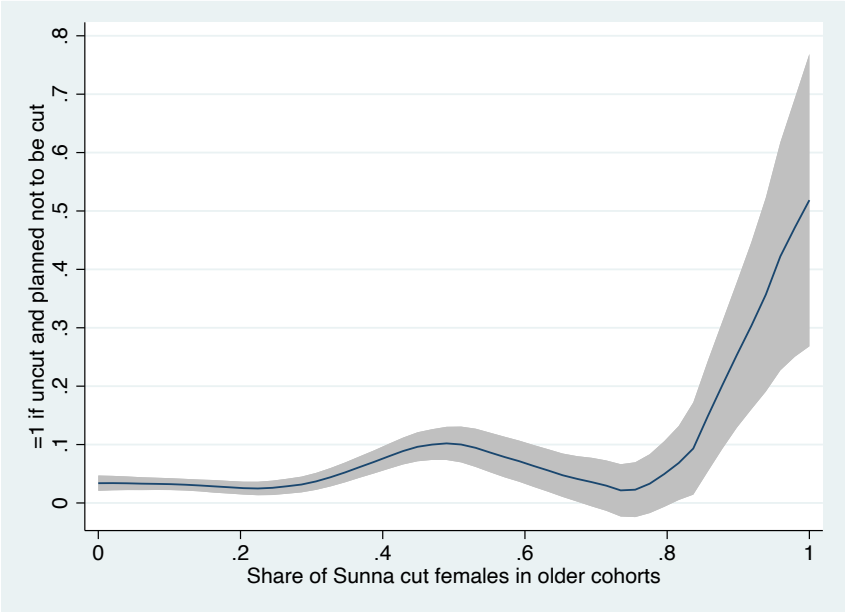
(b) Conditional on u_L and u_M



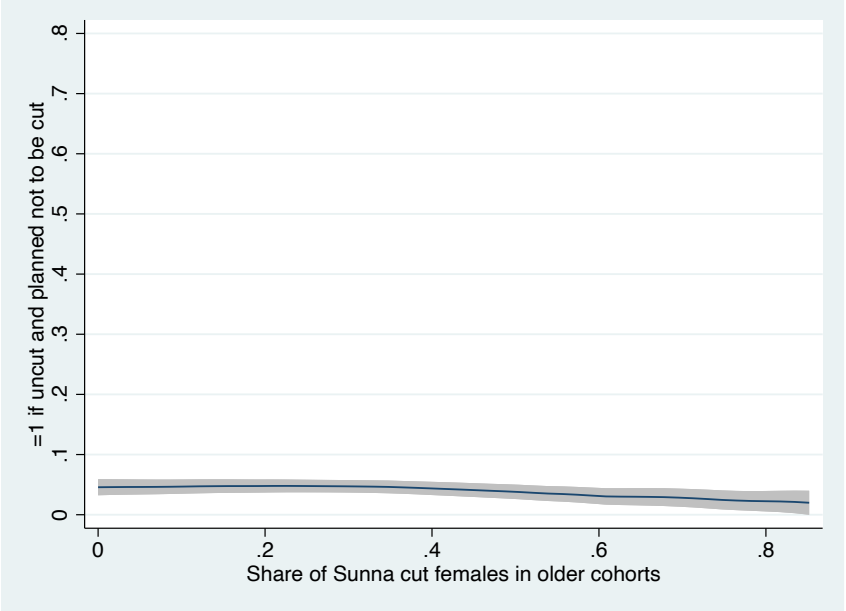
Notes: The variable on the x-axis is s_{LM} – the share of female respondents within the community who thought that parents who chose *Pharaonic* circumcision for their daughter would be unhappy if their son wanted to marry a girl with *Sunna* circumcision. In Panel A, the variable on the y-axis is the share of Sunna-cut girls within the community; the sample includes female respondents' daughters aged 0–18. In Panel B, the variable on the y-axis is the predicted residual from regressing share of Sunna-cut girls on u_L and u_M at the community level where u_L (u_M) is the share of female respondents who perceive any health complications or experienced (directly or indirectly via their daughters) any complications due to Pharaonic (Sunna) circumcision.

Figure 11: Threshold for stepping stone transition

(a) Low s_{MH}



(b) High s_{MH}



Notes: Sample includes female respondents' daughters aged younger than 18. Share of Sunna takes into account females aged 10+ years older than the girl within her community.

girls within her community (on the horizontal axis).[28] We do this separately for communities with below-median s_{MH} (top panel), and above-median s_{MH} (bottom panel). Figure 11a shows that in communities with lower s_{MH}, the likelihood that a girl is uncut begins to rise steeply when the share of Sunna-cut girls in older cohorts within the community reaches around 80 percent. This is in line with proposition 2, which predicts that in communities satisfying the stepping stone conditions, once a threshold q^* of people have already chosen Sunna, the switch to Uncut will be triggered. By contrast, in communities with relatively high s_{MH} we see no discernible relationship between a girl's likelihood to be Uncut and the share of Sunna cut girls (11b). In these communities, it is plausible that Sunna may have become an absorbing state.

