Department of Economics
Discussion Paper Series

The dynamics of costly social norms

Sam Jindani and Peyton Young

Number 883
First Published: October, 2019
Revised: January, 2020

Department of Economics
Manor Road Building, Manor Road
Oxford, OX1 3UQ

ISSN 1471-0498

The dynamics of costly social norms[1]

Sam Jindani[2] and Peyton Young[3]

30 January 2020

Social norms that are harmful for individuals can remain in place for a long time because deviations are punished by members of the community. Examples include female genital cutting, foot binding, and duelling. These and other costly norms are rarely 'all-or-nothing', but take many alternative forms. We develop a general theory of norm dynamics that characterises the intermediate-run behaviour of such systems. The theory identifies conditions under which norms tend to collapse suddenly on the one hand, or transition via lower-cost variants on the other. It also identifies when it is more efficient for policy interventions to target full abandonment and when it may be preferable to nudge the community toward an intermediate norm.

1 Introduction

Early one morning in 1804, Alexander Hamilton and Aaron Burr left Manhattan in separate boats and rowed to New Jersey. At the time, duelling was outlawed in many states and New Jersey was a common duelling spot for New Yorkers because the ban was less strictly enforced. Pistols were hidden in a bag, allowing the rowers to declare under oath that they had not seen any pistols; they also stood with their backs to the duellists. Although Hamilton

1 The authors thank Sam Bowles, Jean-Paul Carvalho, Charles Efferson, and Yuya Kudo for constructive comments on an earlier draft. Sam Jindani gratefully acknowledges financial support from the Economic and Social Research Council (award number ES/J500112/1) and the University of Oxford Department of Economics.
2 Department of Economics, University of Oxford.
Corresponding author. Email: sam.jindani@gmail.com.
3 London School of Economics and Nuffield College, University of Oxford.
Email: peyton.young@economics.ox.ac.uk.

was morally opposed to duelling and felt that it would be irresponsible towards his wife and children, he decided to go ahead with the duel anyway, because of the threat of public discredit. On the day of the duel, Hamilton shot first and missed, perhaps intentionally; Burr hit and mortally wounded Hamilton, who died the following day.

Duelling, along with other costly behaviours such as female genital cutting (FGC) and foot binding, is an example of a costly social norm. A social norm is a behaviour or set of behaviours that agents engage in because it is prescribed by society; a costly social norm is one that is harmful to the agents following it. In the case of duelling, the cost could be very high. It is estimated that more than a third of duels in the early nineteenth century in France ended in the death of one of the duellists.[4] Moreover duelling was generally illegal and decried by institutions such as the Church. Yet people, for the most part, felt compelled to follow the norm. Contemporary accounts make it clear that social pressure was a significant motivation: people were concerned about loss of esteem from their peers if they failed to challenge someone to a duel when one was expected, or failed to accept a duel, or in any way breached the norms surrounding duelling, and this concern was sufficient to drive them to risk their lives.[5] The most striking illustrations of this are the numerous instances of people duelling despite being opposed to the practice, including Hamilton.

Social norms often govern a set of behaviours rather than a single behaviour, and they can take many intermediate forms that are more or less costly. In the case of duelling, convention regulated when it was appropriate to challenge someone to a duel, how the duel should be organised, what type of weapons should be used, how many paces should separate the duellists, and so forth. Moreover, these expectations differed from place to place and changed over time. Until the mid-sixteenth century duellists in France and England predominantly used slashing broadswords and protected themselves with bucklers. Eventually, these were replaced by the use

4 Chesnais 1981, 103.

5 See in particular R. Akerlof (2017), who highlights the importance of social esteem in inducing conformity.

of rapiers, which were far more deadly since they could inflict damage to internal organs whereas broadswords only caused superficial flesh wounds.[6] When pistols became popular in the late eighteenth century the distance separating the duellists, which heavily influenced both accuracy and lethality, was the subject of strong restrictions. The conventional distance was ten or twelve paces, which might be reduced if the duel was particularly serious, or increased if the converse was true.[7] Duellists who duelled at smaller distances were considered reckless and were often the subjects of social opprobrium. It is interesting to note that although duellists who were considered to have followed conventions were rarely the subject of legal pursuits, this leniency never applied to those who had breached them.[8] The comte de Chatauvillard, the author of a popular duelling manual, said that 'honour might require us to risk our lives, but not to play with them'.[9]

Understanding duelling as a complex social norm with many possible variants is essential to explaining its evolution over time. In the United Kingdom, duelling died out suddenly in the mid-nineteenth century. Figure 1 shows that this happened within a few years in the 1840s. In contrast, in France duelling endured up to the turn of the century;[10] importantly, however, the *form* of duelling changed during this period. Swords became popular again,[11] and, increasingly, the risks involved in duelling were made explicit: duels were announced in advance as being *au premier sang*, to serious wounds, or, rarely, *à la mort*.[12] These changes had the effect of drastically reducing fatality rates. As mentioned above, it is estimated that more than a third 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 would be around two percent.[13] The change was so significant that duellists became

6 Stone 1965, 242–44.
7 Hopton 2007, 81–83.
8 Hopton 2007, 30–33.
9 Cited in Nye 1993, 142.
10 Hopton 2007, 323.
11 Nye 1993, 186.
12 Hopton 2007, 79.
13 Banks 2012, 49–50.

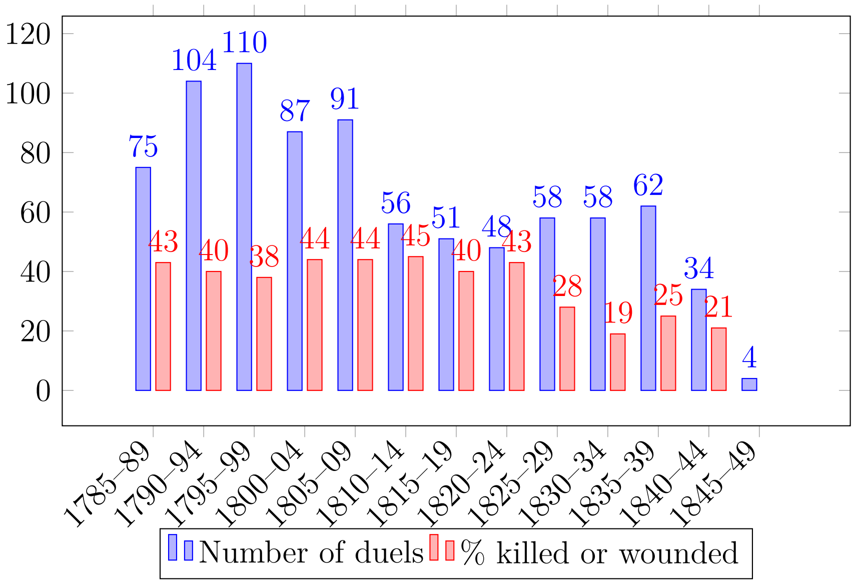


Figure 1: Data on duelling in the UK. Source: Banks 2008.

the subject of ridicule, to the extent that Mark Twain was moved to write in 1880 after visiting France:

> 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.[14]

Like duelling, the norms of FGC and foot binding endured for hundreds of years despite efforts to abolish them. Foot binding was prevalent for over a millennium in China, from the tenth century to the turn of the twentieth century. The process of foot binding took years, starting between the ages of six to eight, before the arches of the feet were fully developed. The feet were broken, folded on themselves, and bound tightly with bandages to mould them to the desired shape: a compressed arched foot around ten centimetres in length.[15] The bandages were redone and tightened regularly for the next six to ten years. The whole procedure was extremely painful, and frequently

14 Twain (1880) 1997, 38.
15 Mackie 1996, p. 1000.

led to infection. It is thought that as many as ten percent of girls subjected to the process did not survive. Those who did were usually effectively hobbled. There was considerable social pressure to conform to the norm: foot binding was considered attractive, a sign of wealth and status; parents considered it necessary for a proper marriage and feared that they would be unable to find a match for their daughters if they did not have their feet bound.[16]

