Preprint

This is the submitted version of a paper published in .

Citation for the original published paper (version of record):

Asymmetries in punishment propensity may drive the civilizing process
https://doi.org/10.1038/s41562-017-0278-z

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:mdh:diva-38850
Title
The Civilizing Process Revisited: How Asymmetries in Punishment Propensity May Drive Norm Change

Authors
Pontus Strimling\textsuperscript{a,b,c}, Mícheál de Barra\textsuperscript{b,d}, and Kimmo Eriksson\textsuperscript{b,e}

Affiliations
\textsuperscript{a} Institute for Analytical Sociology, Linköping University
\textsuperscript{b} Centre for the Study of Cultural Evolution, Stockholm University
\textsuperscript{c} Institute for Future Studies, Stockholm
\textsuperscript{d} Institute of Applied Health Sciences, University of Aberdeen
\textsuperscript{e} School of Education, Culture and Communication, Mälardalen University
Abstract

Norms about hygiene and violence both show a tendency to become increasingly strict through recorded history, in the sense that the handling of bodily fluids and the use of violence have become increasingly restricted. This directional change over many generations and across a large number of societies is a puzzling phenomenon first documented by Norbert Elias. Here we propose a theory for directional norm change that is based on the aggregation of everyday interactions. This theory posits that directional norm change can come about if there is an asymmetry in punishment propensity between the people who prefer stricter norms and those who prefer looser norms. An asymmetry in punishment can arise from an underlying asymmetry in the threat perceived, where a stricter than preferred behaviour is perceived as inherently less threatening than a looser one. We demonstrate the logic of the theory using a formal model and test some of its assumptions through survey experiments on 848 subjects.
Introduction

Social norms are informal rules about how to behave. The human mind seems to be particularly good at understanding, remembering and internalizing social norms (e.g., Campbell, 1964; Cosmides, 1989; O’Gorman et al., 2008). Experiments show that social norms can be powerful determinants of people’s behaviour (e.g., Cialdini & Trost, 1998). Thus, social norms tie together cognition with behaviour.

Theorists often treat social norms as if they are inherently stable, in equilibrium (Binmore & Samuelson, 1994). Indeed, there are several mechanisms that could serve to stabilize norms, such as sanctions of norm violators (Brauer & Chekroun, 2005) and psychological mechanisms that justify status quo (Jost et al., 2004; Eriksson et al., 2015). Nonetheless, history is replete with examples of norm change across the behavioural repertoire. For instance, the oldest Westerners living today have in their lifetime witnessed dramatic change in norms about what to wear, how to speak, and what is expected of different genders. Although these norm changes are well-known, it is a mystery why they occur (Bloom, 2010).

Hygiene, defined as how one behaves with respect to bodily fluids, is an important domain where norm change has been systematically documented. Sociologist Norbert Elias (1939) analysed 400 years of European etiquette books in his classic work Über den Prozess der Zivilisation (“The Civilizing Process” in English translation). According to Elias’s analysis, the ideal for how to handle bodily fluids have tended to become more restrictive in each generation. In other words, hygiene norms have become increasingly strict. Elias argued that this phenomenon was due to an increase in the perceived value of self-control. (We describe his ideas in more detail below.) Although the work of Elias has received some criticism (e.g., Duerr, 1988, 1994, 1995, 1997; see Elias, 2008, for a response), later empirical work has confirmed the existence of long-term changes in norms about hygiene and violence, not just in Europe but throughout the world (Mennell & Goudsblom, 1997).

Elias also suggested that the domain of this civilizing process extended to norms about the use of violence (Elias, 1982). This thread has been picked up by other researchers (e.g., Eisner, 2001; Klingenstein et al., 2014), most notably in an ambitious treatment of the civilizing process by Pinker (2011). Pinker reviews a large body of evidence on how norms regarding violence have become stricter over time, both in the sense that there are fewer situations where violence is allowed and that the violence itself, when it is allowed, tends to
become less brutal. He goes on to argue that the decline in violence seems to be part of the same process as the tightening of hygiene norms. For instance, temporary reversals of the general trend, such as the increase in violence in the US starting in the 1960s, tend to coincide with similar reversal of the general trend for hygiene related norms.

These trends towards stricter hygiene and violence norms have not been studied for each society throughout history, but they are known to have occurred across a range of societies with different formal institutions and different cultures (Ikegami, 1995; Jarrick & Söderberg, 1993; Pinker, 2011). Moreover, any reversals in the trend have tended to last only for a short time (e.g. Pah et al., 2017; Suri, 2009), and have tended to occur in societies undergoing large destabilizing changes (Elias, 1996). Thus, we believe that there is strong empirical evidence for the claim that there is a general pattern of hygiene and violence norms changing over time to become increasingly strict—and this pattern has been observed at least for several hundred years, and possibly throughout human history (Cieri et al., 2014; Pinker, 2011).

