From Peer Pressure to Biased Norms

Moti Michaeli & Daniel Spiro*

Abstract

This paper studies a coordination game between a continuum of players with heterogeneous tastes who perceive peer pressure when behaving differently from each other. It characterizes the conditions under which a social norm – a mode of behavior followed by many – exists in equilibrium and the patterns of norm compliance. The emergent norm may be biased compared to the average taste in society, yet endogenously upheld by the population. Strikingly, a biased norm will under some circumstances be more sustainable than a non-biased norm, which may explain the bias of various social and religious norms.

Keywords: Peer pressure, Social norm, Coordination.

JEL: D03, D70, Z10, Z12.

*Michaeli: Department of Economics, European University Institute, Italy. Email motimich@gmail.com. Spiro: Corresponding author, Department of Economics, University of Oslo, Norway. daniel.spiro@econ.uio.no, Tel +47 22855137, Fax +47 22855035. We wish to thank Jacopo Bizzotto, Martin Dufwenberg, Martin Dumav, Tore Ellingsen, Joan Esteban, Bård Harstad, Paul Klein, Gilat Levy, Charles Manski, Arie Michaeli, Kalle Moene, Andrew Oswald, Paolo Piacquadio, Andrew Postlewaite, Debraj Ray, Gerard Roland, Moses Shayo, Jörgen Weibull, two anonymous referees and seminar participants at George Washington University, Hebrew University, Oslo University, Tilburg University, BI, EUI and the ASREC, WEHIA and, NCBEE conference for valuable comments.

Coordination games with heterogeneous tastes, such as the Battle of the Sexes, are often used to represent the individual trade off between coordinating with others and following one's own taste. These games are particularly useful for studying social norms as, when there are many players, a common interpretation is that the point of coordination constitutes a social norm or convention (e.g., Schelling 1960, Lewis 1969, Granovetter 1978, Young 1993). This applies to many social settings such as dress codes, political declarations and cultural customs. In these situations, it seems reasonable to assume that the number of possible actions is large, that individuals in society have a large variety of tastes and that partial gains of coordination can be achieved also if two individuals behave partially, but not exactly, the same. The purpose of this paper is to study the existence of a social norm (i.e., coordination) in such circumstances and to study what norms can be upheld in equilibrium. In particular, we are interested in the sustainability of biased norms, i.e., norms that are misrepresentative of the private tastes in society.

For this purpose, we use a model of pairwise interaction between a large number (a continuum) of individuals who can choose actions from a continuum. All individuals differ in their private bliss points (or tastes), and the cost for an individual of deviating from her bliss point is increasing with the size of the deviation. At the same time, the individual feels peer pressure when deviating from the action of another individual. Hence, the individual gains by behaving similarly to others (i.e., coordinating) and this gain is increasing the more similar she behaves to each other person in society. This means that coordination with one person may imply miscoordination with another person. In this setup, whether a norm exists or not depends on whether many individuals choose to behave the same in equilibrium, despite having different tastes, despite having the option to choose actions from a continuum and despite the fact that partial gains of coordination are attained even without behaving exactly the same as others.[1]

[1]Naturally, analyzing existence of coordination between many players, with gains from partial coordination, rules out the usage of the common binary action model (e.g., Granovetter 1978, Kuran 1995, and Neary 2012). A setup similar to ours has been used to analyze other questions under a more restrictive set of functional form assumptions (Manski

As an illustration, consider a Muslim girl who needs to choose headwear when going to school. Her choice set contains at least seven forms of veiling (ranging from burqa to hijab, see BBC 2010 for an illustration) and of course a large number of non-veiling headwears (a scarf, a cap, no headwear, etc.). When choosing headwear, she will need to trade off how similar her headwear is to that of the other girls in class. For example, wearing a hijab would imply full coordination gains with a classmate wearing a hijab as well, partial coordination gains with a classmate wearing a burqa, a niqab, or a scarf, and possibly no coordination gains with a classmate with no headwear. At the same time, she needs to take into account her own private preference with respect to headwear. Each of the other girls is of course facing a similar problem. When all girls in class have made their choices, and none of them is inclined to change headwear, we have an equilibrium. If there exists a headwear on which (at least some) girls coordinate, despite their different tastes, we refer to it as a social norm. Otherwise, if every girl chooses a unique headwear, we say that no norm exists.[2]

The explicit modeling of pairwise interaction differentiates our paper from

and Mayshar, 2003; Kuran and Sandholm, 2008). Furthermore, it is important to note the difference between our line of modeling and models of status and effort (e.g., Clark and Oswald, 1998; Kandel and Lazear, 1992) or network externalities (see Jackson and Zenou, 2014, section 4.3). In these models there is agreement about what is right (to work hard or achieve status) but there is an individual effort cost of getting there. In our model, on the other hand, there is disagreement about the right action since tastes are heterogeneous, but individuals gain by coordinating. Hence, we are interested in what sociologists call a *descriptive* norm (something people do) while models of status and work effort have a *prescriptive* norm (something people should do). See Cialdini et al. (1991), Cialdini (2003) and Blumenthal et al. (2001) for a further discussion.