There was also variation in the norm across time and place. Stories suggest that foot binding originated from a court dancer wrapping her feet in silk to perform, which others sought to imitate (Mackie 1996, p. 1001). As time went on, the practice became more extreme and its association with dancing was eventually lost. Archaeological evidence supports this view: remains from the 13th century show women with bound feet, but not as small as the later ten-centimetre ideal.[17] There were also distinct regional variants of the norm, for instance in Sichuan, where people practiced a less severe form of foot binding that did not involve distorting the heel.[18]

Foot binding died out abruptly at the turn of the twentieth century.[19] Governments attempted to outlaw the practice in 1665 and 1847, but were unsuccessful. Throughout the nineteenth century, foot binding fell increasingly out of favour in public opinion and seen as incompatible with modern Chinese society. It was condemned by Chinese intellectuals and Western missionaries, and by the end of the nineteenth century, a number of anti-footbinding societies had formed. By all accounts, the demise of foot binding was fast. In 1908, the majority of public opinion was opposed to foot binding, and by the 1920s, it had all but disappeared in the new generation.

FGC endures to this day, principally in and around the Sahel. The extent of the practice varies considerably across different countries and ethnic groups; altogether, it is estimated that 100–200 million women have undergone some form of FGC.[20] The United Nations distinguishes three main

16 Mackie 1996, p. 1002.
17 Ko 2007, p. 187-91.
18 Gates 2015, p. 7.
19 Mackie 1996, p. 1001.
20 For different perspectives on FGC practices and what sustains them, see Mackie 1996, Shell-Duncan *et al.* 2011, Efferson *et al.* 2015, and Kudo 2019.

types: type I refers to partial or total removal of the clitoris and/or the prepuce; type II also includes removal of the labia minora; type III, also known as infibulation, is the severest, and involves cutting and stitching the sides of the labia together.[21] In its most severe form, FGC is extremely painful and can have far-reaching health consequences. The operation itself can lead to urinary infection, septicaemia, and tetanus amongst other complications. After healing, it can still cause frequent pain, difficulty in urinating, difficulties in childbirth and painful sexual intercourse.[22] For type I and type II, long-term health consequences appear to be more limited.[23]

As with duelling and foot binding, social acceptance appears to be a key factor holding the norm in place: In surveys, the most commonly cited benefit of FGC is social acceptance.[24] Circumcision admits a girl to the social network of older circumcised women in the community, whereas non-circumcision may lead to exclusion and harassment. Marriageability is also cited as a factor but it appears to be less important than social acceptance by other women (Shell-Duncan *et al.* 2011). In many countries where it is practiced, the majority of women think it should be abolished but conform to it nonetheless.[25]

The prevalence and type of FGC vary considerably, both between countries and within countries. Variations in the prevalence across regions depend on ethnic composition as well as community of origin, suggesting that FGC is practiced and enforced at the community level (see in particular Kudo 2019 and United Nations Children's Fund 2013, p. 31). In some countries, there has been a decline in the prevalence of FGC over the past three decades, particularly the most severe forms of the practice.[26] For instance in Djibouti, 83 percent of women aged 45 to 49 report having had type III FGC, compared to only 42 percent of girls aged 15 to 19.[27]

21 United Nations Children's Fund 2013, p. 7.
22 Mackie 1996, p. 1003.
23 Wagner 2015; Efferson *et al.* 2015; Kudo 2019.
24 United Nations Children's Fund 2013, p. 67.
25 United Nations Children's Fund 2013, p. 54.
26 United Nations Children's Fund 2013, p. 94 and ff.
27 United Nations Children's Fund 2013, p. 111.

A growing trend in some countries is the medicalisation of FGC, whereby the procedure is increasingly carried out by healthcare professionals in healthcare facilities rather than by traditional cutters at home. In Egypt in particular, the percentage of girls cut by traditional practitioners decreased from 42 percent in 1995 to 22 percent in 2008. FGC operations were banned in state hospitals in 1995, but a loophole allowed the practice to continue; a stricter ban was introduced in 2008, although studies suggests this was not effective. The United Nations argues that the trend towards medicalisation reflects the desire of parents to mitigate the harm caused by FGC to their daughters, in an environment where social pressure in favour of the practice endures.[28]

The present paper develops a general theory of the dynamics of costly norms such as duelling, foot binding, and FGC. The theory identifies conditions under which norms tend to collapse suddenly on the one hand (as was the case for duelling in the United Kingdom and foot binding in China), or transition via lower-cost variants on the other. In the former case, we also show that norms can sometimes transition to a very low-cost form instead of collapsing completely (as was the case for duelling in France). The conditions also have implications for policy intervention: in the former case it is more efficient for policy interventions to target full abandonment, whereas in the latter case it may be preferable to nudge the community toward an intermediate norm.[29]

The model extends Akerlof's model of social distance (1997) to a dynamic setting: agents choose a package of behaviours with an associated *individual cost*, and suffer a *social cost* (loss of esteem, ostracism) if their chosen behaviour is different from other agents' behaviours. The further the behaviour is from that of other agents, the more severe the sanction. We show that the shape of the social cost function is a key determinant of the norm dynamics. Throughout the paper we distinguish between two cases related to concavity

28 United Nations Children's Fund 2013, pp. 107-10.

29 According to recent research, FGC appears to be eroding in some countries, but it is too early to tell whether the transition is mainly to lower-cost variants (including symbolic 'nicking') or complete abandonment (Shell-Duncan *et al.* 2011; Efferson *et al.* 2015).

and convexity of the sanctions function, respectively.

We first analyse the best-response process without stochastic perturbations, whereby agents update their actions periodically and pick best responses to the current state. We show that, with probability one, the process will converge to a *norm*, meaning a homogenous, stable state. Using martingale theory we also derive an explicit upper bound on the expected waiting time to reach a norm.

The best-response model also allows us to look at the impact of exogenous societal changes. In the case of duelling, scholars have argued that the rapid modernisation of the Industrial Revolution led to a shift in social attitudes towards honour and contributed to the decline of duelling (Banks 2008). Similarly, the decline in traditional patriarchal values, expanded labour opportunities for women, and the influence of Western imperialism have all been put forward as factors in the demise of foot binding. We model these societal changes as changes in the strength of social sanctions relative to the cost of the norm that causes the current norm to become unstable. We show that depending on the shape of the sanctions function, the norm will either abruptly die out or to shift to the closest stable norm.

Finally, we turn to the perturbed best-response process, which allows us to study the impact of bottom-up shocks as well as top-down policy interventions. We assume agents' actions are subject to idiosyncratic preference shocks or policy pressure and we characterise the intermediate-run behaviour of the resulting systems. When the strength of social sanctions relative to the cost of the norm is decreasing, we show that the process will tend to by-pass intermediate norms; instead, norms tend to die out completely. They can also collapse to a low-cost norm, if its cost is sufficiently small. Moreover, once a low-cost norm becomes established it is difficult to displace, so the existence of low-cost intermediate norms may significantly retard eventual abandonment. In contrast, when the strength of social sanctions relative to the cost of the norm is increasing, we show that the process will tend to gradually decline via intermediate norms. For policy interventions, we show that in the concave case it is more efficient to target full abandonment, whereas in the convex case it is more efficient to target a shift to an intermediate norm. We provide some *a priori* reasons why concavity might be plausible in some settings (*e.g.* duelling and foot binding) and convexity in others (*e.g.* FGC), but ultimately we note that the shape of the sanctions function is likely to be an empirical question.