To sum up, the predictions of our model regarding transition dynamics seem consistent with three patterns that emerge from the data. First, a transition has taken place from Pharaonic circumcision to Sunna. This occurred in an environment where, in the face of moderate social sanctions for moving from Pharaonic to Sunna, the health costs of the former greatly exceed those of the latter. Communities with lower social sanctions s_{LM} exhibit a higher share of Sunna-cut women.

Second, no transition from Sunna to Uncut has materialized in communities where the social sanctions s_{MH} are relatively high (the health costs associated to Sunna are perceived as uniformly low). In these communities, Sunna appears to be an absorbing state.

Third, in communities that have relatively low social sanctions for transitioning from Sunna to Uncut (s_{MH}) and where the share of Sunna-cut women has reached a certain threshold (in our data this is around 80 percent), a switch to Uncut seems to be underway. These are the communities where Sunna plays the role of a stepping stone, following the threshold dynamics predicted by our model.

4 Conclusions

We have proposed a model that allows us to analyze the intermediate-run dynamics of the evolution of social norms and to assess the conditions under which 'intermediate' versions of prevailing norms may act as a 'stepping stone' to transition to a superior norm in the long run, or else become an absorbing state. The key parameters governing the transition are the social sanctions imposed on those who abandon the prevailing option in favor of another, relative to the difference in intrinsic utilities of the two options. We have also characterized the waiting time associated with these transitions and their

28 The sample on the vertical axis includes daughters of female respondents. Since in our model decisions at a point in time depend on the shares of people cut until that point in time, each daughter in the graph is associated with the share of girls/women in her community who are 10 or more years older than herself (horizontal axis).

welfare consequences.

The evidence presented in the paper shows how, in the Somali context, the transition dynamics characterized by our model can help interpret the evolution of FGC over the past three decades. First, starting from virtually universal Pharaonic circumcision, a change in the social sanctions and/or in the perceived health costs associated with Pharaonic relative to Sunna triggered a dramatic shift in norms. Nowadays Sunna has replaced Pharaonic as the dominant trait for younger generations. Second, whether this new condition represents a stepping stone in the transition to not cutting or a new absorbing state depends on the characteristics of the community. While perceived health costs from Sunna are low everywhere, there is significant variation in attitudes towards people who choose Uncut over Sunna – which we interpret as a proxy for social sanctions. In communities with relatively low sanctions, the threshold property predicted by our model seems to hold in the data, leaving room for a transition to Uncut.

These patterns provide insights into potential policy responses. Since the key condition for transitioning from Sunna to Uncut requires that sanctions are low relatively to perceived utility gains, policymakers may work on two fronts. On the one hand, they may try to reduce sanctions associated with the decision not to cut, for example by changing the narrative around the value of circumcision in the marriage market. This is what NGOs like Tostan have been proposing, for example through 'public declarations' by community members that pledge not to cut their daughters and not to marry their sons to cut girls. On the other hand, policymakers may work on changing perceptions and knowledge of the health costs of FGC, e.g., through health information campaigns. Clearly, the two approaches are not exclusive and rather complement each other, as is evident from our theoretical predictions.

Finally, while our main example and empirical application concerns FGC in Somalia, our theoretical framework is general and can be applied to a variety of settings.

An interesting historical example is the case of dueling. In the United Kingdom, dueling died out suddenly in the mid-nineteenth century (Banks, 2008), while in France it endured up to the turn of the century (Hopton, 2007, 323). Part of the reason why this harmful norm persisted longer in France is that the form of dueling changed during this period, allowing for a less harmful alternative to replace the costlier norm. Swords became popular again, replacing pistols (Nye, 1993, 186), and, increasingly, the risks involved in dueling were made explicit.[29] These changes had the effect of drastically reducing fatality rates. It is estimated that more than 1/3 of duels in the early nineteenth century ended in the death of one of the duellists; during the second half of the century the number was

29 Duels were announced in advance as being *au premier sang*, to serious wounds, or, rarely, *à la mort* (Hopton, 2007, 79).

around 2 percent (Banks, 2012, 49–50).[30] These changes are consistent with a shift to a mild absorbing intermediate norm in our model.

A different, contemporary application of the stepping stone concept concerns child marriage. Most countries outlaw marriage before adult age (typically 18 years). This approach has failed to eradicate child marriage in developing countries, most notably in South Asia. Our stepping stone approach suggests that an intermediate step may be that of reducing the legal age for marrying to, say, 16 and seeing if this triggers a transition. Of course there could be a risk that this becomes the new norm. But this can only be assessed with data. An interesting avenue for future research is to explore these implications in countries that have indeed changed the legal age of marriage at different points in time (e.g., Bangladesh).

Finally, another setting in which our model could be applied is that of cigarette smoking. Recent years have seen the introduction of electronic cigarettes, and the share of consumers that have substituted tobacco with e-cigarettes has risen sharply. Will e-cigarettes ultimately lead people to quit smoking for good, or will they become the new norm for consumers who would have otherwise smoked tobacco? Again, as more data becomes available, it may be possible to explore transition patterns and give an answer to this question.

References

Abdalla, R. H. D. 1982. *Sisters in Affliction: Circumcision and Infibulation of Women in Africa.* Connecticut: Lawrence Hill and Co.