[2]In a recent paper, Carvalho (2012) studies veiling under peer pressure. In his paper, veiling entails a positive peer effect stemming from all religious types, regardless of their own veiling choices, and a negative peer effect from all secular types, regardless of *their* veiling choices. In contrast, in our model pairwise pressure stems from differences in the chosen *actions* of peers, which to us seems more natural for studying public expressions such as veiling. This approach is pursued by Carvalho in a subsequent paper (Carvalho 2014) which studies the integration of groups with different norms. However, there the gains from coordination are assumed to be binary (i.e., arise only when two individuals behave exactly the same). Hence, the relation between the gains from full coordination vs. partial coordination, which is shown in the current paper to be crucial for the existence of norms, cannot be investigated in that model.

a branch of the literature that assumes that a norm exists and equals the (weighted) average of what people do (see Glaeser and Scheinkman 2000 and Ozgur 2011 for reviews and Michaeli and Spiro 2015 for a recent paper). In this literature social pressure (or loss of miscoordination) is assumed to be the lowest when a person completely conforms to the norm and, importantly, this is independent of whether anyone else follows the norm. It could even be that what is considered a norm in that setup would actually be the maximal point of pressure in a model of pairwise interactions like in our paper.[3] Moreover, social pressure is after all a form of disutility. Hence, using a von Neumann-Morgenstern (vNM) approach seems the most natural way to aggregate pressure. To see this formally, denote the action of one individual by s and the action of another individual she interacts with by s'. When the individual interacts with many others, like in our model, the vNM aggregate pressure she feels when stating s is $E_{s'}\left[p\left(|s-s'|\right)\right]$, where p is the pairwise pressure between two individuals. This is not the same as the pressure she feels when interacting with a person who takes the average action in the population, $p\left(|s-E_{s'}\left[s'\right]|\right)$, as modelled in the previous literature just mentioned.[4]

We start by showing that existence of norms hinges on the curvature of pairwise pressure (i.e., on the gains from full coordination relative to the gains from partial coordination). When pairwise pressure is convex, a norm cannot exist in equilibrium. The crude intuition for this is that in this case, slight deviations from full coordination are inconsequential, so there is no need for a person to act exactly as her peers, hence the person will want to deviate toward her bliss point. Thus, when tastes are fully heterogeneous, then also on the

[3] As an easy example, consider a case where actions can be chosen in the whole range $[0,1]$ but half of the population chooses 0 while the other half chooses 1. Here, treating 0.5 as the social norm would not only be somewhat unconvincing from a descriptive point of view, but also, if pairwise pressure happens to be concave, contradictory to the nature of the norm as a pressure-minimizing choice, as 0.5 would be the point where aggregate pressure is maximized.

[4] In another branch of the literature individuals are punished for the private taste they are perceived to have and the punishment increases the more deviant this perceived taste is from an exogenous norm (e.g., Bernheim, 1994). This leads to a signaling model where some types try to hide their tastes by behaving similarly to others. In our model pressure is applied to actions of individuals.

aggregate level individuals do not act the same, which implies a norm does not exist. Previous papers with a model similar to ours (Manski and Mayshar 2003 and Kuran and Sandholm 2008) are nested in this case as they assume a quadratic pairwise pressure. In contrast, when pairwise pressure is concave, the marginal benefit from coordination is increasing as individuals approach each other, and hence equilibria with endogenous norms exist, provided that concerns for coordination are sufficiently strong.[5]

To see what the curvature of social pressure represents, consider the earlier headwear problem, where, on an axis of strictness, burqa is stricter than hijab, which itself is stricter than a scarf. Suppose now that a girl considers changing headwear from scarf to hijab. A concave pressure implies that this change will mainly reduce pressure arising from a girl already wearing hijab, while the reduction of pressure arising from a girl wearing burqa will be smaller. A convex pressure (e.g., quadratic costs) implies the opposite: when changing from scarf to hijab, the reduction of pressure arising from the girl wearing hijab is smaller than the reduction of pressure from the girl wearing burqa. We show that a concave pairwise pressure is necessary and sufficient for the existence of a norm when tastes are fully heterogeneous (as long as individuals care sufficiently about coordination).

Furthermore, we characterize what patterns of behavior are self-sustaining *and* imply norm existence in equilibrium. Two prototypical types of equilibria that sustain a social norm exist. These differ in the fundamental feature of who upholds the norm (i.e., which individuals coordinate). In the first type of equilibrium society, the norm is upheld by individuals with private tastes close to the norm. We call this an alienating society, as potential non-conformers are those with tastes far from the norm – in that sense they are alienated. In the second type of equilibrium society, those upholding the norm are individuals whose tastes are far from it. By conforming, these individuals unwillingly help to strengthen the norm, by making it more of a focal point. Meanwhile, those

[5]Showing this existence is not trivial, as most equilibria contain also non-conformers. Aggregating over the pairwise pressures is thus complex, as the non-conformers put pressure on others, with different tastes, to follow their non-conforming choices rather than the norm.

who only slightly disagree with the norm choose to follow their private tastes. We call this an inverting society, as actions are inverted relative to private tastes.[6]

As mentioned, our second research question is what norms (i.e., which points of coordination) a society can sustain. We are particularly interested in analyzing the emergence of a biased norm in society – a coordination point that is far from the average taste – and in understanding whether biasness makes the norm stronger or weaker. Biased norms are commonplace in social and political life. This has been documented in excessive drinking among college students (for a review see Borsari and Carey 2001), in attitudes toward alcohol prohibition (Robinson 1932, Cohen 2001) and toward racial segregation (O'Gorman 1975, Fields and Schuman 1976, Miller and Prentice 1994), among religious communities (Schank 1932) and vegetarians (Kitts 2003), in honor cultures and honor killings (Colson 1975, Gladwell 2000, Milgram 1992, Wilson and Kelling 1982, Centola et al. 2005), and in norms of violence (Cohen et al. 1996, Vandello and Cohen 2000).