The plan of the paper is as follows. In section 2, we give an overview of related literature. In section 3 we introduce the basic model, which treats norms as packages of behaviours that have individual costs as well as social costs from norm violation. In section 5 we characterise the dynamics of the process when society begins in an out-of-equilibrium situation and players best-respond to the behaviour of others. In section 6 we look at the impact of exogenous changes in preferences. Finally, in section 7 we consider bottom-up shocks and top-down interventions. Section 8 concludes.

2 Related literature

There is a wide-ranging literature on the dynamics of norms that spans economics, sociology, philosophy, and political science.[30] Schelling (1978) was one of the first to introduce a game-theoretic explanation of norms in terms of coordination equilibria, and to analyse norm shifts using the concept of tipping points. Mackie (1996) extended this framework to analyse the persistence of costly norms, such as FGC and foot binding, and argued that targeted intervention at the level of subgroups can be particularly effective in overturning such norms. Another key contributor to the economic analysis of norms is G. A. Akerlof (1980, 1997), who argued that social interactions and the attendant pressure to adhere to established norms of behaviour can lead to inefficient equilibrium traps from which it is difficult to escape.

More recently, evolutionary game theory has emerged as a useful tool for analysing norm dynamics (Young 1993, 1998; Kandori, Mailath and Rob 1993; Bowles 2006).[31] This methodology allows one to analyse how sys-

30 See among others Schelling 1978; G. A. Akerlof 1980; Axelrod 1986; Skyrms 1996, 2003; Mackie 1996; Young 1998, 2015; Blume and Durlauf 2001; Brock and Durlauf 2001; Bowles 2004; Bicchieri 2005; and Acemoglu and Jackson 2015.

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

tems of interacting agents converge to normative behaviours from out-of-equilibrium conditions, and how shifts between norms are induced by exogenous or endogenous variations in payoffs. Applications to specific cases include contractual norms in agriculture (Young and Burke 2001), the evolution of property rights (Bowles and Choi 2013, 2019), norms of medical practice (Burke, Fournier and Prasad 2010), inferior institutions and forms of organisation (Belloc and Bowles 2013), and signalling norms such as veiling among Muslim women (Carvalho 2013).

The present paper also adopts this framework, but unlike much of the literature we focus on intermediate instead of long-run behaviour. We are particularly interested in norms that are costly in the sense that individuals may suffer harm from following the norm, but they persist in doing so because of the negative sanctions that are applied to deviants. We show how evolutionary models can be used to explain the intermediate-run dynamics of norm shifts in these settings that are consistent with actual behaviour in cases where norm shifts have been documented.

3 A model of costly norms

There is a finite set of feasible actions $A = \{a_0, a_1, \ldots, a_n\}$, where $a_0 = 0$ is the *null* action. This set of actions captures behaviours along multiple dimensions. In the case of duelling, agents might be more or less prone to challenging other people to duel, and more or less prone to accepting challenges; when they fight duels, they could tend to choose one weapon over another; if they fight with pistols, they could push for a small or large number of paces. In the case of foot binding, decisions are made by the family about how many of their daughters to subject to the procedure, or how small to make the feet. In the case of FGC, agents might decide which type of procedure to perform and whether to have it performed in a traditional setting or not.

In adopting an action, an agent is faced with two types of cost: individual and social. The *individual cost* of an action is its inherent disutility to the agent. Let c_i be the cost of the ith action. We will assume that $c_0 = 0$ and that costs are distinct and indexed so that

$$c_0 < c_1 < \cdots < c_n. \tag{1}$$

The *social cost* of an action represents the pressure by others to conform. They punish agents who adopt actions that differ from their own, either by expressing disapproval, breaking off relations, or perhaps threatening them with violence. The severity of the punishment will depend on the perceived deviation from the normative behaviour.

We assume that if one agent chooses action a while another agent chooses action $b \neq a$, the degree of disapproval or sanction that each imposes on the other is proportional to the perceived distance between two the actions, as measured by the difference in their costs. Specifically, we will assume that there is a monotone nondecreasing, non-negative function $s : [0, c_n] \to \mathbb{R}_+$ and a real number $\lambda > 0$ such that $s(0) = 0$ and for all actions a, b, the sanction is equal to

$$\lambda s(|c(a) - c(b)|). \tag{2}$$

In our analysis, it will be convenient to distinguish the *strength* of sanctions, represented by λ, from the *shape* of the sanctions function, represented by s. In particular, we will be interested in the behaviour of the process as λ varies. Shifts in social attitudes towards honour, or the role of women in society, for instance, can be captured by a decrease in λ, meaning that social costs become less important relative to individual costs.

The population consists of m agents, $m \geq 2$. At a given point in time, the *state* z specifies the number of agents playing each action:

$$(z_0, z_1, \ldots, z_n) \in \mathbb{Z}^n, \quad \sum_{i=0}^{n} z_i = m. \tag{3}$$

The *payoff* of an agent playing action a_i in state z is

$$u(a_i, z) = -c_i - \lambda \sum_j z_j \cdot s(|c_i - c_j|). \tag{4}$$

Note that λ reflects the strength of social sanctions exerted between individuals, which depends on how frequently they interact. In a small group, such as a village, λ may be much larger than in a larger more diffuse group. Indeed, as a first approximation one might expect that λ varies inversely with m. We will assume as much in the examples that follow; although the formal results do not depend on this assumption.

There is a trade-off between individual and social costs: an individually costly action may be preferable to a zero-cost one if there are more people playing it or actions close to it. This leads us to introduce the concept of a norm. We say that a state z is *homogenous* if all agents play the same action; that is, there exists i such that $z_i = m$. We shall denote the state in which all agents play action i by z^i. A state z is *stable* if no agent strictly prefers to change action. Formally, for all i such that $z_i > 0$ and for all j, $u(i, z) \geq u(j, z)$. A *norm* is a homogenous, stable state; that is, a strict Nash equilibrium.

Let Z^* be the set of norms. State z^0 is always a norm, and for every $i > 0$, z^i is a norm if and only if for all $j < i$,

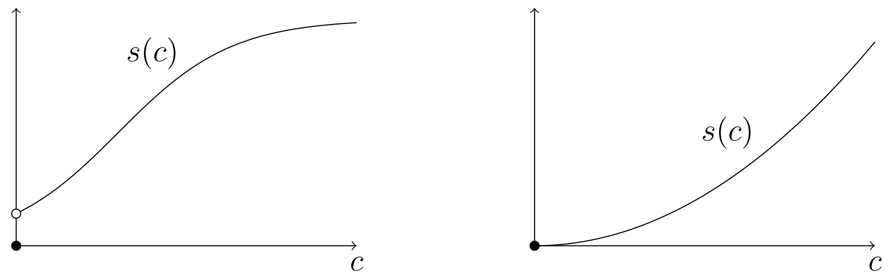
$$\frac{s(c_i - c_j)}{c_i - c_j} \geq \frac{1}{\lambda(m-1)}. \tag{5}$$

Let $\kappa = |Z^*|$ be the number of norms, which we assume to be finite.

4 The shape of the sanctions function

The shape of the sanctions function will play a key role in the dynamics of the model. In particular, we will consider the case where $s(c)/c$ is decreasing and the case where it is increasing. We do not assume that s is continuous.

When average sanctions are decreasing, small differences in cost have a relatively *greater* impact than do large differences in costs. This condition is implied by concavity of s, but is substantively weaker, for instance if there is a discontinuous jump at zero (as shown in figure 2a). In the case of duelling, this seems plausible. Small deviations from the prevailing norm, such as requesting additional paces, may be viewed as cowardice and trigger severe



(a) $s(c)/c$ strictly decreasing.

(b) $s(c)/c$ strictly increasing.

Figure 2: Two sanctions functions.

sanctions. Small deviations may also evidence willingness to violate norms in general.

When average sanctions are increasing, small differences in cost have a relatively *smaller* impact than do large differences in costs. This condition is implied by convexity of s (as shown in figure 2b), but is weaker. An increasing $s(c)/c$ seems appropriate in cases where small deviations from the norm may be difficult to observe (as with FGC). In this situation minor deviations may not be detected, so the impact of sanctions will be low in expectation. In contrast, large deviations will be easier to detect and yield relatively large sanctions in expectation.