This general pattern requires an explanation: Why do norms about hygiene and violence tend to develop in this manner? Several possibilities may spring to mind, such as technological change, functionality, and institutional change. Below we argue that these suggestions cannot explain the general scope of the phenomenon. Elias himself offered a more sophisticated explanation, but again we argue that it does not seem to account for the generality of the empirical pattern. Instead we propose a novel theory of directional norm change as an emergent phenomenon arising from human psychology.

**Technological change**

Some changes in hygiene behaviour are clearly associated with technological change. As the costs of hot water, detergents, and other hygiene related technologies declined, people could afford to buy more of them. A norm for daily showers is unlikely to develop in a cold climate without affordable hot water. It is therefore a reasonable suggestion that norm change in hygiene is in fact driven by technological change. However, this can hardly serve as a general explanation. There are many examples of hygiene norm change that are unrelated to technology, such as the introduction of norms against spitting indoors, eating with mouth closed, and playing with snot in front of others (Elias, 1939). For a full account of the relationship between civilization and technology see Elias (2008).

Trends in violence norms could also rely on technological change. For instance, technological change has increased people’s access to things like novels and movies, which
enable them to view things from another person’s perspective. This could lead to more empathy and, by extension, less violence (Jarrick & Söderberg, 1994) However, experimental evidence show that the effect can just as well go in the other direction, with media exposure that depicts violence making people more, not less, willing to allow violent behaviour (Funk et. al., 2004). Thus, to the extent that media drives the change of violence norms, a separate explanation is still required for why the change would be directed towards more restrictive norms.

**Functionality**

Another possibility is that changes in hygiene norms might have been driven by an increased awareness of the role of body fluids in infectious disease transmission, that is, for functional reasons. The transition to coughing in one’s elbow instead of one’s hand is a recent example of a norm change that was arguably driven by explicit awareness of transmission pathways. But, dating as it does from the late 19th century, germ theory cannot bear much explanatory burden for long-term changes in hygiene norms. We know from Elias’s work that hygiene norms have been changing for centuries. Even today, germ theory plays little role in many people’s interpretation of common infectious diseases like the common cold (attributed to temperature changes rather than hygiene; Helman, 1978) or food poisoning (often attributed to spoilage rather than hygiene; Wilcock et al., 2004). In conclusion, it does not seem that the rise of stricter hygiene norms in general can be adequately explained by functionality.

**Institutional change**

Formal institutions may play a role in norm change in general and in particular in the domain of violence as it is typically regulated in formal law. Could change in norms about violence simply be driven by a corresponding change in laws about violence? In general, we think it could not. For one thing, violence seems to have started decreasing long before the arrival of formal laws (Pinker, 2011). For another, we would still require an explanation for why formal laws tend to become increasingly restrictive about violence.

Institutions such as trade could also drive change in violence norms indirectly, by increasing people’s dependence on others and thereby make acts of violence more costly (Wright, 2001). However, this mechanism seems unable to explain the broad spectrum of changes in violence norms. For instance, physical punishment within the family is still being practised in a majority of countries in the world (Zolotor & Puzia, 2010) whereas states
hardly every allow individuals to hit strangers except under very special circumstances, such as self-defence or defence of property. Interdependence is clearly higher within the family than between strangers. Thus, it does not seem to be the case that interdependence generally decreases violence.

**Elias’s civilizing process**

As we have seen above, there are several factors that could affect some norm changes but fail to address the generality of the trends that we study. Moreover, none of these factors seem to explain why hygiene and violence norms tend to move together. By contrast, Elias’s own explanation for the consistent trend in norm change addresses both the generality of the trends and the link between violence and hygiene. It therefore deserves a more thorough analysis.

Elias’s theory is about the interaction between social change in power structures and individual psychology (Dunning & Hughes, 2012; Elias, 1939) He regarded norms about hygiene and violence as part of a larger concept that he named “civilization”, by which Western societies distinguished themselves from, in their view, more “primitive” societies. A society’s level of civilization included its norms about hygiene and violence, but also its patterns of class, gender, sexuality, sports, and many other things. While we here focus just on hygiene and violence norms, Elias’s own studies covered several of these related phenomena too (Elias, 2008). Basically, Elias let the civilization concept include anything for which one could say there was a civilized and a non-civilized way of doing it. It is important to note that Elias did not use civilized in a normative sense, but rather as the term by which Western societies have tried to distinguish themselves from the rest of the world (Elias 1939).