We find that the difference between the alienating and inverting societies in terms of who upholds the norm has implications for the existence of biased norms. First note that the more biased the norm is, the larger is the overall misalignment between the tastes of individuals and the norm – there are more private bliss points far from the norm. Hence, in the alienating society, where norm-deviators are those who strongly disagree with the norm, a biased norm will be less sustainable than a central norm. Conversely, in inverting societies, the norm draws its strength from those who privately disagree with it the most as they are the ones who adhere to it. Thus, a biased norm will have more

[6] The patterns of individual behavior that characterize the alienating and inverting societies can emerge also in a model with an exogenous norm (see Michaeli and Spiro 2015). However, when the norm is exogenous, the patterns of behavior do not affect the strength of the norm and, more generally, are not required to be self-enforcing. Hence, such reduced-form modeling cannot be used to investigate the two main research questions of the current paper, about the conditions for the existence of a norm and about the strength and sustainability of biased norms. Thus, we view the current paper as providing the microfoundations for that earlier paper and more broadly for the strand of the literature that simply assumes a norm exists (surveyed in Glaeser and Scheinkman 2000 and in Ozgur 2011).

adherence and will survive under *weaker* conditions than a non-biased norm. It will also be dynamically more stable. Our model thus predicts that inverting societies will be particularly well suited for upholding biased norms as those exemplified. It also points at the potential history-dependence of societies: If a group of individuals, possibly a long time ago, established a particular norm, this norm can be expected to persist long after the group is gone and private tastes have changed.

The next section presents the model and analyzes the existence of a norm in equilibrium. Sections 2 and 3 analyze the strength of biased norms and the patterns of norm conformity in the alienating and inverting societies respectively. Section 4 concludes. The appendix presents some auxiliary results and all formal proofs.

1 A model of peer pressure and single-norm equilibria

Consider a society with a continuum of individuals, each having a different bliss point $t \in T \subseteq \mathbb{R}$, i.e., some private preference, ideology or opinion, referred to also as the individual's type. One can think of t as a position on a political scale. Let $f(t)$ denote a continuous probability density function of types. Each individual chooses a publicly observable action (or stance), denoted by $s \in \mathbb{R}$. The inner disutility of an individual choosing action s in public, $D(|t-s|)$, increases in the distance between that action and the individual's type, representing the cognitive dissonance or displeasure felt by her.

In addition, the individual feels social pressure. When choosing action s, the social pressure arising from another individual choosing s' is

$$p = p(|s - s'|)$$

which is increasing in the distance between s and s' (p can also be thought of as the loss from miscoordination when two individuals behave differently). Such pressure arises between each pair of individuals, hence we refer to p as pairwise pressure. This means that, given the actions of all types in society,

$s' : T \to \mathbb{R}$, the aggregate pressure felt by an individual taking action s is given by

$$P(s; s') \equiv E_{s'(\tau)}[p(|s - s'(\tau)|)] = \int_{\tau \in T} p(|s - s'(\tau)|) f(\tau) d\tau. \quad (1)$$

This formulation captures the essence of the coordination problem: when a person takes an action s which is similar to some s'_1 taken by another person, it may imply miscoordination with another person who takes s'_2. Hence, the individual has to trade off conformity (coordination) between different individuals.[7]

The objective of the individual is to choose an action s which minimizes the total loss that arises from the inner disutility and the aggregate social pressure

$$L(s; t, s') \equiv D(|t - s|) + P(s; s'). \quad (2)$$

All individuals move simultaneously and hence take the actions of others as given. An equilibrium is a mapping from the type space to the action space, $s^* : T \to \mathbb{R}$, such that, for each $t \in T$

$$s^*(t) = \arg\min_s \{D(|t - s|) + P(s; s^*)\}. \quad (3)$$

That is, each individual optimally chooses her action $s^*(t)$, given the choices of all others, such that the chosen actions recreate the ones taken as given by the individual. Being interested in studying the emergence of a norm in society and in the conditions under which this norm may be biased, we first define what we mean by a norm.

Definition 1 *A social norm is an action $\bar{s}$ taken by a non-zero mass of agents. If the social norm is not equal to the average type in society, the norm is said to be biased.*

[7] There are two ways to interpret equation (1). Either s is a statement or action made in public, implying that $P(s; s')$ is an actual pressure felt when choosing s. Or, alternatively, $P(s; s')$ is the expected pressure felt when not knowing whom one is about to interact with under random pairwise matching (as, for instance, in Kuran and Sandholm 2008).