Lemma 1 characterises the behaviour of the set of norms Z^* under both cases.

Lemma 1. *If $s(c)/c$ is weakly decreasing on $(0, c_n)$ then there exists $c \in [0, c_n]$ such that*

$$Z^* = \{z^i : c_i \leq c\}. \tag{6}$$

If $s(c)/c$ is weakly increasing on $(0, c_n)$ then a_i for $i > 0$ is stable if and only if

$$\frac{s(c_i - c_{i-1})}{c_i - c_{i-1}} \geq \frac{1}{\lambda(m-1)}. \tag{7}$$

Proof. For the first part, suppose $s(c)/c$ is weakly decreasing on $(0, c_n)$. Suppose z^i is a norm and consider z^j $j < i$. We need to show that z^j is a norm.

Since z^i is a norm, we have for all $k < j$,

$$\frac{s(c_i - c_k)}{c_i - c_k} \geq \frac{1}{\lambda(m-1)}.$$

But since $s(c)/c$ is weakly decreasing, we have for all $k < j$,

$$\frac{s(c_j - c_k)}{c_j - c_k} \geq \frac{s(c_i - c_k)}{c_i - c_k} \geq \frac{1}{\lambda(m-1)},$$

So z^j is a norm.

For the second part, suppose $s(c)/c$ is weakly increasing on $(0, c_n)$. Action z^i is a norm if and only if for all $i < j$,

$$\frac{s(c_i - c_j)}{c_i - c_j} \geq \frac{1}{\lambda(m-1)}.$$

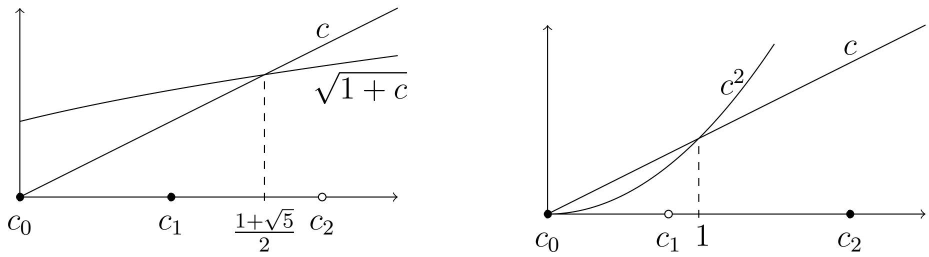
But since $s(c)/c$ is weakly increasing, $\frac{s(c_i - c_j)}{c_i - c_j}$ is minimised for $c_j = c_{i-1}$. Therefore z^i is a norm if and only if

$$\frac{s(c_i - c_{i-1})}{c_i - c_{i-1}} \geq \frac{1}{\lambda(m-1)}. \tag{8}$$

□

Lemma 1 follows from the fact that when $s(c)/c$ is decreasing, the best action for a player to deviate to (starting from a homogenous state) is always the null action a_0. This is because a_0 is where social cost is least important relative to individual cost. In contrast, when $s(c)/c$ 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 cost. This dynamic is the core of the model and will be important for many of the following results.

Example (Decreasing averages sanctions)**.** Suppose there are three actions labelled a_0, a_1, a_2, with costs $c_0 = 0, c_1 = 1, c_2 = 2$. Suppose $s(c) = \sqrt{1+c}$ for $c > 0$ and $s(0) = 0$, and $\lambda = 1/(m-1)$. Action a_i for $i > 0$ is stable if and only if $\frac{s(c_i)}{c_i} \geq 1$. Thus any action with cost lower than $(1+\sqrt{5})/2$ is



(a) $s(c)/c$ strictly decreasing.

(b) $s(c)/c$ strictly increasing.

Figure 3: Actions and norms in the examples.

a norm. In particular, actions a_0 and a_1 are stable, but a_2 is not. This is illustrated in figure 3a. Graphically, the stability threshold is the point at which curves c and $\sqrt{1+c}$ cross. When a player considers deviating from a norm at c to a_0, c is the individual cost she avoids, while $\sqrt{1+c}$ is the social cost she incurs. As long as the latter is larger, the player has no incentive to deviate.

Example (Increasing average sanctions). Now suppose that the actions a_0, a_1, a_2 have costs $c_0 = 0, c_1 = 0.8, c_2 = 2$. Suppose $s(c) = c^2$ for all c and $\lambda = 1/(m-1)$. Action a_i for $i > 0$ is stable if and only if $\frac{s(c_i - c_{i-1})}{c_i - c_{i-1}} \geq 1$. Thus, actions a_0 and a_2 are stable, but a_1 is not. This is illustrated in figure 3b.

5 Best-response dynamics

To fix ideas, we start with the case where agents best respond to the current state. We will later introduce noise, leading to richer dynamics.

Assume that time is continuous and that agents update their actions via independent Poisson arrival processes with unit expectation. Thus, every agent updates once per unit of time in expectation, and the probability is zero that agents update simultaneously. When an agent updates at time t, she chooses a best response to the current state $z(t)$. (If there are multiple best responses, she chooses among them with uniform probability.)

We begin by observing that the game defined by $u_i(a_i, z)$ in expression (4) is a potential game with potential function

$$\rho(z) = -\sum_i z_i c_i - \lambda \sum_{0 \leq i \leq j \leq n} z_i z_j s(|c_j - c_i|) \tag{9}$$

Indeed, suppose that an agent playing i changes to j. Then the change in the agent's payoff is

$$c_i - c_j - \lambda(z_i - z_j - 1)s(|c_i - c_j|) - \lambda \sum_{k \neq i,j} z_k[s(|c_j - c_k|) - s(|c_i - c_k|)].$$

This is precisely the change in ρ, hence ρ is a potential function. It follows that, starting from any initial state $z(0)$, the process converges in finite time to a pure Nash equilibrium with probability one. This follows from two facts: (i) pure Nash equilibria exist, and (ii) there is a positive number α such that, in any nonequilibrium state z, there is at least one agent that can improve his payoff by at least α.

We show next that the only pure Nash equilibria are the norms, and establish an upper bound on the expected waiting time to converge to a norm from an arbitrary initial state.

Let

$$\gamma = \min_{0 \leq i \leq j \leq n} s(c_j - c_i) \tag{10}$$

Theorem 1. *Starting from any initial state $z(0)$, the process converges to a norm in finite time with probability one, and the expected waiting time to reach a norm is at most*

$$(n - \kappa) \left[\frac{|\rho(z(0))|}{\gamma} + 1 \right]. \tag{11}$$

Before giving the proof, let us observe that the adjustment process can be quite complex. Consider the example in figure 4. There are ten agents distributed among two actions a_1 and a_2, with costs c_1 and c_2. Depending on the shape of $s(c)$, it could be that agents at a_1 would prefer to move to a_2 even though this increases their individual cost by $c_2 - c_1$. At the same time, agents at a_2 might prefer to move to a_0 because the cost savings outweigh

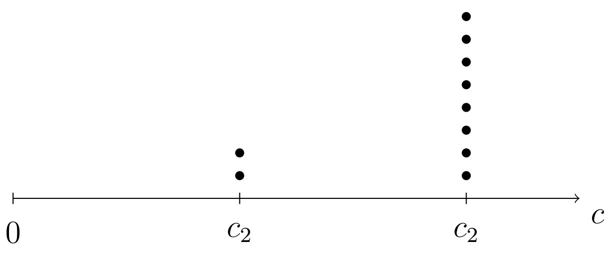


Figure 4: Adjustment of the best-response dynamic.

the increased sanctions from those remaining at a_2. Thus, depending on the order in which agents update, the process could move to a homogenous but more costly state at a_2, or it could unravel to the zero norm.