Adam, T., Bathija, H., Bishai, D., Bonnenfant, Y., Darwish, M., Huntington, D., & Johansen, E. 2010. Estimating the obstetric costs of female genital mutilation in six African countries. *Bulletin of the World Health Organization*, **88**, 281.

Ahmed, S. A., Hassan, S. M., & Maruf, H. 2018. *Somaliland Fatwa Forbids FGM.* https://www.voanews.com/africa/somaliland-fatwa-forbids-fgm.

Akerlof, George A. 1980. A theory of social custom, of which unemployment may be one consequence. *Quarterly Journal of Economics*, **94**(4), 749–775.

Akerlof, George A. 1997. Social distance and social decisions. *Econometrica*, **65**(5), 1005–1027.

30 The change was so significant that duellists became the subject of ridicule, to the extent that Mark Twain wrote in 1880: "Much as the modern French duel is ridiculed by certain smart people, it is in reality one of the most dangerous institutions of our day. Since it is always fought in the open air, the combatants are nearly sure to catch cold." Twain (1880) 1997, 38.

Akerlof, George A., & Kranton, Rachel E. 2000. Economics and identity. *Quarterly Journal of Economics*, **115**(3), 715–753.

Akerlof, George A, & Kranton, Rachel E. 2010. *Identity economics: How our identities shape our work, wages, and well-being*. Princeton: Princeton University Press.

Alesina, A., Giuliano, P., & Nunn, N. 2013. On the Origins of Gender Roles: Women and the Plough. *Quarterly Journal of Economics*, **128**(2), 469–530.

Ambrus, A., & Field, E. 2008. Early Marriage, Age of Menarche and Female Schooling Attainment in Bangladesh. *Journal of Political Economy*, **116**(5), 881–930.

Anderson, S. 2007. The economics of dowry and brideprice. *Journal of Economic Perspectives*, **21**(4), 151–174.

Ashraf, Nava, Bau, Natalie, Nunn, Nathan, & Voena, Alessandra. 2020. Bride price and female education. *Journal of Political Economy*, **128**(2), 591–641.

Banks, Stephen. 2008. Killing with courtesy: The English duelist, 1785–1845. *Journal of British Studies*, **47**(3), 528–558.

Banks, Stephen. 2012. *Duels and duelling*. Oxford: Shire.

Becker, A. 2018. *On the Economic Origins of Female Genital Cutting*. Working paper.

Bellemare, M. F., Novak, L., & Steinmetz, T. 2015. All in the Family: Explaining the Persistence of Female Genital Cutting in West Africa. *Journal of Development Economics*, **116**, 252–265.

Belloc, Marianna, & Bowles, Samuel. 2013. The persistence of inferior cultural-institutional conventions. *American Economic Review*, **103**(3), 93–98.

Benaim, Michel, & Weibull, Jörgen W. 2003. Deterministic approximation of stochastic evolution in games. *Econometrica*, **71**(3), 873–903.

Bicchieri, C., & Marini, A. 2016. *Female Genital Cutting: Fundamentals, Social Expectations and Change*. MPRA working paper No. 72927.

Bicchieri, Cristina. 2005. *The grammar of society: The nature and dynamics of social norms*. Cambridge: Cambridge University Press.

Blume, Lawrence E. 1993. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, **5**(3), 387–424.

Blume, Lawrence E. 1995. The statistical mechanics of best-response strategy revision. *Games and Economic Behavior*, **11**(2), 111–145.

Blume, Lawrence E, Brock, William A, Durlauf, Steven N, & Ioannides, Yannis M. 2011. Identification of social interactions. *Pages 853–964 of: Handbook of social economics*, vol. 1. Elsevier.

Bowles, Samuel. 2004. *Microeconomics: Behavior, institutions, and evolution*. Princeton: Princeton University Press.

Bowles, Samuel. 2006. Group competition, reproductive leveling, and the evolution of human altruism. *Science*, **314**.

Bowles, Samuel, & Choi, Jung-Kyoo. 2013. Coevolution of farming and private property during the early Holocene. *Proceedings of the National Academy of Sciences*, **110**(22), 8830–8835.

Bowles, Samuel, & Choi, Jung-Kyoo. 2019. The Neolithic agricultural revolution and the origin of private property. *Journal of Political Economy*, **127**(5).

Brock, William A, & Durlauf, Steven N. 2001. Discrete choice with social interactions. *Review of Economic Studies*, **68**(2), 235–260.

Camilotti, G. 2016. Interventions to Stop Female Genital Cutting and the Evolution of the Custom: Evidence on Age at Cutting in Senegal. *Journal of African Economies*, **25**(1), 133–158.

Carvalho, Jean-Paul. 2013. Veiling. *Quarterly Journal of Economics*, **128**(1), 337–370.

Corno, L., Hildebrandt, N., & Voena, A. 2017. *Age of marriage and the Direction of marriage payments*. NBER working paper No. 23604.

Corno, L., La Ferrara, E., & Voena, A. 2020. *Female Genital Cutting and the Slave Trade*. Working paper.

Diop, N. J., Faye, M.M., Moreau, A., Cabral, J., Benga, H., Cissé, F., Mané, B., Baumgarten, I., & Melching, M. 2004. *The TOSTAN program. Evaluation of a community based education program in Senegal*.