Elias’s claim is that changes in civilization level go hand in hand with an increase in social pressure on people to exercise more self-control over their emotions and over an increasing set of behaviours that include both hygiene and violence (Dunning & Hughes 2012, p. 88). According to Elias, this increase in social pressure derived from two sources. On the one hand a change in power structure, namely, the centralization of power to the courts during state formation. On the other hand an increase in interdependence and population density, caused by the move towards capitalist societies, industrialisation and urbanisation. These shifts, says Elias, created a need among the influential classes (first monarchs, later the bourgeoisie) for people with self-control, and they therefore used their influence to ensure stricter norms about self-control.
Elias’s theory offers a simultaneous explanation for trends in both hygiene and violence norms. Nonetheless, the theory suffers from several weaknesses that will lead us to propose a different explanation for the phenomenon below. For instance, at the core of Elias’s proposal lies the notion that strict hygiene norms are associated with impulse control. For many individuals, however, the direct opposite is true. People who prefer strict hygiene must use self-control to do things they consider unhygienic. In fact, it is enough to see things that one considers disgusting to deplete one’s self-control (Martijn et al., 2002). Moreover, while Elias’s theory is general in scope, it relies on specific historical changes in Western Europe during the last five hundred years. Violence and hygiene norms have been found to become stricter in many other societies and historical settings (Ikegami, 1995; Jarrick & Söderberg, 1993; Mennell & Goudsblom, 1997; Pinker, 2011). It is understandable that Elias designed his theory to explain the changes for which he had data (i.e., the period in which etiquette books had been published in Europe). Data that have emerged since then indicate that the civilizing process does not look exactly like Elias envisioned. Contrary to the original conception of the process, several of the behavioural domains that Elias considered has gone through a process of informalization during the last half century (Wouters, 2007). For instance, manners for social interaction, such as greetings, have become more allowing and so have norms surrounding sexual expression (Collins, 2014; Wouters, 2004).

For the domains of hygiene and violence, however, the evidence so far suggests a continuing directional norm change. These are the domains on which we focus here. It is possible that the real scope of the process that both we and Elias try to pin down also include some of the other behavioural domains that he suggested.

Our approach

In this paper, we investigate a different explanation for the civilizing process. Rather than explaining this process as the consequence of other societal changes, we propose that it can be viewed as an emergent phenomenon arising from the way people’s psychological biases shape their social interactions. To ensure that this explanation is logically sound we have represented it in a mathematical model and analysed its dynamics. To ensure the validity of the assumptions about psychology and interaction patterns that form the basis of our explanation, we investigated these assumptions using structured surveys. In the coming sections we present the proposed explanation, followed by the mathematical model and the survey data.
Before laying out our theory in detail we want to highlight several challenges to our approach of explaining long-term directed norm change as an emergent phenomenon. First, deviations from a predominant norm tend to be met with social sanctions—even when the deviation is in the direction of being more considerate of others, more moral, and more useful to the group (Exline & Lobel, 1999; Fisher, Nadler, & Whitcher-Alagna, 1982; Monin, Sawyer, & Marquez, 2008; Parks & Stone, 2010). Indeed, the empirical data we present later in this paper indicate that both too strict and too loose hygiene and violence behaviour may be met with social sanctions. Hence, one theoretical challenge is why sanctioning of deviations does not stop norm change. Second, there are many domains where norms change over time but not in any particular direction—clothes fashion being a prototypical example. This poses the challenge to identify what is special about the domains of hygiene and violence, such that norms in these domains move in a consistent direction.

A theory of the civilizing process as an emergent phenomenon

The first building block of our theory is diversity in behavioural options. This diversity arises as a consequence of people innovating new behaviours. We assume that novel behaviours regularly appear and that they are small variations on the currently available behavioural repertoire: sometimes a bit on the loose side, other times a bit on the strict side.

Given this diversity of behavioural options, our theory focuses on individuals and their interactions. We conceive of norm change at the level of society as the aggregation of norm changes at the individual level. By a norm at the individual level we mean a preference for behaviour that applies both to their own and others’ behaviour. By this definition, a population may be subdivided into groups of individuals holding competing norms. For example, some people prefer that children are smacked by their parents, whereas a competing norm is held by people who prefer that children are not smacked. There are other conceptions of norms in the literature that do not allow for heterogeneity of preferences; for instance, Bicchieri (2005, 2016) conceives of social norms as rules of behaviour that people will comply with only if they expect other people to comply too. However, it is well-documented that individuals differ in their preferences for how people should behave, and some of the cultural and psychological roots of this heterogeneity have been studied (e.g., Haidt, 2012; Inglehart, 1997; Nisbett & Cohen, 1996; Schnall et al., 2008). The studies we present below specifically document such heterogeneity in preferences for behaviours, that is, competing norms, in the domains of hygiene and violence.
We are interested in norm shifts at the level of society, specifically, the process whereby a society can move from a state where the vast majority of individuals prefers a certain behaviour to a state where the vast majority of individuals prefers another, stricter, behaviour. We posit that this process could be driven by social sanctions as follows.

One aspect of social sanctions is what elicits them. Although social sanctions could be strategically motivated, our working assumption is that they should typically be understood mainly as an emotionally driven response to others’ behaviour (e.g., Yamagishi et al., 2009). We propose that an important determinant of an individual’s use of social sanctions with regards to hygiene and violence is whether or not he/she feels threatened by the other’s behaviour. When someone feels threatened, a fight-or-flight response is triggered in the autonomic nervous system. This response is involuntary and includes several hormone releases that prepare a body for a situation where it either needs to fight or flee (Cannon, 1929; Jansen et al., 1995). In a social situation, this would translate into either confronting or avoiding the person perceived as behaving in a threatening way. Both confrontation and avoidance serve as social sanctions.