The proof of theorem 1 makes use of the following lemma:

Lemma 2. *Inhomogeneous states are unstable. Moreover, if* $z_i, z_j > 0$*, and* $z_k = 0$ *for all* $i < k < j$*, then some agent at* a_i *or* a_j *can increase her payoff by at least*

$$s(|c_i - c_j|) \geq \gamma \tag{12}$$

Proof. Consider some inhomogeneous state z. Assume $0 < i < j$ are such that such that $z_k = 0$ for all $i < k < j$. Let $c = c_j - c_i$ and $s = s(c_j - c_i)$. Let

$$\forall k > j \quad \Delta_k = z_k(s(c_k - c_i) - s(c_k - c_j)) \tag{13}$$

$$\forall k < i \quad \Delta'_k = z_k(s(c_j - c_k) - s(c_i - c_k)) \tag{14}$$

First, suppose someone at a_i switches to a_j. The change in payoff is:

$$\Delta = -c + \sum_{k>j} \Delta_k - \sum_{k<i} \Delta'_k - (z_i - 1)s + z_j s \tag{15}$$

For someone at a_j switching to a_i, the payoff change is:

$$\Delta = c - \sum_{k>j} \Delta_k + \sum_{k<i} \Delta'_k - z_i s + (z_j - 1)s \tag{16}$$

Let

$$\alpha = -c + \sum_{k>j} \Delta_k - \sum_{k<i} \Delta'_k - z_i s + z_j s \tag{17}$$

Then $\Delta = \alpha + s$ and $\Delta' = -\alpha + s$. Therefore $\Delta \geq s$ or $\Delta' \geq s$ (or both).

Therefore some agent can move and increase their payoff by at least s. This establishes lemma 2. □

We remark that this argument holds for any non-negative sanctions function $s(c)$ irrespective of whether $s(c)/c$ is monotone decreasing.

Proof of theorem 1. We first bound the expected waiting time until the process reaches a homogenous state. The argument will be more transparent if we consider the associated discrete-time process in which one agent is drawn uniformly at random 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.

Lemma 2 implies that if $z(t)$ is inhomogenous, the expected change in ρ satisfies

$$E_t[\rho(z(t+1)) - \rho(z(t))] \geq \gamma/m \tag{18}$$

In general we can write

$$\rho(z) = -\sum_i c_i z_i - D(z), \tag{19}$$

$$\text{where } D(z) = \sum_{0 \leq i \leq j \leq n} z_i z_j s(c_j - c_i) \tag{20}$$

Starting from $\mathbf{z}(0)$, let the random variable T be the first time such that $D(z(T)) = 0$. In other words T is the first time that the process is in a homogenous state.

Define the function

$$h(t) = \rho(z(t)) - t\gamma. \tag{21}$$

By inequality (18),

$$E_t[h(t+1) - h(t)] \geq 0. \tag{22}$$

Hence $h(t)$ is a submartingale. The value of the stop time T is finite with probability one, hence Doob's optional sampling theorem implies that

$$E[h(T)] \geq h(0) \tag{23}$$

$$\Rightarrow \quad E[\rho(z(T))] - \rho(z(0)) \geq \gamma E[T]$$

$$\Rightarrow \quad E[T] \leq \frac{E[\rho(z(T))] - \rho(z(0))}{\gamma/m}$$

$$\Rightarrow \quad E[T] \leq -\frac{\rho(z(0))}{\gamma/m}$$

The latter because ρ is weakly negative. Therefore

$$E[T] \leq \frac{m|\rho(z(0))|}{\gamma} \tag{24}$$

The state $z(T)$ is homogenous but it might be unstable. If this is the case, then in one more period some agent will move, and the resulting state will be inhomogenous. Now repeat the argument starting at state $z(T)$ and let T' be the first time such that $z(T')$ is homogenous. Since ρ is always increasing,

$$|\rho(z(T'))| < |\rho(z(T))| \tag{25}$$

Continue in this manner until the first time T^* that $z(T^*)$ is homogenous and stable. Inequality (25) implies that each unstable homogenous state is visited at most once, and by inequality (24) the waiting time between homogenous states is at most $|\rho(z(0))|/\gamma$. Hence

$$E[T^*] \leq (n-\kappa)m\left[\frac{|\rho(z(0))|}{\gamma} + 1\right] \tag{26}$$

The expected waiting time in the continuous process is $E[T^*]/m$. This concludes the proof of of theorem 1. $\square$

6 Exogenous societal change

The preceding section shows how a norm becomes established from an arbitrary state. We now discuss what causes norms to shift. There are various possible mechanisms: one is exogenous societal change, another is idiosyncratic variation in agents' behaviours, and yet a third is coordinated intervention. We begin by analysing the effect of exogenous societal change, then in section 7 we look at endogenous (bottom-up) norm shifts that are induced by idiosyncratic variation. We also show how the framework can be adapted to the analysis of top-down interventions.

We model societal changes by changes in the strength of sanctions λ: the higher λ, the more important sanctions are; the lower λ, the more important individual costs are. In particular, suppose there is some change in society that causes the parameter λ to drop. A downward shift in λ may reduce the set of norms, making some norms unstable. If the current norm stays stable, this should not have an impact in the short run (although it may influence intermediate-run dynamics, as discussed in the following section). What if the current norm becomes unstable? It turns out that if $s(c)/c$ is strictly decreasing, the best-response process will converge to the zero-cost norm. In contrast, when $s(c)/c$ is strictly increasing, the process converges to the closest norm.

Lemma 3. *Suppose the current state is some norm z^i with $i > 0$, and suppose λ decreases making z^i unstable. Let λ' be the new value of λ and let z^j, $j < i$, be the highest norm below z^i under λ'. If $s(c)/c$ is strictly decreasing then the best-response process converges to z^0. If $s(c)/c$ is strictly increasing then the best-response process converges to z^j.*

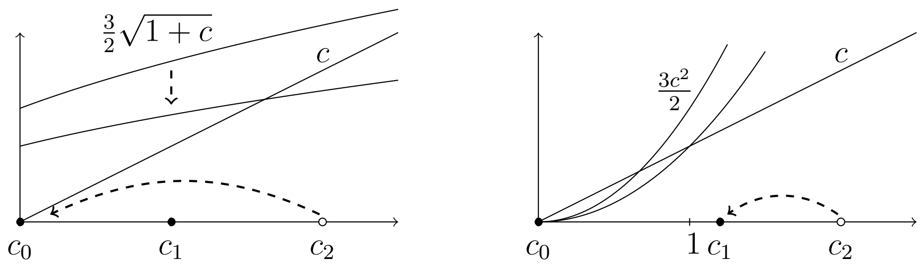
Proof. For the first part, let v_i be the utility agents get from playing the unstable action a_i under λ'. Let v_k be the utility agents get from playing some action a_k with cost $c_k < c_i$. Then

$$\begin{aligned} v_i &= -c_i, \text{ and} \\ v_k &= -c_k - \lambda' s(c_i - c_j). \end{aligned}$$

Note that

$$\begin{aligned} v_k - v_i &= c_i - c_k - \lambda' s(c_i - c_k), \\ &= (c_i - c_k)\left(1 - \lambda' \frac{s(c_i - c_k)}{c_i - c_k}\right). \end{aligned}$$

If $s(c)/(c)$ is strictly decreasing, then $v_k - v_i$ is uniquely maximised when



(a) $s(c)/c$ strictly decreasing.

(b) $s(c)/c$ strictly increasing.

Figure 5: Collapse after exogenous societal change.

$c_i - c_k$ is maximised. So $a_k = a_0$ is the unique best response.