Efferson, C., Vogt, S., Elhadi, A., Ahmed, H., & Fehr, E. 2015. The economics of female genital cutting. *Science*, **349**(6255), 1446–1447.

Efferson, Charles, Vogt, Sonja, & Fehr, Ernst. 2020. The promise and the peril of using social influence to reverse harmful traditions. *Nature Human Behaviour*, **4**(1), 55–68.

El Dareer, A. 1982. *Woman, Why do you Weep?* London, U.K.: Zed Press.

Ellison, Glenn. 2000. Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *Review of Economic Studies*, **67**(1), 17–45.

Fernandez, R., & Fogli, A. 2009. Culture: An Empirical Investigation of Beliefs, Work, and Fertility. *American Economic Journal: Macroeconomics*, **1**(1), 146–177.

Fernandez, R., Fogli, A., & Olivetti, C. 2004. Mothers and Sons: Preference Formation and Female Labor Force Dynamics. *Quarterly Journal of Economics*, **119**(4), 1249–1299.

Foster, Dean P, & Young, H Peyton. 1990. Stochastic evolutionary game dynamics. *Theoretical Population Biology*, **38**(2), 219 – 232.

Goyal, Sanjeev. 2012. *Connections: an introduction to the economics of networks*. Princeton: Princeton University Press.

Gulesci, S., La Ferrara, E., Smerdon, D., & Sulaiman, M. 2020. *Changing Harmful Norms through Information and Coordination: Experimental Evidence from Somalia*. mimeo.

Hombrados, J. G., & Salgado, E. 2020. *Female Genital Cutting and Education: Evidence from a Legal Reform in Senegal*. mimeo.

Hopton, Richard. 2007. *Pistols at dawn: A history of duelling*. London: Piatkus.

Jackson, Matthew O. 2008. *Social and economic networks*. Princeton: Princeton University Press.

Kandori, Michihiro, Mailath, George J., & Rob, Rafael. 1993. Learning, mutation, and long run equilibria in games. *Econometrica*, **61**(1), 29–56.

Kudo, Yuya. 2019. *Political efforts to eliminate female genital cutting: Long-term effects on women's health and marriage in West Africa*. Working paper. Institute of Developing Economies (IDE-JETRO), Japan.

Mackie, G. 1996. Ending footbinding and infibulation: A convention account. *American Sociological Review*, **61**(6), 999–1017.

Manski, Charles F. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, **60**(3), 531–542.

McKelvey, Richard D, & Palfrey, Thomas R. 1995. Quantal response equilibria for normal form games. *Games and Economic Behavior*, **10**(1), 6–38.

Moffitt, Robert A. 2001. Policy interventions, low-level equilibria, and social interactions. *In:* Durlauf, Steven N, & Young, H Peyton (eds), *Social dynamics*. Cambridge, MA: MIT Press.

MOLSA. 2009. *National Policy For The Abandonment Of Female Genital Mutilation (FGM)*.

Monderer, Dov, & Shapley, Lloyd S. 1996. Potential games. *Games and Economic Behavior*, **14**(1), 124–143.

Newell-Jones, K. 2016. *Empowering communities to collectively abandon FGM/C in Somaliland: Baseline Research Report*.

Novak, Lindsey. 2020. Persistent norms and tipping points: The case of female genital cutting. *Journal of Economic Behavior & Organization*, **177**, 433 – 474.

Nye, Robert A. 1993. *Masculinity and male codes of honor in modern France*. Oxford: Oxford University Press.

Samuelson, Larry. 1997. *Evolutionary games and equilibrium selection*. Cambridge, MA: MIT Press.

Sandholm, William H. 2010. *Population games and evolutionary dynamics*. Cambridge, MA: MIT Press.

Schelling, Thomas C. 1978. *Micromotives and macrobehavior*. New York: W. W. Norton.

Shell-Duncan, Bettina, Wander, Katherine, Hernlund, Ylva, & Moreau, Amadou. 2011. Dynamics of change in the practice of female genital cutting in Senegambia: testing predictions of social convention theory. *Social science & medicine*, **73**(8), 1275–1283.

Slack, A. 1988. Female Circumcision: A Critical Appraisal. *Human Rights Quarterly*, **10**(4), 437–486.

Spisma, H. L., Chen, P. G., Ofori-Atta, A., Ilozumba, U. O., Karfo, K., & Bradley, E. H. 2012. Female genital cutting: current practices and beliefs in western Africa. *Bulletin of the World Health Organization*, **90**, 120–127F.

Twain, Mark. (1880) 1997. *A tramp abroad*. London: Penguin.

UNICEF. 2008. *Long-term evaluation of the Tostan programme in Senegal: Kolda, Thiès and Fatick regions*. www.unicef.org/evaldatabase/index_59605.html.

UNICEF. 2013. *Female genital mutilation/cutting: A statistical overview and exploration of the dynamics of change*. https://www.unicef.org/publications/index_69875.html.

UNICEF. 2016. *Female Genital Mutilation/Cutting: A Global Concern*. https://data.unicef.org/wp-content/uploads/2016/04/FGMC-2016-brochure_250.pdf.