We will analyze the existence and characteristics of the following type of equilibrium.

Definition 2 *A single-norm equilibrium is an equilibrium with one and only one social norm.*

Note that the continuity of $f(t)$ excludes cases where a norm exists simply because it represents the private preference of a mass of people. The single-norm equilibrium is not the only type of equilibrium that may exist, as there may be more than one norm in equilibrium. However, we confine our analysis to the single-norm equilibrium and to inexistence of a norm in equilibrium.

Wherever applicable, we will perform the analysis for power functions of the form

$$D = |s - t|^{\alpha}, \tag{4}$$

$$p = K |s - s'|^{\beta}, \tag{5}$$

where $\alpha > 0$ and $\beta > 0$ represent the curvature of cognitive dissonance and pairwise pressure respectively. K represents the relative weight of the peer pressure, and so captures the extent to which individuals care about social pressure (or coordination). In our analysis, α, β and K are identical across individuals in a given society. The heterogeneity is in individual tastes. We let the distribution of types be uniform: $t \sim U(-1, 1)$. This of course makes the problem more tractable but it also ensures that a biased norm, following the above definition, does not arise as an artefact of the distribution of types being non-symmetric.[8] With the uniform distribution, following (1) and (5), the aggregate pressure function becomes

$$P(s; s') \equiv \frac{1}{2} K \int_{-1}^{1} |s - s'(\tau)|^{\beta} \, d\tau.$$

[8] The results of the first proposition, about the necessary conditions for existence and the properties of conformity in single-norm equilibria, hold generally for any continuous distribution of types. We illustrate and discuss in Appendix A how the later results (about biasness of the norm) translate to other distributions of types.

We start the analysis by characterizing what kind of peer pressure is needed for the existence of single-norm equilibria and by characterizing the main properties of these equilibria when they do exist.[9]

Proposition 1 *For any $\alpha > 0$:*

1. *If $\beta > 1$, there exists no single-norm equilibrium.*

2. *If $\beta \leq 1$ and $\beta < \alpha$, single-norm equilibria exist, provided that K is sufficiently large. In all single-norm equilibria, the types closest to the norm fully conform and hence non-conformers, if they exist, are only types sufficiently far from the norm.*

3. *If $\beta \leq 1$ and $\alpha < \beta$, single-norm equilibria exist, provided that K is sufficiently large. In all single-norm equilibria, the types closest to the norm follow their hearts and hence only types sufficiently far from the norm fully conform.*

The proof of the proposition appears in the appendix and the results are depicted in Figure 1. As can be seen in the figure, the proposition spans the entire parameter space. The two main results of the proposition are that single-norm equilibria can exist if and only if $\beta \leq 1$ and that there are two mutually exclusive types of single-norm equilibria depending on whether α or β is greater.[10]

The case of $\beta > 1$ is represented by the upper region in the figure. Loosely speaking, it portrays a society where individuals are liberal in how they perceive others' opinions, in the sense that tension (p) arises in between two individuals only when they choose distant actions. For two such liberal individuals, there will never be a reason to take the same action (unless they happen to privately agree). Hence, also at the aggregate level, there will not

[9] In the following proposition, and throughout the paper, "following one's heart" means $s(t) = t$.

[10] In the limit case, $\alpha = \beta$, the condition for existence is the same (K has to be sufficiently large), but both types of equilibria may arise.

be any one mode of behavior that many follow, and this in fact holds generally for any continuous distribution of types.[11] Note also that $\beta > 1$ nests the special case of a double-quadratic function as has been analyzed by Manski and Mayshar (2003), Kuran and Sandholm (2008) and Acemoglu and Jackson (2014). Their analyses have different focus than ours, but Proposition 1 implies that no norm can be sustained in the double-quadratic case.

As an illustration of the workings of a society where $\beta \leq 1$, consider two individuals of types t_1 and t_2 such that $t_1 < t_2$. Suppose that both individuals start by following their hearts but they consider compromising on an intermediate action $\tilde{s} \in (t_1, t_2)$. If t_1 changes her chosen action from $s(t_1) = t_1$ to $s(t_1) = \tilde{s}$, she makes it easier for type t_2 to choose $\tilde{s}$ too (because $p(|\tilde{s} - s(t_1)|)$ decreases from $p(|\tilde{s} - t_1|)$ to $p(|\tilde{s} - \tilde{s}|)$). However, at the same time, t_1 also makes it easier for t_2 to follow her heart (because $p(|t_2 - s(t_1)|)$ decreases from $p(|t_2 - t_1|)$ to $p(|t_2 - \tilde{s}|)$). If peer pressure is concave, the decrease of $p(|\tilde{s} - s(t_1)|)$ is greater than the decrease of $p(|t_2 - s(t_1)|)$, which incentivizes t_2 to choose $\tilde{s}$ as well – conformity by leftists helps conform rightists. This description suggests that concave peer pressure can facilitate clustering. However, in order to create a norm (i.e., coordination) in a society with many individuals, it is the aggregate pressure P that has to be concave around the norm, thus incentivizing individuals to fully conform. Indeed, if the pairwise pressure p is concave and some individuals do cluster at a point $\bar{s}$, then the aggregate pressure P will also be concave around $\bar{s}$, which facilitates the clustering in the first place.[12] For individuals in this cluster not to deviate from it, it has to contain sufficiently many of them. Hence, K has to be sufficiently large to make full conformity worthwhile for many.