The second part is similar and is left to the reader. □

Example (Decreasing average sanctions)**.** Suppose that as before, there are three actions labelled $a_0, a_1, a_2,$, with costs $c_0 = 0, c_1 = 1, c_2 = 2$. Suppose $s(c) = \sqrt{1+c}$ for $c > 0$ and $s(0) = 0$. When $\lambda(m-1) = 3/2$, all three actions are stable. Now suppose that λ falls so that $\lambda'(m-1) = 1$, as illustrated in figure 5a. Then action a_2 becomes unstable, and the process converges to norm z^0, jumping over the intermediate norm z^1.

Example (Increasing average sanctions)**.** Suppose that the as before $s(c) = c^2$ and $c_0 = 0$ and $c_2 = 2$. However suppose that now $c_1 = 1.2$. Suppose we start with $\lambda(m-1) = 3/2$ as above, so that all three actions are stable. Suppose $s(c) = c^2$ for all c and $\lambda = 1/(m-1)$. Now suppose that λ falls so that $\lambda'(m-1) = 1$. Then action a_2 becomes unstable, but this time the process converges to the intermediate norm z^1. This is illustrated in figure 5b.

7 Bottom-up shocks and top-down interventions

We saw in section 6 how norm shifts can be brought about by exogenous societal changes. We now consider two other causes of norm shifts: bottom-up idiosyncratic variation and top-down policy interventions.

In practice, people don't always best-respond to the current state and may deviate for a number of reasons, including lack of information, inattention, or idiosyncratic variation in preferences. If enough deviations accumulate, this can cause the process to tip to a new norm. Policy interventions can also lead to people deviating from the norm. Crucially, an intervention doesn't need to target everyone in order to displace a norm – it just needs to create enough deviations for the process to tip. In both cases, the key question is how many deviations are necessary to make the norm unstable. We capture this using the concept of *resistance*. From a given norm, we show that depending on the shape of the sanctions function, the process transits to different intermediate norms, leading to different intermediate-run dynamics.

We begin by considering norm shifts driven by bottom-up stochastic perturbations. Assume that each time an agent updates her action, she normally best responds to the current state, but with probability $\varepsilon \geq 0$ she chooses an action at random.[32] This defines a stochastic process P^{ε} over the state space Z. Suppose the process starts at some costly norm. In the short run, some players will deviate but this will not be enough to displace the norm. In the intermediate run, enough deviations may accumulate to make some other action optimal, at which point players will start to switch and this action will become the new norm. The time it takes to make such a transition depends on such factors as the probability with which agents make mistakes, the topology of group interactions, and the resistance to making the transition, a concept that we consider next.[33]

Consider two norms $z^i, z^j \in Z^*$. The *resistance* r_{ij} is the minimum number of mistakes required for the process to go directly from z^i to z^j without visiting intermediate norms. As we show in theorem 2, when $\varepsilon \to 0$, transitions that have lower resistances become arbitrarily more likely than transitions that have higher ones. In some cases, indirect transitions may require fewer total mistakes, *e.g.* we may have $r_{ik} + r_{kj} < r_{ij}$. We define the *global*

32 This is the uniform error model (Young 1993; Kandori, Mailath and Rob 1993; Ellison 1993; Jackson 2008). Comparable results hold for the logit model, but are omitted for the sake of conciseness.

33 For recent work on waiting times in evolutionary models, see Kreindler and Young (2013, 2014) and Arieli, Peretz and Young (2019). In experimental studies, departures from best response behaviour have been estimated to occur 3–5% of the time (Lim and Neary 2016; Mäs and Nax 2016; Hwang *et al.* 2018).

resistance $\tilde{r}_{ij}$ to be the minimum number of mistakes to transition from z^i to z^j possibly through intermediate norms. It is the minimum number of group members who must be persuaded to change their behaviour in order to make it in the interest of the rest of the group to follow suit. In particular, note that the *difficulty of inducing a norm shift by a top-down intervention* may be better captured by the global resistance of the transition than the simple resistance, depending on the specifics of the intervention.

Define the minimum outgoing resistance or *radius* of $z^i \in Z^*$ to be

$$r_i := \min_{j \neq i} r_{ij}.$$

The radius r_i is a measure of how difficult it is to exit z^i. Given some norm $z^i \in Z^*$, we can also define the set of *most likely successors*; that is,

$$L_i = \{z^j \in Z^* : r_{ij} \leq r_{ik}; \forall k \neq i, j\} \tag{27}$$

Theorem 2 establishes that the concepts of resistance and radius can be used to answer the question of what transitions are likely in the intermediate run.

Theorem 2. *If the state at time t is norm $z^i \in Z^*$, the probability of transitioning to a norm in L_i is at least a factor of $O(1/\varepsilon)$ more likely than transitioning to any norm not in L_i.*

Theorem 2 establishes that, provided noise is sufficiently small, when the process shifts, it will likely shift to a norm whose transition has minimal resistance.[34] This means that a good approximation of the intermediate-run dynamics of the process when noise is small is that from norm z^i we should expect the process to transition to some norm in L_i.

We now analyse the resistances in more detail. Theorem 3 provides a formula for the resistance of a transition, and characterises the set of most likely successors under the strictly decreasing and strictly increasing average

34 This follows from standard results on first exit times (see for example Ellison 2000, lemma 5); for the sake of completeness we give the proof in the appendix.

sanctions cases.

Theorem 3. *Consider any monotone increasing sanctions function $s(c)$ with $s(0) = 0$.*

(a) For any norms z^i, z^j such that $i > j$,

$$r_{ij} = \left\lceil \frac{m-1}{2} - \frac{c_i - c_j}{2\lambda s(|c_i - c_j|)} \right\rceil.$$

(b) If $s(c)/c$ is strictly decreasing for $c \neq 0$, then for any norm z^i with $i > 0$ there is some $\bar{c}_i < c_i$ such that

$$L_i = \{z^j \in Z^* : c_j \leq \bar{c}_i\}.$$

(c) If $s(c)/c$ is strictly increasing for $c \neq 0$, then for any norm z^i with $i > 0$ there is some $\bar{c}_i < c_i$ such that

$$L_i = \{z^j \in Z^* : c_j \geq \bar{c}_i\}.$$

Proof. See appendix. □

Theorem 3 implies that under decreasing average sanctions, the process will tend to tip suddenly to the null norm (or one close to it) in the intermediate run. In contrast under increasing average sanctions, the process will tend to decline gradually via intermediate norms. The following example illustrates the decreasing average sanctions case.

Example (Decreasing average sanctions)**.** Suppose $s(c) = \sqrt{c}/4$, $\lambda = 1/(m-1)$, and $m = 11$. Suppose that costs are evenly spaced at unit intervals: $c_i = i$, $i = 0, 1, \ldots, n$. Action a_i is stable if and only if

$$c_i \leq 6.25. \tag{28}$$

For any norms z^i, z^j such that $i > j$,

$$r_{ij} = \left\lceil 5 - 2\sqrt{c_i - c_j} \right\rceil. \tag{29}$$

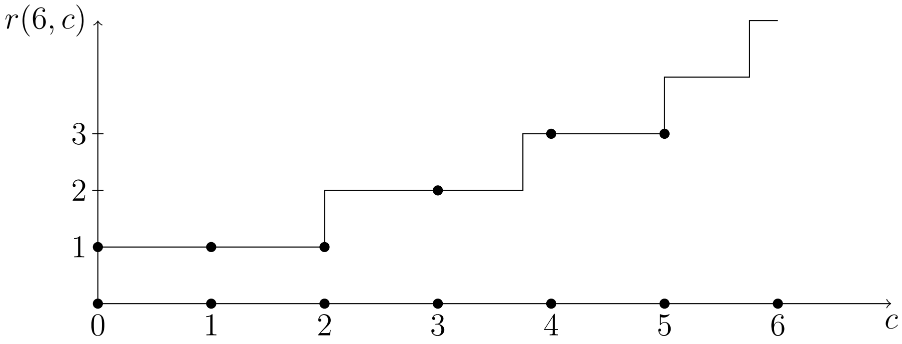


Figure 6: Resistances from norm 6.