Vega-Redondo, Fernando. 1996. *Evolution, games, and economic behaviour*. Oxford: Oxford University Press.

Vogt, S., Zaid, N.A.M., Ahmed, H.E.F., Fehr, E., & Efferson, C. 2016. Female Circumcision: A Critical Appraisal. *Nature*, **538**, 506–509.

Wagner, Natascha. 2015. Female Genital Cutting and Long-Term Health Consequences – Nationally Representative Estimates across 13 Countries. *The Journal of Development Studies*, **51**(3), 226–246.

Weibull, Jörgen W. 1995. *Evolutionary game theory.* Cambridge, MA: MIT Press.

Williams, David. 1991. *Probability with martingales.* Cambridge: Cambridge University Press.

Yoder, P. S., Wang, S., & Johansen, E. 2013. Estimates of female genital mutilation/cutting in 27 African countries and Yemen. *Studies in Family Planning*, **44**(2), 189–204.

Young, H Peyton. 1993. The evolution of conventions. *Econometrica*, **61**(1), 57–84.

Young, H Peyton. 1998. *Individual strategy and social structure.* Princeton: Princeton University Press.

Young, H Peyton. 2015. The evolution of social norms. *Annual Review of Economics*, **7**(1), 359–387.

A Proofs

A.1 Proof of claim 1

We need to show that for any $p \in \tilde{\Delta}$ and $i, j \in A$,

$$v_j(p + e^{ij}) - v_i(p) = \rho(p + e^{ij}) - \rho(p); \tag{47}$$

that is, the change in utility from switching from i to j is the same as the change in potential.

First, we have

$$\begin{aligned} v_j(p + e^{ij}) - v_i(p) &= u_j - \sum_{k \in A} s_{jk}(p_k + e^{ij}_k) - u_i + \sum_{k \in A} s_{ik} p_k && (48)\\ &= u_j - u_i - \sum_{k \in A} (s_{jk} - s_{ik}) p_k - \sum_{k \in A} s_{jk} e^{ij}_k && (49)\\ &= u_j - u_i - \sum_{k \in A} (s_{jk} - s_{ik}) p_k + \delta s_{ij}. && (50) \end{aligned}$$

Second, we have

$$\begin{aligned} \rho(p + e^{ij}) - \rho(p) &= m \sum_{k \in A} (p_k + e^{ij}_k) u_k - \frac{m}{2} \sum_{k \in A} \sum_{l \in A} (p_k + e^{ij}_k)(p_j + e^{ij}_l) s_{kl} \\ &\quad - m \sum_{k \in A} p_k u_k + \frac{m}{2} \sum_{k \in A} \sum_{l \in A} p_k p_l s_{kl} && (51)\\ &= m \sum_{k \in A} e^{ij}_k u_k - \frac{m}{2} \sum_{k \in A} \sum_{l \in A} (p_k e^{ij}_l + p_l e^{ij}_k + e^{ij}_k e^{ij}_l) s_{kl} && (52)\\ &= u_j - u_i - m \sum_{k \in A} p_k \sum_{l \in A} e^{ij}_l s_{kl} - \frac{m}{2} \sum_{k \in A} e^{ij}_k \sum_{l \in A} e^{ij}_l s_{kl} && (53)\\ &= u_j - u_i - m \sum_{k \in A} p_k \delta (s_{jk} - s_{ik}) - \frac{m}{2} \sum_{k \in A} e^{ij}_k \delta (s_{jk} - s_{ik}) && (54)\\ &= u_j - u_i - \sum_{k \in A} p_k (s_{jk} - s_{ik}) - \frac{m}{2} (-2\delta^2 s_{ij}) && (55)\\ &= u_j - u_i - \sum_{k \in A} p_k (s_{jk} - s_{ik}) + \delta s_{ij}. && (56) \end{aligned}$$

Hence $\mathscr{G}$ is a potential game with potential function ρ.

A.2 Proof of theorem 2

Let $q^* \in (0, 1)$ be as defined in equation (19); that is, q^* is the proportion of M-players at which agents start to switch to H.

Define

$$\lambda = u_H - u_L + s_{LH} \tag{57}$$

Note that λ is the maximum possible absolute change in payoff from switching from one action to another. Recall that γ is the smallest nonzero increase in payoff from switching actions. Given $\nu \in (0, 1-q)$, define

$$\varepsilon = \frac{\nu\gamma}{1+\lambda+\nu\gamma}. \tag{58}$$

Choose α_ν such that for all $p \in \tilde{\Delta}$ and $\alpha \in [0, \alpha_\nu]$, $\sigma(p, \alpha)$ puts probability at least $1-\varepsilon$ on the set of best replies.

Consider the stochastic process $(p^t)_{t\geq 0}$ that starts in state p^L at time $t = 0$. For any $t \geq 0$, define the stochastic increment

$$D(p^t) = \rho(p^{t'}) - \rho(p^t), \tag{59}$$

where $t' > t$ is the first time after t when some agent updates.

We claim that for every $p^t \in \tilde{\Delta}$ such that $p^t_H < 1-\nu$,

$$\mathbf{E}[D(p^t)] \geq \varepsilon. \tag{60}$$

That is, the expected change in potential is bounded from below by ε.