A different angle on what the curvature of p means is attained by consid-

[11] In fact, the proof of the proposition also rules out the existence of multiple norms in equilibrium when $\beta > 1$. The exception is the special case of $\alpha = 1$, which cannot sustain a single norm equilibrium but where there can potentially exist more than one norm.

[12] Note that a concave p is not a sufficient condition for a concave P as, depending on the distribution of actions, a concave p may also imply a convex P. For instance, in the appendix (Lemma 5) we show that if all types follow their hearts ($s(t) = t$), then a convex P would arise independently of the curvature of p. Hence, it is the concave p along with clustering that creates the concave P.

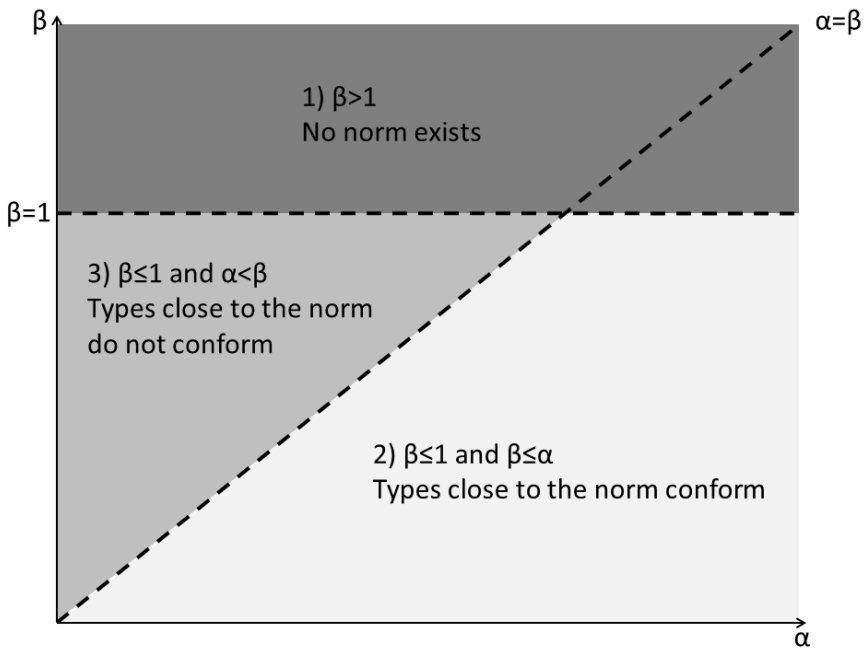

Figure 1: Graphical illustration of Proposition 1: Existence and main properties of single-norm equilibria depending on values of α and β.

ering support of sport teams. Suppose one person publicly supports team s_1, another person publicly supports team s_2, while a third person publicly supports team s_3, where the closeness between the teams is given by $s_1 < s_2 < s_3$. Now suppose the third person changes his support from s_3 to s_2. A concave p implies that this change of support reduces pressure from the person also supporting s_2 more than it reduces pressure from the person supporting s_1. The proposition says that this concavity of p is what enables the existence of social norms. A convex p, on the other hand, would mean that the change of support from s_3 to s_2 mainly reduces pressure from the person supporting s_1.

As is further expressed in statements 2 and 3, when $\beta \leq 1$ there exist two qualitatively different kinds of single-norm equilibria. These equilibria are mutually exclusive as each exists for a different set of parameters and no other kinds of single-norm equilibrium exist. This result is in fact independent of the distribution of types. The two equilibria are treated in great detail in the next two sections and they differ in whether or not they induce conformity by

types close to the norm. This in turn depends on whether β or α is smaller, i.e., whether p or D is more concave. To see why, note that a concave p, which is the prerequisite for a norm to exist, implies that the properties of the aggregate pressure (P) at some point $\tilde{s}$ are mainly determined by types who take stances close to $\tilde{s}$. Hence, close to the norm, P is mainly determined by the norm conformers and thus P has the same curvature as p. Now consider a type very close to the norm. Since P is very steep near the norm when p is concave, this person will essentially compare the two corner solutions, that is, compare the cost of speaking her mind with the cost of following the norm. This comparison boils down to whether P or D is steeper for small deviations from the norm and blisspoint respectively, as determined by the relative sizes of β and α.[13]

When $\beta < \alpha$, P is steeper than D for types near the norm, and deviating from the norm becomes more painful than deviating from the bliss point. Hence, in the equilibrium (or society) that arises, types close to the norm fully conform. It is represented by the lower right region in Figure 1. As is further expressed by Proposition 1 and will be shown in Section 2, the fact that those closest to the norm fully conform implies that, if anyone does not conform, these must be types far from the norm. In fact, they greatly deviate from the norm and are in a sense alienated. Hence we call this an *alienating* society.