Figure 6 illustrates the resistance of different transitions from norm z^6. As c decreases, r_{6c} also decreases, but in a stepwise manner. Note that there are a number of ties between transitions, and in particular the least-resistance set is

$$L_6 = \{0, 1, 2\}. \tag{30}$$

Figure 7 shows the least-resistance sets for each norm and the corresponding resistances. Norm z^6 is relatively unstable, because its radius is 1, so we would expect it to be displaced relatively quickly. However, the least-resistance set contains three norms, so the process could end up transitioning to a non-zero norm. If, for instance, the process transitions to norm z^2, it may stay there for a long time, since z^2 has a radius of 3. In contrast, norm z^4 has the same radius as norm z^6, but only one norm in the least-resistance set: the zero norm.

8 Conclusion

This paper has developed a general framework for studying the dynamics of norms that pose harm to individuals but remain in place due to social pressure. We have taken three particular cases to illustrate the theory – foot binding, FGC, and duelling – but the framework is general and has potential application to many other examples. The model is based on concepts from evolutionary game theory, but unlike much of this literature, which characterises the long-run behaviour of the evolutionary process, our focus here has

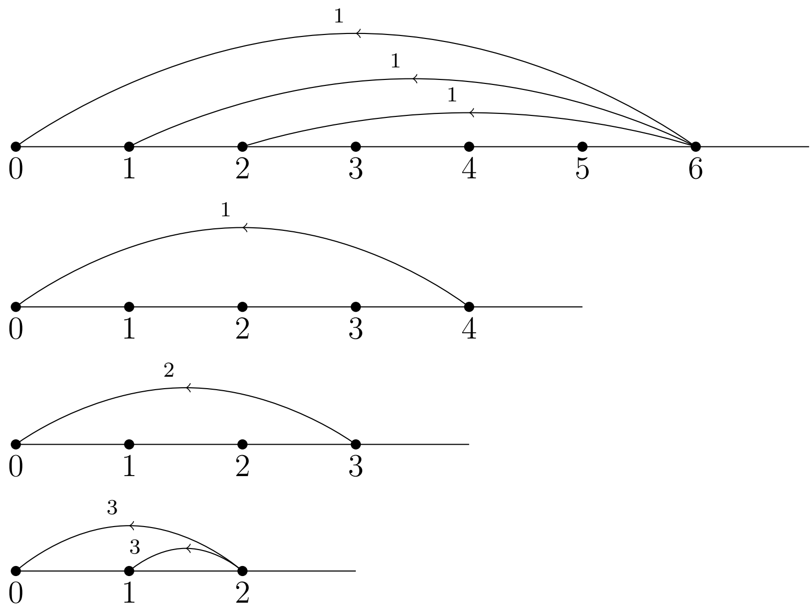


Figure 7: Least-resistance sets.

been on the intermediate-run dynamics. Furthermore, our model takes account of the fact that norms often prescribe a package of behaviours, where many variants of the norm are possible.

The theory allows us to examine the impact of several different factors that induce norm shifts. One is exogenous change in the social environment that attenuates the force of sanctions. Such changes seem to have played a significant role in the decline of duelling: as European societies became larger, more urbanised and interactions more diffuse, traditional codes of honour gradually lost their force. Similarly, as Chinese society evolved over the course of the nineteenth century, foot binding came to be seen as an antiquated procedure that was not in step with modern and progressive ideas. Our model shows that such exogenous changes can destabilise an existing norm, and in fact cause it to collapse completely.

Another factor inducing norm shifts is idiosyncratic variations in behaviour: if enough of these variations accumulate the process may tip into a new norm. The theory shows that high-cost norms are the most susceptible to being displaced by such forces. Depending on the shape of the sanctions function, norms may tend to collapse altogether or to ratchet down gradually. It remains to be seen what happens in the case of FGC: although there is some evidence that the most costly forms of the practice have been in decline over the past few decades, it is still too early to tell.

Finally the theory allows us to draw implications for policy interventions aimed at ending costly social norms like FGC. We argue that the concept of resistance can be interpreted as a measure of the difficulty of inducing a norm shift by a top-down intervention. The theory then implies that the shape of the sanctions function is a key determinant of the effectiveness of a policy: under some conditions policies targeting transition to an intermediate variant of a norm are optimal, whereas under other conditions policies targeting full abandonment are preferable.

Appendix

Theorem 2. *If the state at time t is norm $z^i \in Z^*$, the probability of transitioning to a norm in L_i is at least a factor of $O(1/\varepsilon)$ more likely than transitioning to any norm not in L_i.*

Proof. Let $r = r_i$ be the radius of z^i. Starting in norm z^i at time t, let $t+T_1$ be the first time the process reaches a state where exactly $m-r$ agents are choosing a_i. Note that all such states are equiprobable because non-best replies are always chosen with the same probability. Let p be the probability that $z(t+T_1)$ is in the basin of attraction of some norm in L_i. Note that p is a function of $|L_i|$ and $|A|$, but does not depend on ε.

If $z(t+T_1)$ is not in the basin of attraction of some norm in L_i, let $t+T_1+T_2$ be the first time such that someone makes an error or the norm z^i is reached with no one making an error. Note that since $z(t+T_1)$ is not in the basin of attraction of some norm in L_i, action a_i is the unique best reply. Each period, either the number of a_i-players goes up by one (event A), an error is made (event B), or neither occurs (event C). We have $P(A) \geq (1-\varepsilon)/(n+1)$, and $P(B) \leq \varepsilon$. Therefore the probability that r A-events occur before any

B-event occurs is at least

$$g(\varepsilon) = \left[\frac{(1-\varepsilon)/(n+1)}{1-\varepsilon)/(n+1)+\varepsilon}\right]^r.$$

The probability that an error is made after $t + T_1$ and before norm z^i is reached is at most $(1-p)(1-g(\varepsilon))$. Note that this is the only case when the process could enter the basin of a norm not in L_i.

Therefore the probability that the process enters the basin of a norm in L_i, relative to the probability of entering a basin of a norm not in L_i is at least

$$\frac{p}{(1-\varepsilon)(1-g(\varepsilon))} \approx \frac{p}{(1-\varepsilon)\varepsilon(n+1)} = O(1/\varepsilon).$$

□

Theorem 3. *Consider any monotone increasing sanctions function $s(c)$ with $s(0) = 0$.*

(a) For any norms z^i, z^j such that $i > j$,

$$r_{ij} = \left\lceil \frac{m-1}{2} - \frac{c_i - c_j}{2\lambda s(|c_i - c_j|)} \right\rceil.$$

(b) If $s(c)/c$ is strictly decreasing for $c \neq 0$, then for any norm z^i with $i > 0$ there is some $\bar{c}_i < c_i$ such that

$$L_i = \{z^j \in Z^* : c_j \leq \bar{c}_i\}.$$

(c) If $s(c)/c$ is strictly increasing for $c \neq 0$, then for any norm z^i with $i > 0$ there is some $\bar{c}_i < c_i$ such that

$$L_i = \{z^j \in Z^* : c_j \geq \bar{c}_i\}.$$

Proof. Suppose that $s(c)/c$ is strictly decreasing. First we establish claim *(a)*. Define

$$r_{ij} := \left\lceil \frac{m-1}{2} - \frac{c_i - c_j}{2\lambda s(|c_i - c_j|)} \right\rceil$$

We will show that r_{ij} is in fact the resistance of the transition $z^i \to z^j$. The proof is in two steps.

First, consider 'direct' paths from z^i to z^j that only involve mistakes to a_j. Suppose k agents have moved (in error) to a_j. Action a_j is preferred by each remaining agent at a_i if and only if

$$\begin{aligned} & -c_j - \lambda(m-k-1)s(|c_i - c_j|) > -c_i - \lambda k s(|c_i - c_j|) \\ \Leftrightarrow\ & k > \frac{m-1}{2} - \frac{c_i - c_j}{2\lambda s(|c_i - c_j|)}. \end{aligned}$$

The quantity r_{ij} is the least such k. Therefore, r_{ij} is the minimum cost among paths that only involve mistakes to a_j.