The key idea of our theory is that there may be a systematic asymmetry in the elicitation of social sanctions between looser behaviour and stricter behaviour. Namely, we expect social sanctions to be elicited more strongly by looser behaviour. Those who are seen as unhygienic or violent should be more likely to elicit a sense of disgust or physical threat and hence a more severe social sanction, whereas overly hygienic or overly timid people are likely to elicit only annoyance. The reason for such asymmetry is that we expect the presence of violence and bodily fluids to be inherently more threatening than their absence. Bodily fluids and pathogens, as well as interpersonal violence, are generally seen as potent selection pressures in the evolution of human cognition, and hence peoples’ strong emotive responses to them have deep evolutionary roots (Bolles & Fanselow, 1980; Curtis & Biran, 2001). The studies we present below empirically establish the expected asymmetry in the use of social sanctions in the domains of hygiene and violence.

This brings us to the consequences of social sanctions. It is well-established that being targeted by social sanctions can make individuals change behaviour (e.g., Fehr & Gächter, 2000). Here we assume that such change in an individual’s behaviour will over time be internalized into a change also in the individual’s preferences (Perry & Perry, 1983). This means that social sanctions may change norms at the individual level.

Under the set of assumptions outlined above, the dynamics of each step of the civilizing process should unfold as follows. Most of the population prefers a certain behaviour, but due
to innovation a small subset of the population holds a stricter norm. When individuals holding competing norms interact, the individual with the looser behaviour will be more likely to receive a strong social sanction than the individual with the stricter behaviour. As social sanctions may make individuals change their norms, the society-level norm will over time shift towards the stricter behaviour. Due to innovation, an even stricter norm will then take up the competition, etc. In short, we propose a cultural evolutionary process in which stricter norms tend to outcompete looser norms.

Below we compare our theoretical approach with Elias’s. We then develop a mathematical model to verify the logical consistency of our proposed explanation. We then present data from surveys designed to test the validity of some of the assumptions underlying our theory: (1) heterogeneity in preferences, and (2) asymmetric use of social sanctions, due to (3) asymmetry in what behaviours are perceived as threatening.

A comparison with Elias's theory

Our theoretical approach differs from Elias’s on several points, especially in its emphasis on social sanctions instead of demand for self-control. Our theory is also silent about the role of power in these norm changes, although we do not deny that it may play a role. Indeed, it seems likely that power will influence the effectiveness of social sanctions. For the process of norm change, however, this will matter only if there is an unequal distribution of power between the proponents of the different behaviours. A priori, we see no reason to make any such assumption. If new norms tend to emerge first in the more powerful classes, as claimed by Elias (1939), this might not be due to class differences in preferences but simply due to the powerful being more effective at sanctioning and therefore more effective at spreading new norms.

There are also noteworthy similarities between our theory and the one proposed by Elias. They are both general in scope, explaining norm change for a general class of behaviours rather than for specific behaviours. Further, both theories explain the correlation between violence and hygiene norms, not as a direct effect but rather as a pattern that emerges from the same underlying force acting in both domains. Finally, both explanations are focused on the process of change through interaction between individuals. Thus, both theories are attempts at understanding the relation between individual level psychology and societal change.
A mathematical model of directional norm change

The theory outlined above concluded that the sum of all social sanctions may amount to a net force moving norms in the stricter direction. The condition for this occurring is that the asymmetry between stricter and looser individuals in their preparedness to deploy social sanctions must be strong enough to overcome the disadvantage a novel stricter behaviour faces by initially being uncommon. To ensure that this is theoretically plausible, we will formalize the process into a mathematical model. The model will thereafter be used to show how the spread of stricter norms could look over several generations. Prior work on modelling norm change has employed a variety of paradigms. Some authors model norm change as a move between different equilibria in a coordination game, made possible by some randomness in behaviour (Young, 1993). Other authors model norm change as a selection of the more favourable equilibrium between groups that have different norms (Boyd & Richerson, 2002), or by looking at how changes in the underlying payoff structure or in the individual make-up of society affect the equilibrium in a model where individuals care about their personal utility as well as about acting according to the current norm (Azar, 2004, 2008). These modelling approaches have some common features that make them unsuitable for our theory. One such feature is the assumption that norms can be represented by equilibria, that is, stable states. There is no empirical evidence that stable states are achieved in the population, either for hygiene or violence norms. Game theoretic models are unsuitable for our theory also because they do not easily account for processes in which there is endogenous change of individuals’ preferences. Some models deal with exogenous change of preferences, in that they study how equilibria depend on different distributions of preferences in the population (Azar, 2008). Our theory, however, assumes that people’s preferences shift within the process, such that an individual’s interaction with others can alter what behaviour they prefer.