The opposite case, represented by the lower left region in Figure 1, is where $\alpha < \beta$. Here D is steeper than P near the norm and hence for types close to the norm it is more costly to deviate from the bliss point than to endure the pressure when deviating from the norm. Hence, this society *does not* induce conformity by types close to the norm. Instead, these types follow their hearts. Since in a single-norm equilibrium somebody has to uphold the norm, the fact that those close to the norm do not conform implies that there is a cutoff beyond which some types do conform. That is, and perhaps surprisingly, individuals who dislike the norm are the ones upholding it. We call this an *inverting* society since the private tastes and the public actions are inverted

[13] Analytically, $(t-s)^{\alpha}$ is steeper and thus larger than $K(s-\bar{s})^{\beta}$ when the arguments approach zero if and only if $\alpha < \beta$ (and K is finite).

between those close to the norm and those far from it. The full intuition for this pattern of conformity and further properties of this case will be explained in Section 3, where we study a special case in more detail.

As parts 2 and 3 of Proposition 1 state, there is potential for multiple single-norm equilibria for each set of parameters that allows their existence. However, these equilibria differ only in the location of the norm, and share the same basic society characteristics regarding who upholds the norm. The possible location of the norm, and how this location affects the sustainability of the norm, is the second main question of this paper. Hence, the next two sections concentrate on a comparative-statics analysis of the different equilibria (different norm locations), showing whether equilibria with biased norms require larger or smaller weight of peer pressure (K). Performing this analysis for any combination of α and β is very difficult. We therefore look at two simplified special cases which capture the essential properties of our two society types. As expressed in Proposition 1, which of the two kinds of societies will emerge crucially hinges on which of α and β is smaller. Hence, we will let the smaller of the two parameters approach zero. I.e., the case of β being smaller than 1 and smaller than α will be illustrated by letting β approach 0, implying that p is a step function. The case of $\alpha < \beta \leq 1$ will be illustrated by letting α approach 0, implying that D is a step function.

2 Alienating societies

The purpose of this section is to further examine the case represented by point 2 of Proposition 1 and by the lower right region in Figure 1. The fundamental characteristic of this case is that p is concave and, in particular, more concave than D. To capture this very concave pairwise pressure, we let p be a step function,

$$p\left(|s-s'|\right) = \begin{cases} K \text{ if } |s-s'| \neq 0 \\ 0 \text{ if } |s-s'| = 0 \end{cases} \tag{6}$$

while $D = |s-t|^{\alpha}$ for some $\alpha > 0$. A first useful result thus follows.

Lemma 1 *Suppose that p is given by (6), D is given by (4) with $\alpha > 0$ and a*

single norm $\bar{s}$ exists and is followed by a share x of the population. Define

$$y \equiv (xK)^{1/\alpha} . \tag{7}$$

Then for an individual of type t, the optimal action is given by

$$s^*(t) = \begin{cases} \bar{s} \text{ if } |t - \bar{s}| \leq y \\ t \text{ otherwise} \end{cases} . \tag{8}$$

This is a partial equilibrium result showing which action an individual will take given the existence of a norm $\bar{s}$. Since (6) implies that the only way to avoid being pressured by someone is to fully agree with her, the only way to lower *aggregate* pressure to any meaningful extent is by choosing a mode of behavior followed by many. When a single norm exists, this can be achieved only by following this norm. Furthermore, since all actions except for following the norm yield the same pressure, the only effect of the pressure is in determining how unpleasant it feels to take any of these actions relative to following the norm. This is determined by the share of norm followers x:

$$P = \begin{cases} K \text{ if } s \neq \bar{s} \\ (1-x)\,K \text{ if } s = \bar{s} \end{cases} . \tag{9}$$

Given such a social pressure function P, the only sensible thing to do for an individual is to either follow the norm (thereby lowering pressure) or to follow her heart (thereby not feeling cognitive dissonance). Any other choice will induce some cognitive dissonance while not reducing social pressure. Moreover, two individuals of different types face the same reduction in pressure when following the norm, but differ in the cognitive dissonance that accompanies it. Thus follows the behavior depicted by the lemma – a type close to the norm will conform to it while a type far from the norm will follow her heart, in a sense being *alienated*. The parameter y in Lemma 1 captures the distance between the norm and the type who is indifferent between the two corner solutions. That the norm will be upheld by those closest to it thus echoes the result for the more general case of β smaller than 1 and smaller than α, as

stated in point 2 of Proposition 1.

The previous lemma implies that if we assume that individuals divide into two distinctive kinds – those who follow the norm and those who follow their hearts – then that same qualitative division is obtained after inducing the individual choices. This hints at the possibility of an equilibrium. However, the actual existence of an equilibrium hinges on the share of norm followers implied by (8) being equal to the value of x that is assumed in the lemma. In order to establish this relation, the following lemma presents the share of norm followers given the individual optimization in (8).

Lemma 2 *Suppose $s^*(t)$ is according to (8), for a given value of y. Then the share of individuals following the norm $\bar{s}$ is*

$$x = \begin{cases} y \text{ if } y \leq 1 - |\bar{s}| \\ \frac{y+1-|\bar{s}|}{2} \text{ if } 1 - |\bar{s}| < y < 1 + |\bar{s}| \\ 1 \text{ if } y \geq 1 + |\bar{s}| \end{cases} . \tag{10}$$

Furthermore, x is increasing in y and decreasing in $|\bar{s}|$.