Second, we need to consider 'indirect' paths from z^i to z^j that involve mistaken choices of actions other than a_j. (However we do not need to consider paths that involve actions other than a_i or a_j being a best response at any time.) From any given state, suppose one player switches from a_i to some other action $a_k \neq a_j$. The payoff from playing a_i decreases by $\lambda s(|c_i - c_k|)$, while the payoff from playing a_j decreases by $\lambda s(|c_j - c_k|)$. Let

$$\Delta_k = s(|c_i - c_k|) - s(|c_j - c_k|).$$

Now suppose a player switches from a_i to a_j. The payoff from playing a_i decreases by $\lambda s(|c_i - c_j|)$ while the payoff from playing a_j increases by $\lambda s(|c_i - c_j|)$. Let

$$\Delta_j = 2s(|c_i - c_j|).$$

A path that involves mistakes to $a_k \neq a_j$ can only be a minimal-resistance path if $\Delta_k > \Delta_j$. We argue that $\Delta_k < \Delta_j$:

1. First, suppose c_k is between c_i and c_j. Then $s(|c_i - c_k|) < s(|c_i - c_j|)$, so $\Delta_k < \Delta_j$.

2. Second, suppose c_k is not between c_i and c_j and is closer to c_i. Then $\Delta_k < 0$, so $\Delta_k < \Delta_j$.

3. Third, suppose c_k is not between c_i and c_j and is closer to c_j. Then

$|c_i - c_j| + |c_j - c_k| = |c_i - c_k|$. Since $s(c)/c$ is strictly decreasing, it satisfies the triangle inequality strictly. Hence $s(|c_i - c_j|) + s(|c_j - c_k|) > s(|c_i - c_k|)$, which in turn implies $\Delta_k < \Delta_j$.

Next we shall establish claim *(b)*, namely $r_{ki} \leq r_{kj}$ for any norms z^i, z^j, z^k such that $i < j < k$. Since $c_k > c_j > c_i$ we have

$$r_{ki} = \left\lceil \frac{m-1}{2} - \frac{c_k - c_i}{2\lambda s(c_k - c_i)} \right\rceil$$
$$r_{kj} = \left\lceil \frac{m-1}{2} - \frac{c_k - c_j}{2\lambda s(c_k - c_j)} \right\rceil$$

Since $s(c)/c$ is strictly decreasing and $c_k - c_i > c_k - c_j$, the result follows immediately.

Finally, claim *(c)* follows immediately from the preceding result. □

References

Acemoglu, D., and M. O. Jackson. 2015. 'History, expectations, and leadership in the evolution of social norms'. *Review of Economic Studies* 82: 423–456.

Akerlof, G. A. 1980. 'A theory of social custom, of which unemployment may be one consequence'. *Quarterly Journal of Economics* 94: 749–775.

———. 1997. 'Social distance and social decisions'. *Econometrica* 65: 1005–1027.

Akerlof, R. 2017. 'Value formation: The role of esteem'. *Games and Economic Behavior* 102: 1–19.

Arieli, I., R. Peretz and H. P. Young. 2019. 'The speed of innovation diffusion in social networks'. Forthcoming in *Econometrica*.

Axelrod, R. 1986. 'An evolutionary approach to norms'. *American Political Science Review* 80: 1095–1111.

Banks, S. 2008. 'Killing with courtesy: The English duelist, 1785–1845'. *Journal of British Studies* 47: 528–558.

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

Belloc, M., and S. Bowles. 2013. 'The persistence of inferior cultural-institutional conventions'. *The American Economic Review* 103: 93–98.

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

Blume, L., and S. N. Durlauf. 2001. 'The interactions-based approach to socioeconomic behavior'. Chap. 2, edited by S. N. Durlauf and H. P. Young, 15:15–44. Cambridge, MA: MIT Press.

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

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

Bowles, S., and J.-K. Choi. 2013. 'Coevolution of farming and private property during the early Holocene'. *Proceedings of the National Academy of Sciences* 110: 8830–8835.

———. 2019. 'The Neolithic agricultural revolution and the origin of private property'. *Journal of Political Economy* 127.

Brock, W. A., and S. N. Durlauf. 2001. 'Discrete choice with social interactions'. *The Review of Economic Studies* 68: 235–260.

Burke, M. A., G. M. Fournier and K. Prasad. 2010. 'Geographic variations in a model of physician treatment choice with social interactions'. *Journal of Economic Behavior & Organization* 73: 418–432.

Carvalho, J.-P. 2013. 'Veiling'. *Quarterly Journal of Economics* 128: 337–370.

Chesnais, J.-C. 1981. *Histoire de la violence.* Paris: Robert Laffont.

Efferson, C., S. Vogt, A. Elhadi, H. E. F. Ahmed and E. Fehr. 2015. 'Female genital cutting is not a social coordination norm'. *Science* 349: 1446–1447.

Ellison, G. 1993. 'Learning, local interaction, and coordination'. *Econometrica:* 1047–1071.

———. 2000. 'Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution'. *Review of Economic Studies* 67: 17–45.

Gates, H. 2015. *Footbinding and women's labor in Sichuan.* Abingdon and New York: Routledge.

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

Hwang, S.-H., W. Lim, P. Neary and J. Newton. 2018. 'Conventional contracts, intentional behavior and logit choice: Equality without symmetry'. *Games and Economic Behavior* 110: 273–294.

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

Kandori, M., G. J. Mailath and R. Rob. 1993. 'Learning, mutation, and long run equilibria in games'. *Econometrica* 61: 29–56.

Ko, D. 2007. *Cinderella's sisters: A revisionist history of footbinding.* Berkeley and Los Angeles: University of California Press.

Kreindler, G. E., and H. P. Young. 2013. 'Fast convergence in evolutionary equilibrium selection'. *Games and Economic Behavior* 80: 39–67.

———. 2014. 'Rapid innovation diffusion in social networks'. *Proceedings of the National Academy of Sciences* 111: 10881–10888.

Kudo, Y. 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.

Lim, W., and P. R. Neary. 2016. 'An experimental investigation of stochastic adjustment dynamics'. *Games and Economic Behavior* 100: 208–219.

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

Mäs, M., and H. H. Nax. 2016. ‘A behavioral study of ‘noise’ in coordination games’. *Journal of Economic Theory* 162: 195–208.

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

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

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

Shell-Duncan, B., K. Wander, Y. Hernlund and A. Moreau. 2011. ‘Dynamics of change in the practice of female genital cutting in Senegambia: testing predictions of social convention theory’. *Social science & medicine* 73: 1275–1283.

Skyrms, B. 1996. *Evolution of the social contract.* Cambridge: Cambridge University Press.

———. 2003. *The stag hunt and the evolution of social structure.* Cambridge: Cambridge University Press.

Stone, L. 1965. *The crisis of the aristocracy, 1558–1641.* Oxford: Oxford University Press.

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

United Nations Children’s Fund. 2013. *Female genital mutilation/cutting: A statistical overview and exploration of the dynamics of change.* New York: UNICEF.

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

Wagner, N. 2015. ‘Female genital cutting and long-term health consequences: nationally representative estimates across 13 countries’. *Journal of Development Studies* 51: 226–246.

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

Young, H. P. 1993. 'The evolution of conventions'. *Econometrica* 61: 57–84.

———. 1998. *Individual strategy and social structure.* Princeton: Princeton University Press.

———. 2015. 'The evolution of social norms'. *Annual Review of Economics* 7: 359–387.

Young, H. P., and M. A. Burke. 2001. 'Competition and custom in economic contracts: A case study of Illinois agriculture'. *American Economic Review* 91: 559–573.