For these reasons, we instead need a model that focuses on the process by which frequencies of behaviour change over repeated interactions, in which each interaction can change the individual’s behavioural type. Such models have previously been used to model cultural evolution as a competition between cultural variants that differ in how good they are at spreading and at being retained (Strimling et al., 2009). Here we use a similar model to capture the role of sanctions in enabling a norm shift.
A model of norm shift through the use of sanctions

Consider a population of size \( N \), and consider a situation where individuals in the population can choose between a stricter and a looser behaviour. Let \( q_t \) denote the proportion of the population that uses the strict behaviour at time \( t \). We assume that the others use the loose behaviour. We shall examine when a norm shift, from a predominantly loose population (\( q_t \) close to 0) to a predominantly strict population (\( q_t \) close to 1), can arise through an asymmetry in the use of sanctions. The process by which people change their behaviour is modelled as a sequence of interaction events. Each interaction event consists of three steps:

1. Step 1. Two individuals observe each other’s behaviour.
2. Step 2. In case their behaviours differ, each individual may sanction the other one.
3. Step 3. An individual who was sanctioned may switch to the other behaviour.

The important parameters to formalise are the sanctioning probabilities in step 2 and the switching probabilities in step 3. Exactly how pairs of individuals are matched up in step 1 is not of great importance, as long as it sometimes happens that individuals with different behaviours are matched.

**Sanctioning probabilities.** We expect there are a multitude of factors creating individual differences in the propensity to use sanctions. For the purpose of this model, however, only a systematic difference between strict and loose individuals is relevant. Thus, let \( P_l \) denote the probability that a loose individual will sanction an observed strict individual. This probability may be frequency dependent. Specifically, we assume that uncommon behaviours may be more likely to be sanctioned. Assuming a linear effect of frequency, we then have

\[
P_l = b - c \times q_t, \quad \text{for some parameters satisfying } 0 \leq c \leq b \leq 1.
\]  

(1)

Similarly, let \( P_s \) denote the probability that a strict individual will sanction an observed loose individual. Our theory assumes an asymmetry in the form of a higher propensity for sanctioning against loose behaviours than against strict behaviours. This asymmetry can be modelled by the inclusion of an additional term:

\[
P_s = s + b - c \times (1 - q_t), \quad \text{for some parameter } 0 \leq s \leq 1 - b.
\]  

(2)

Note that the expression involves \( (1 - q_t) \) in place of \( q_t \), as it here represents the frequency of the loose behaviour in the population.

**Switching probabilities.** Again, we consider only such individual differences in switching probabilities that are systematic between strict and loose individuals. Let \( U_l \) denote
the probability of a strict individual switching to loose when sanctioned. Similarly, \( U_s \) denotes the probability of a loose individual switching to strict when sanctioned.

*The net effect of sanctions.* When a loose and a strict individual are matched up in step 1, the expected net effect on behaviour from steps 2 and 3 is given by \( E[q_{t+1} - q_t] = \Delta / N \), where \( \Delta \) is the difference between the probability of the loose individual both being sanctioned and switching to strict, and the probability of the strict individual both being sanctioned and switching to loose:

\[
\Delta = P_s \times U_s - P_l \times U_l .
\] (3)

Thus, depending on whether \( \Delta \) is positive or negative, the strict behaviour will tend to increase or decrease in frequency in the population. Plugging equations (1) and (2) into (3) we obtain

\[
\Delta = (s + b - c \times (1 - q_t)) \times U_s - (b - c \times q_t) \times U_l ,
\]

which is positive when

\[
s > c + b \times (U_l / U_s - 1) - c \times q_t \times (U_l / U_s + 1).
\] (4)

*When will a predominantly loose population shift towards strictness?* If inequality (4) is satisfied for a predominantly loose population \( (q_t \) close to 0), then it will remain satisfied as the population grows stricter. Thus, we obtain the condition for when a predominantly loose population will evolve towards strictness by letting \( q_t \) tend to 0 in (4), yielding the inequality:

\[
s > c + b \times (U_l / U_s - 1) .
\] (5)

In case there is no systematic difference in switching probabilities (i.e., \( U_l = U_s \)), the condition simplifies to \( s > c \), that is, the value of the asymmetry parameter must be greater than the frequency. If switching to strict is more likely than switching to loose (i.e., \( U_l < U_s \)), then it is even easier for the strict behaviour to take over in the population.

\( U_l \) and \( U_s \) can be interpreted as a reverse measure of the strength of the currently held preference, as an individual who finds the other behaviour particularly bad is unlikely to switch. These parameters can be used to account for exogenous processes. For instance, a technological change that makes it easier to adopt strict behaviour could be represented as an increase in the value of \( U_s \) relative to \( U_l \). Similarly, a societal collapse that makes it harder to uphold the stricter behaviour could be represented as an increase in \( U_l \) relative to \( U_s \), which could lead to a shift towards looser behaviour.

Figure 1 (left panel) illustrates how the norm changes over time in a simulation where step 1 is implemented such that the two individuals of each interaction event are drawn uniformly at random from the entire population.
It is straightforward to extend the model to demonstrate how the same process may make the norm change through a sequence of ever stricter behaviours. Consider a set of behaviours that can be ordered according to how strict they are (e.g., spit on the floor, spit in spitoon, spitting outdoors only, etc.). For each behaviour $i$ in the set, assume that with probability $U_i$ an agent who is punished for that behaviour will change to the next behaviour in the sequence, in the direction towards the behaviour of the punisher. As illustrated in Figure 1 (right panel), an initial loose norm may then be replaced by a stricter norm, which is in turn replaced by an even stricter norm.