This lemma presents the share of the population (x) that follows the norm as a function of y (the distance between the norm and the indifferent type). It builds on the previous result that those close to the norm fully conform while those far from it follow their hearts. This directly implies that the further from the norm the indifferent type is, the greater is the number of individuals conforming to the norm. The use of a uniform distribution at $[-1, 1]$ implies that when $\bar{s} = 0$ we automatically get that $x = y$, but when $\bar{s} \neq 0$ the mapping from y to x is not one-to-one for every y, as expressed in (10).

A static equilibrium of the model is essentially a fixed point, defined by a triplet $(x, y, \bar{s})$ that satisfies Lemma 1 and Lemma 2 simultaneously. The conditions for the existence of such an equilibrium are presented in the following proposition.

Proposition 2 *Suppose that pairwise pressure is according to (6) and D is given by (4) with $\alpha > 0$. Then:*

1. *For each value of* $\bar{s} \in [-1, 1]$ *there exists a single-norm equilibrium with* $\bar{s}$ *as the norm if and only if* K *is sufficiently large.*

2. *Denote by* $K_{\min}(|\bar{s}|)$ *the infimum value of* K *that supports a single-norm equilibrium with* $\bar{s}$ *as the norm. Then* $K_{\min}(|\bar{s}|)$ *is weakly increasing in* $|\bar{s}|$.

This proposition expresses two main results which extend part 2 of Proposition 1. Firstly, there exist single-norm equilibria for any $\bar{s}$, i.e., the norm may be biased. This holds as long as individuals care sufficiently about social pressure – K has to be greater than $K_{\min}(|\bar{s}|)$.[14] Secondly, the more biased the norm is, the larger is the K needed to sustain it in equilibrium. This last result is a key result. It essentially says that in order to uphold a biased norm, individuals in society need to care about social pressure more than is needed in order to uphold a more central norm. The intuition for this result is that the strength of the norm depends on the number of followers, where potential deviators are types with tastes far from the norm. When the norm is biased, there are more private tastes further away from the norm and hence more potential deviators. To sustain the norm this has to be compensated for by a heavier weight of pressure (i.e., a stronger emphasis on coordination).

Figure 2 depicts this equilibrium. The two graphs on the left show the case of a central norm, where the distribution of actions is shown in the upper left schedule and the mapping of types to actions in equilibrium is shown on the lower left. In this particular case all individuals conform fully to the norm. The right graphs show the case of a biased norm. Here, a group of extreme objectors express their heterogeneous private tastes

The previous results imply that, for a given value of K, there can be multiple equilibria, as the norm can be located anywhere along a continuous range, but these equilibria share the same pattern of norm conformity – alienation. Moreover, even for given values of K and $\bar{s}$ there can be multiple single-norm

[14] In the case of $\alpha > 1$ we get that $K_{\min} = 0$. This is a direct consequence of p being a step function. If one were to assume a less concave p (i.e., not a step function), $K_{\min}$ would have been greater than zero also when $\alpha > 1$. The rest of the results presented are not specific to the step function assumption: they hold more generally when $\beta < \alpha$ and $\beta \leq 1$.