To establish the claim, let $r(p)$ be the proportion of players that are currently not playing best replies in state p. Let

$$r^* = \min_{\{p\in\tilde{\Delta}:\, p_H<1-\nu\}} r(p). \tag{61}$$

Suppose the minimum is attained at p^*. Since M is a stepping stone, L-players are never at a best reply. If $p^*_M + p^*_H > q^*$, M-players prefer to switch to H, and hence $r^* \geq p^*_M + p^*_L = 1 - p^*_H > \nu$. If $p^*_M + p^*_H \leq q^*$ then $r^* \geq p^*_L > 1 - q^* > \nu$. Thus, in any case, $r^* > \nu$.

Consider any $p^t \in \tilde{\Delta}$ such that $p^t_H < 1-\nu$. At the next updating, the probability is at least $r^* > \nu$ that the updating agent is not currently playing a best reply. Such a player will increase her payoff by at least $(1-\varepsilon)\gamma - \varepsilon\lambda$ in expectation. The probability is less than $1-\nu$ that the agent is already playing a best reply, in which case the agent's payoff decreases by at most $\varepsilon\lambda$ in expectation. The expected change in potential is equal

to the expected change in payoffs. Hence for any $t \geq 0$,

$$\begin{aligned}
\mathbf{E}[D(p^t)] &\geq \nu((1-\varepsilon)\gamma - \varepsilon\lambda) - (1-\nu)\varepsilon\lambda && (62)\\
&= \gamma(1-\varepsilon)\nu - \lambda\varepsilon && (63)\\
&= \varepsilon, && (64)
\end{aligned}$$

the latter by equation (58). This establishes the claim in equation (60).

The expected number of updates until the process first reaches a state p satisfying $p_H \geq 1-\nu$ is bounded by the total change in potential divided by ε. In expectation there are m updates per time period. Therefore

$$\begin{aligned}
T_\nu &\leq \frac{1}{m}\frac{\rho(p^H) - \rho(p^L)}{\varepsilon} && (65)\\
&\leq \frac{(u_H - u_L)(1+\lambda+\nu\gamma)}{\nu\gamma} && (66)\\
&\approx O(1/\nu). && (67)
\end{aligned}$$

A.3 Proof of theorem 3

The argument will be more transparent if we consider the associated discrete-time process in which one agent is drawn uniformly at random to update each period. This process is m times slower than the original process and amounts to looking at the embedded chain of updates in the original process.

Note that in any given state, potential can only weakly increase. Recall that γ is the smallest nonzero increase in payoff from switching from one action to another. We begin with the following lemma:

Lemma 1. In any inhomogenous state $p \in \tilde{\Delta}$, the expected change in potential is at least $\delta\gamma$.

Proof. First, we show that in any inhomogenous state $p \in \tilde{\Delta}$, there is at least one agent who can increase her payoff by at least γ. Since γ is the smallest nonzero increase in payoff, it is sufficient to show that there is at least one agent who can increase her payoff. Since p is inhomogenous, let $i, j \in A$ be two actions played by a positive proportion of agents. Suppose to the contrary that neither i-players nor j-players can increase their payoff. Then $u_i - s_{ij}p_j \geq u_j - s_{ij}(p_i - \delta)$ and $u_j - s_{ij}p_i \geq u_i - s_{ij}(p_j - \delta)$. But this implies $u_i - s_{ij}p_j \geq u_i - s_{ij}(p_j - \delta)$, which is a contradiction. Hence, there exists at least one agent who can increase her payoff by at least γ. The probability of selecting that agent is δ, and any other agent will either increase the potential or keep it constant. Therefore the expected change in potential is at least $\delta\gamma$, as required. □

Thus if p^t is inhomogenous, the expected change in ρ viewed at t satisfies $\mathbf{E}_t[\rho(p^{t+1}) -$

$\rho(p^t)] \geq \delta\gamma$. But since potential will not decrease in the following period, we also have $\mathbf{E}_t[\rho(p^{t+2}) - \rho(p^t)] \geq \delta\gamma$.

Now suppose that p^t is homogenous but is not strictly stable. If it is unstable, then some agent will move at t. The expected increase in potential is at least γ. Hence $\mathbf{E}_t[\rho(p^{t+2}) - \rho(p^t)] \geq \gamma$. If p^t is homogenous and weakly stable, agents are indifferent between two or more options. Then with probability at least $1/2$, the updating agent switches action. Although the potential does not increase, p^{t+1} is now inhomogenous, so from $t+1$ to $t+2$ the expected increase in potential is at least $\delta\gamma$ by lemma 1. Therefore $\mathbf{E}_t[\rho(p^{t+2}) - \rho(p^t)] \geq \frac{\delta\gamma}{2}$.

It follows that for any $p^t \in \tilde{\Delta}$ that is not strictly stable,

$$\mathbf{E}_t[\rho(p^{t+2}) - \rho(p^t)] \geq \frac{\delta\gamma}{2}. \tag{68}$$

Starting from $p^0 \in \tilde{\Delta}$, let the random variable T be the first even time such that the process is in a strictly stable homogenous state.