The purpose of the model is not to provide an exact description of the world but to provide a transparent argument supporting a general qualitative conclusion: The presence of a sufficiently strong asymmetry in the willingness to use social sanctions can drive a process of directional norm change.

**Figure 1. Simulations of norm change from looser to stricter behaviours.** Left panel: With two behaviours and parameter values $b=0$, $s=0.4$, $c=0.2$ and $U_1=U_2=0.5$, the stricter behaviour spreads at the expense of the looser behaviour. Right panel: With three behaviours and parameter values $b=0$, $s=0.6$, $c=0.2$ and $U_i=0.5$ for all $i$, the middle behaviour first spreads at the expense of the loosest behaviour, whereupon the strictest behaviour spreads at the expense of the middle behaviour.

**Study 1**

Our first study investigates the key assumption of asymmetry: Given that some people prefer a looser behaviour and some prefer a stricter behaviour (in the domains of violence and
hygiene), is the propensity to sanction the disliked behaviour greater among those who prefer the stricter behaviour than among those who prefer the looser behaviour?

**Participants**

Three hundred and two US participants (mean age 30 years, 67% male) were recruited from Amazon Mechanical Turk.

**Method**

Participants were first asked what they prefer other people do in six different contexts. There were three questions regarding hygiene: handwashing before eating, providing fresh linen for guests, and sneezing into a tissue. There were also three questions regarding violence: smacking a misbehaving child, punching someone who insults your mother, and shooting an intruder. The full text of each item and the response options can be seen in the supplementary material. To give one example, one of the hygiene questions read: “Some people provide freshly washed linen for visiting guests; other people do not think this is necessary. What would you prefer other people provide?” Participants responded by clicking on a sliding scale with 100 demarcations, anchored by “usual bedsheets” and “freshly washed bedsheets”.

We aimed for questions where we would find substantial support for both positions. The frequency of support for the stricter behaviour ranged between 42% (in the shooting context) to 89% (the hand-washing context).

Which follow-up questions participants received depended on whether they preferred the looser of stricter behaviour. Follow-up questions asked participants about their inclination for three different sanctions of individuals exhibiting their non-preferred behaviour: reprimands (“I would tell her I disapproved of her behaviour”), gossip (“I would tell other people about her behaviour and that I disapproved”), and avoidance (“I would spend less time with her in future”). Participants again responded by clicking on a 0-100 scale anchored with ‘strongly disagree’ and ‘strongly agree’. For each target behaviour, the three sanction scores were averaged to a combined propensity to punish score with good internal consistency (Cronbach's alpha ≥ .78 for all target behaviours).

**Result**

Our research question is whether people who prefer stricter norms regarding violence and hygiene are more inclined to use sanctions than people who prefer more violent behaviour and more relaxed hygiene norms. In each of the six contexts we studied,
participants who preferred the “stricter” behaviour showed greater propensity to punish than those who preferred the “looser” behaviour (smacking a child: t(241.56) = 11.70, p < .001; hitting an insulter: t(300) = 5.76, p < .001; shooting an intruder: t(300) = 6.67, p < .001; handwashing: t(300) = 2.01, p = .05; providing fresh linen: t(300) = 2.15, p = .03; sneezing into tissue: t(149.14) = 3.88, p < .001). These results are illustrated in Figure 2. Descriptive statistics can be found in the supplementary materials.

![Figure 2. Propensity to punish non-preferred behaviour.](image)

There was a greater propensity to punish non-preferred behaviour among those who prefer the stricter behaviour than among those who prefer the looser behaviour. Error bars indicate 95% confidence intervals.

Previous research has found that some individuals are generally more inclined to use punishment than others (Eriksson et al., 2014, 2015). Consistent with such a personality effect, we found punishment scores across behaviours to be highly inter-correlated, suggesting a common underlying factor (Cronbach's alpha = .78). To ensure that the difference in punishment of strict and loose behaviours is not due to this personality effect we checked that the difference remained also when punishment scores were normalized through division by the sum of all the participant’s punishment scores and it was not (smacking a
child: $t(218) = 9.67, p < .001$; hitting an insulter: $t(295) = 5.83, p < .001$; shooting an intruder: $t(295) = 5.68, p < .001$; handwashing: $t(295) = 2.86, p = .005$; providing fresh linen: $t(295) = 3.95, p < .001$; sneezing into tissue: $t(295) = 4.00, p < .001$.

**Discussion**

A key assumption in the proposed theory is that the inclination to punish behaviour that you do not prefer yourself should be stronger among those who prefer a less violent or a more hygienic behaviour than among those who prefer a more violent or a less hygienic behaviour. In this study we found this assumption to be supported in each of six different contexts. The results for violence contexts are particularly noteworthy as individuals who are in favour of shooting an intruder orspanking a child might be expected to be generally pro-punishment, which would work against our hypothesis.

**Study 2**

Another key assumption of our theory is that the way people prefer others to act is also how they act themselves. Moreover, our theory says that for norm shifts to occur the tendency to punish rare behaviours more than common behaviours must not be too strong. We designed a second study to investigate these assumptions, in addition to replicating the fundamental asymmetry that was found in the first study.

**Participants**

Two hundred and sixteen US participants (mean age 31 years, 53% male) were recruited from Amazon Mechanical Turk.

**Methods**

The study was conducted in exactly the same way as Study 1, only it included two additional questions. The participants were asked (1) to estimate what percentage of people engages in the non-preferred behaviour, and (2) if they themselves behave in the way they had stated a preference for in others (e.g., “I always wash my hands before eating”), with responses on a sliding scale from 0 to 100. For the latter question, the response scale was anchored by “strongly disagree” and “strongly agree”.
Results

Using the same analysis as in Study 1 we first replicated our main effect that those who prefer the stricter behaviour are more likely to punish the looser behaviour than vice versa (see supplementary material). We then investigated the model assumption that individuals should exhibit the same behaviour that they prefer in others. Indeed, the correlation between preference and own behaviour was strongly positive for all contexts (insult: $r=.52$, $p < .001$; shoot: $r=.76$, $p < .001$; child: $r=.73$, $p < .001$; linen: $r=.47$, $p < .001$; hand-wash: $r=.52$, $p < .001$; sneeze: $r=.19$, $p=.004$).

Our theoretical model accounts for the possibility that people’s willingness to punish will depend on the perceived frequency of the target behaviour. No strong support for such frequency-dependent sanctions was found in this study. Correlations between punishment scores and perceived frequency was usually in the expected direction but rather weak (insults: $r=.21$, $p = .002$; shoot: $r=.11$, $p =.12$; child: $r=.11$, $p =.12$; linen: $r=.00$, $p =.98$; hand-wash: $r=.05$, $p =.47$; sneeze: $r=-.04$, $p=.58$).

Discussion

In addition to replicating Study 1, the second study supported the assumption that people tend to behave as they prefer others to behave. Moreover, we found little effect on punishment propensity of how common a behaviour was perceived to be. In terms of our model, this suggests that the value of the $c$ parameter should be low, which means that there is little conformity-based resistance to norm shifts.

Study 3

The first two studies provided evidence for an asymmetry in the amount of social sanctions that are directed towards a strict vs. a non-strict behaviour. According to our theory, this asymmetry should arise because of an inherent difference in the threat perceived from loose and strict behaviours in these domains. To test this we constructed a threat index based on four questions about the participant’s non-preferred behaviour. To avoid respondent fatigue when adding four new questions for each scenario, we went down from six to four scenarios. Specifically, we retained the three scenarios that had the most even strict/loose distribution in the previous studies (sneezing into a tissue, smacking a misbehaving child, shooting an intruder) and added a new hygiene scenario (spitting in the kitchen sink). See the supplementary material for exact wordings.
Participants

Three hundred and thirty US participants (mean age 35 years, 55% male) were recruited from Amazon Mechanical Turk.

Methods

The study was modelled on the previous studies. The question whether the participant prefers the strict or the loose behaviour was given as a binary choice between the two behaviours (whereas the previous studies used a continuous scale that was recoded to binary). For their non-preferred behaviour, participants were then asked what percentage of people engages in it and how they would punish people who engage in it, using the same three punishment questions and the same response scale as in the previous studies.

In order to examine the extent to which participants felt threatened by their non-preferred behaviour, we asked whether it harms others, whether it harms you, whether it invokes a sense of disgust, and whether it invokes a sense of threat. Responses were given on a sliding scale anchored with “no harm” / “much harm”, “no disgust” / “much disgust”, and “no threat” / “much threat”. The four responses for each scenario were averaged into a threat index with good internal consistency (Cronbach’s alpha > .8 for each scenario).

Finally, in order to examine the relationship between behavioural norms and social power or class, we asked participants to report their level of education (eleven alternatives ranging from no completed schooling to Doctorate degree), household income (ranging from $10 000 to more than $150 000 a year, and class (choice between lower, working, middle and upper class).

![Graph](image)

**Figure 3. Propensity to punish and feelings of threat from non-preferred behaviour.** There were both a greater propensity to punish non-preferred behaviour (left panel) and greater feelings of threat (right panel) among those who prefer the stricter
behaviour than among those who prefer the looser behaviour. Error bars indicate 95% confidence intervals.

**Results**

Figure 3 illustrates that both punishment scores and feelings of threat were consistently higher among those who prefer the stricter behaviour than among those who prefer the looser behaviour. Our hypothesis is that the difference in punishment of strict and loose behaviour is mediated by feelings of threat. We performed mediation analyses using the basic mediation model of the PROCESS macro in SPSS (Hayes, 2013) with 5000 bootstrapped samples. This macro calculates a series of regression coefficients, \( a, b, c', \) and \( c \). Coefficient \( a \) represents the path from the independent variable \( X \) (strict/loose preference) to the mediator \( M \) (threat); \( b \) represents the path from \( M \) to the dependent variable \( Y \) (punishment); \( c' \) represents the direct effect of \( X \) on \( Y \); and \( c \) represents the total effect of \( X \) on \( Y \). The macro also calculates a bootstrapped bias-corrected 95% confidence interval of the indirect effect of \( X \) on \( Y \) through \( M \) (the product \( ab \)) and the ratio of the indirect effect to the total effect (denoted by \( P_M \)).