Define the function

$$h(t) = \rho(p^t) - t\frac{\delta\gamma}{4}. \tag{69}$$

By inequality (68),

$$\mathbf{E}_t[h(t+2) - h(t)] \geq 0. \tag{70}$$

Hence $h(t)$ is a submartingale on the even periods. The value of the stopping time T is finite with probability one, hence by Doob's optional-stopping theorem (e.g., Williams, 1991, section 10.10)

$$\mathbf{E}[h(T)] \geq h(0) \tag{71}$$

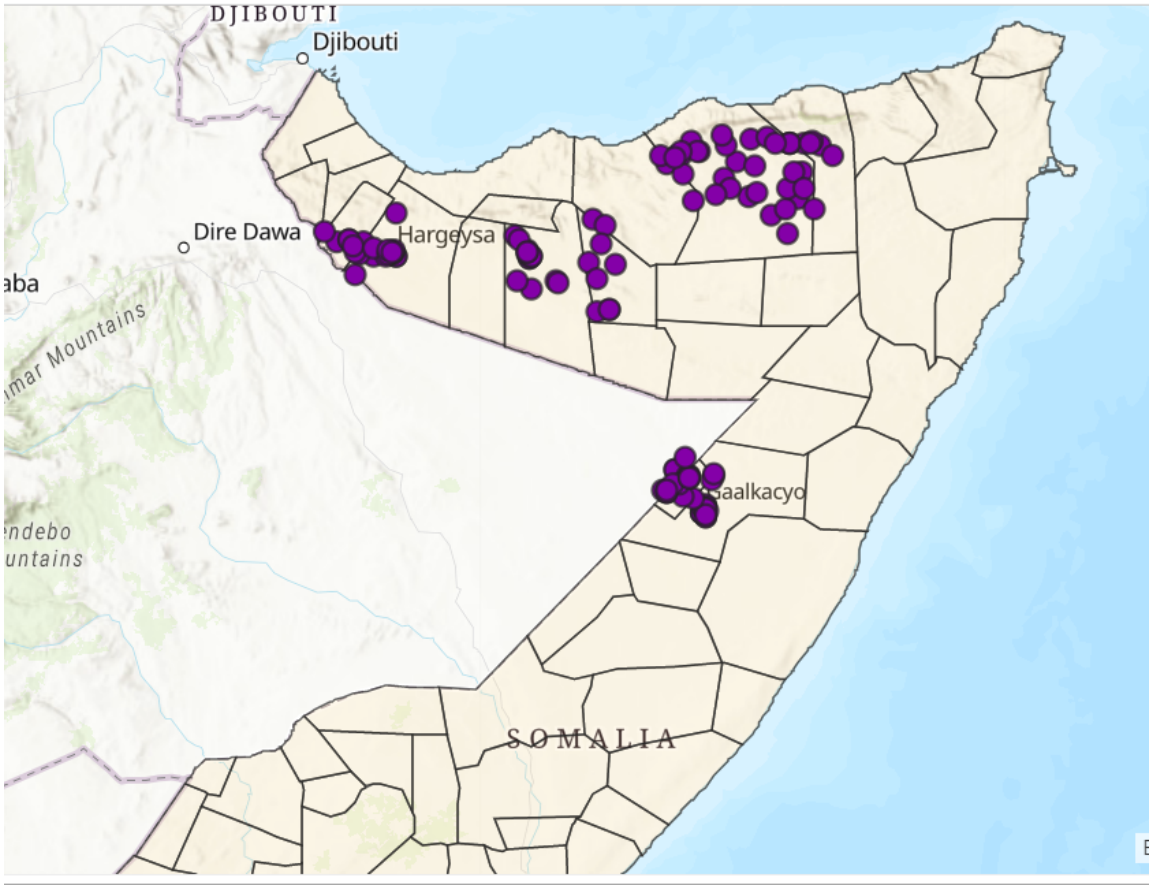
$$\implies \mathbf{E}[T] \leq 4\frac{\rho^* - \rho(p^0)}{\delta\gamma}, \tag{72}$$

where ρ^* is the maximum potential across all states.

The expected waiting time in the continuous process is $\mathbf{E}[T]/m$. This concludes the proof of theorem 3.