Table 1 reports these estimates for every scenario, showing that the indirect effect through the mediator was consistently significant and accounted for most, or all, of the total effect. This is an indication that stronger feelings of threat is the reason people with preference for the stricter behaviour sanction more.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>( a )</th>
<th>( B )</th>
<th>( c' )</th>
<th>( c )</th>
<th>( ab ) [BCa CI]</th>
<th>( P_M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>12.27***</td>
<td>0.75***</td>
<td>4.38†</td>
<td>16.56***</td>
<td>12.18 [8.29, 16.12]</td>
<td>0.74</td>
</tr>
<tr>
<td>Shoot</td>
<td>12.69***</td>
<td>0.73***</td>
<td>0.87</td>
<td>10.15***</td>
<td>9.28 [4.68, 14.60]</td>
<td>0.91</td>
</tr>
<tr>
<td>Sneezing</td>
<td>16.40***</td>
<td>0.61***</td>
<td>−5.26†</td>
<td>4.70***</td>
<td>9.97 [5.02, 14.5]</td>
<td>2.12</td>
</tr>
<tr>
<td>Spit</td>
<td>16.35***</td>
<td>0.85***</td>
<td>9.98***</td>
<td>23.85***</td>
<td>13.87 [8.50, 19.21]</td>
<td>0.58</td>
</tr>
</tbody>
</table>

†: \( p < .1 \), *: \( p < .05 \), **: \( p < .01 \), ***: \( p < .001 \)

None of the measures of social power and class showed a significant (after Bonferroni correction) relation to behavioural preferences. Thus, we found no evidence for social power and class being important factors for understanding current shifts in hygiene and violence norms.
Discussion

In this paper we have proposed a pathway for how social norms about violence and hygiene may become increasingly stricter over time—what Norbert Elias referred to as the civilizing process. The proposed pathway works in many small steps. In each step, a current norm is challenged by a somewhat stricter norm, that is, a behavioural rule that puts further restriction on the use of violence or the exposure of bodily fluids. Our first key assumption is that the human mind tends to exhibit a domain-specific bias such that violence and exposure to bodily fluids are generally perceived as threatening. Accordingly, behaviours that are more restrictive with respect to violence and exposure to bodily fluids should generally be perceived as less threatening. Our second key assumption is that people tend to use social sanctions against behaviour they feel threatened by. Thus, a currently common behaviour may still be socially sanctioned by a small minority because it is perceived as more threatening than their preferred stricter behavioural variant. We propose that these social sanctions could drive norm shifts.

To gain better understanding of what causes shifts in hygiene and violence norms is not only of academic interest. As a case in point, it has been estimated that the vast majority of the human population does not wash hands with soap after contact with excreta, and that handwashing with soap would save approximately 300,000 lives per year (Freeman et al., 2014). A theoretical understanding of norm shifts could provide insight into which interventions are more likely to be successful. For instance, our theory suggests that norm shifts may be speeded up if the threatening aspect of undesirable behaviours is emphasized.

In support of our theory we presented two kinds of evidence. First, a mathematical model demonstrated that the cultural evolutionary process we propose is theoretically feasible. Second, a series of survey studies validated key parts of the theory. Individuals with stricter preferences reported greater willingness to sanction non-preferred behaviour, and this difference was mediated by looser behaviour evoking more feelings of threat. By combining mathematical modelling and survey data we have achieved a higher level of rigor than previous attempts at explaining the civilizing process. Still, our studies have limitations. In particular, we have not tested the plausible assumption that social sanctions may indeed cause behavioural change in the hygiene and violence domains. Another limitation is that we rely on survey data instead of independent observations of behaviour. A further limitation of our
work is that we have surveyed American participants only. These limitations may be addressed in future research.

Finally, let us discuss an assumption of our theory that so far has been tacit. The process we envision relies on society being stable, in the sense that individuals can maintain their current behaviour without making a greater effort in terms of time or resources. When society becomes unstable (e.g., due to war or natural disasters), it may be too effortful to maintain current standards of hygiene and violence behaviour and therefore lead to a temporary reverse process by which stricter norms gradually disappear until a strictness level which is practical, considering the circumstances, is achieved. The possibility of a reverse process was documented already by Elias in his account of Germany post World War One (Elias, 1996). Many similar examples can be found, such as the disappearance of Roman baths from Western Europe after the fall of Rome, or the increase in domestic violence in Iraq post the US-led invasion (Green & Ward, 2009).

As long as a society remains stable, however, our theory predicts that the civilizing process will continue. In a generation from now we should expect norms about hygiene and violence to have become even stricter. We should also expect directional change in any other domain where there is a similar systematic difference in social sanctions of non-preferred behaviour between those who prefer one behaviour and those who prefer another. We believe it may be fruitful for more researchers to adopt this perspective on norm change.

References