B Additional Tables and Figures

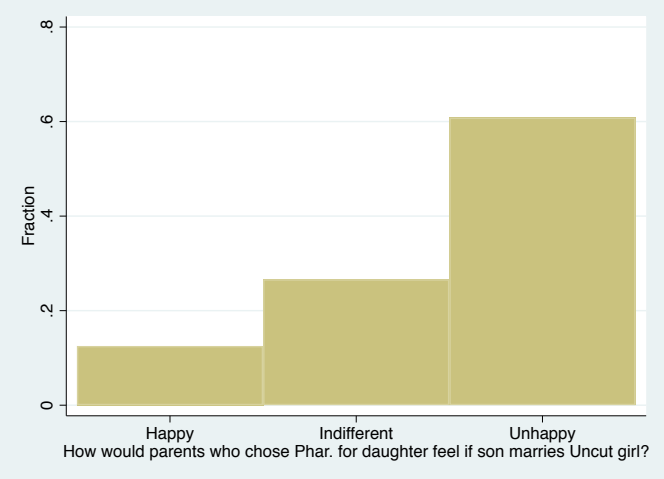
Figure B.1: Location of the Study



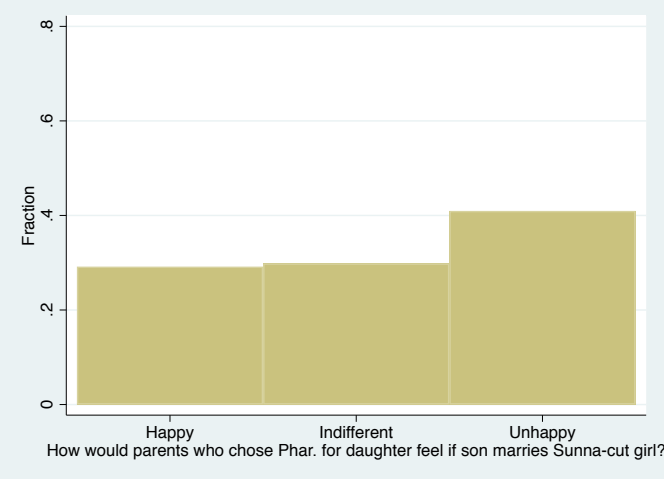
Notes: The map shows the location of communities included in the study, along with district boundaries of Somalia.

Figure B.2: s_{LH} , s_{LM} and s_{MH}

(a) Pharaonic v.s. Uncut s_{LH}



(b) Pharaonic v.s. Sunna s_{LM}



(c) Sunna v.s. Uncut s_{MH}

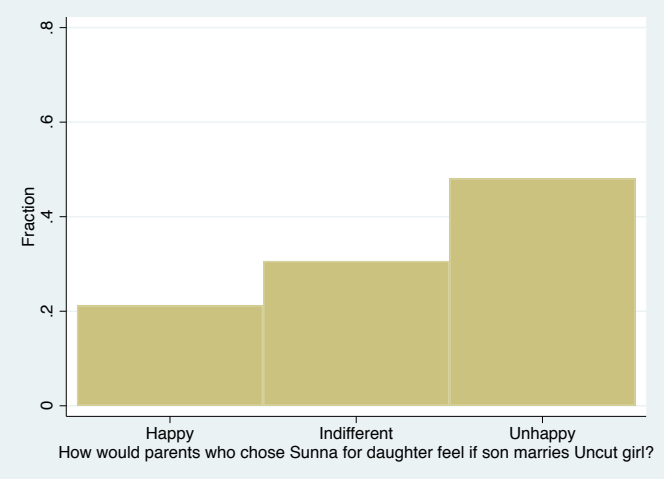
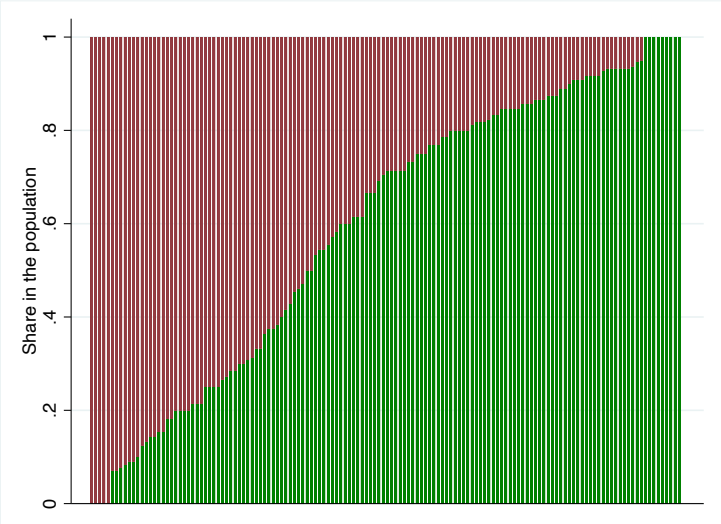
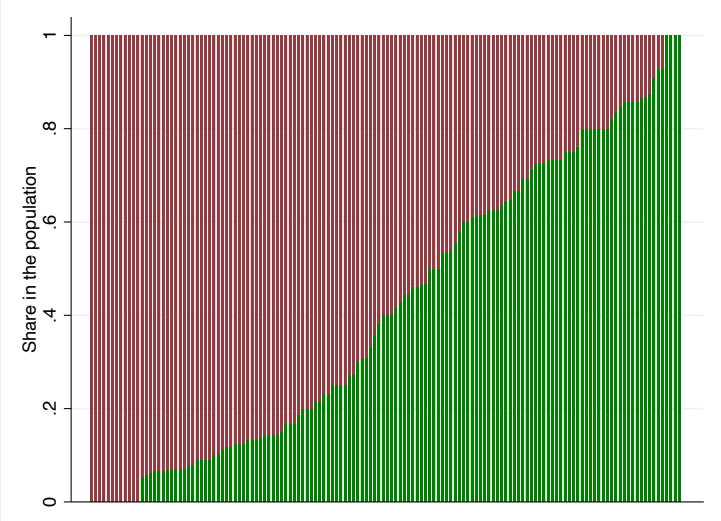


Figure B.3: Distribution of social sanction factors across communities

(a) S_{LH}



(b) S_{LM}



(c) S_{MH}