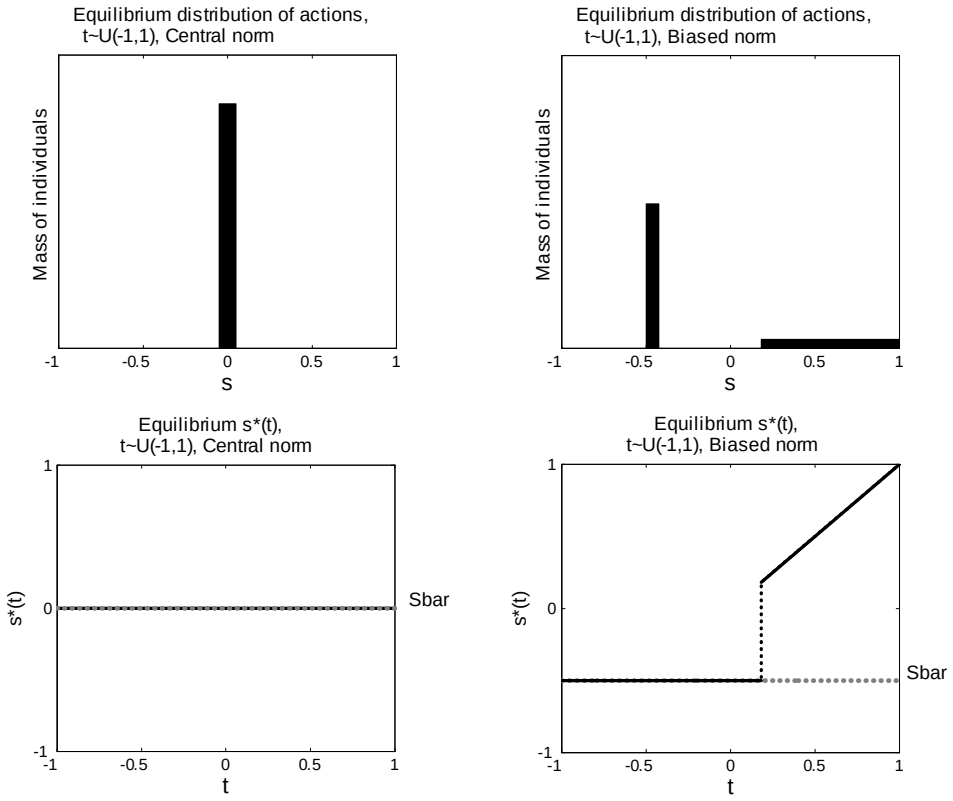

Figure 2: The left graphs show the distribution of actions (top) and $s^*(t)$ (bottom) in equilibrium with a central norm ($\bar{s} = 0$). The right graphs show the distribution of actions and $s^*(t)$ in equilibrium with a biased norm ($\bar{s} = -0.5$). In all figures $\beta = 0.01$, $\alpha = 0.9$ and $K = 1.2$.

equilibria. However, not all equilibria are dynamically stable, in the sense that a small perturbation to the share of norm followers may not lead to convergence back to the same equilibrium. In order to rule out such equilibria that have no gravity and to investigate further the properties of the stable equilibria, we add a simple dynamic structure to the model (as in, for instance, Granovetter 1978 and Kuran 1995). The dynamics considered are such that we perturb the share of norm followers in a single-norm equilibrium and examine whether there is convergence back to that equilibrium. Let i indicate the period of the dynamic process (representing a time period or a generation).

Then, an individual of type t in period i solves the following problem.

$$\min_{s_i} L\left(s_i; t, s'_{i-1}\right) = D\left(|t - s_i|\right) + P\left(s_i; s'_{i-1}\right) \text{ where} \tag{11}$$

$$P\left(s_i; s'_{i-1}\right) \equiv \frac{1}{2}\int_{-1}^{1} p\left(\left|s_i - s^*_{i-1}(\tau)\right|\right) d\tau.$$

This formulation implies that a person in period i plays a best response against the observed behavior in society in period $i-1$. It can be interpreted either as individuals adjusting their actions when observing how others are acting, or as an overlapping generations model, where the actions of the older generation (the parents) create pressure on the younger generation (the kids), who put pressure on the next generation and so on.[15]

In the following proposition (and in Proposition 5 later on) we use $x_{ss}\left(|\bar{s}|, K\right)$ to denote the share of norm followers in a stable single norm steady state. We also present a welfare analysis, where the welfare of an individual with loss L is simply $-L$. This analysis enables us to establish a relationship of *first-order stochastic dominance* between different norms.[16]

Proposition 3 *Consider the dynamic model in (11) with p being a step function as in (6) and D as given in (4) with $\alpha > 0$. Then:*

[15] Implicitly we assume here that the distribution of *types* is stationary between generations. For short to medium-run analysis (say, limited to at most a few decades) this seems reasonable.

[16] For given values of K and $\bar{s}$ there can be up to two different stable single norm steady states, corresponding to two different values of $x_{ss}\left(|\bar{s}|, K\right)$. When there are two such steady states, one is always "degenerate" ($x_{ss}\left(|\bar{s}|, K\right) = 1$) and one is always "non-degenerate" ($x_{ss}\left(|\bar{s}|, K\right) \in \,]0, 1[$). The comparison of $x_{ss}\left(|\bar{s}|, K\right)$ with $x_{ss}\left(|\bar{s}'|, K\right)$ in statement (2) of the proposition is thus applied as follows: When there exist two $x_{ss}\left(|\bar{s}|, K\right)$ and two $x_{ss}\left(|\bar{s}'|, K\right)$, the proposition compares the non-degenerate with each other and the degenerate with each other; when there exist two steady states for $\bar{s}$ and one for $\bar{s}'$ (or vice versa), the proposition compares $\max\left\{x_{ss}\left(|\bar{s}|, K\right)\right\}$ with $\max\left\{x_{ss}\left(|\bar{s}'|, K\right)\right\}$; finally, when there is only one steady state for each norm, the proposition compares the unique $x_{ss}\left(|\bar{s}|, K\right)$ with the unique $x_{ss}\left(|\bar{s}'|, K\right)$. Moreover, whenever there exist two stable steady states for a given norm, the welfare distribution in the degenerate steady state first-order stochastically dominates the welfare distribution in the non-degenerate steady state (see see Lemma 19 in the appendix). Hence, we apply the same comparison rule to the comparison of welfare distributions in statement (3) of the proposition.

1. *For any* $\bar{s} \in [-1, 1]$*, there exists a stable single norm steady state if and only if* $K > K_{\min}(|\bar{s}|)$*.*

2. $x_{ss}(|\bar{s}|, K) \geq x_{ss}(|\bar{s}'|, K)$ *if and only if* $|\bar{s}| \leq |\bar{s}'|$*.*

3. *Consider a norm* $\bar{s}$ *and suppose* $K > K_{\min}(|\bar{s}|)$*. Let* x_i *denote the share of norm followers in period* i*. Then there exists a value* $x_{conv}(|\bar{s}|, K)$ *such that if* $x_i > x_{conv}(|\bar{s}|, K)$*, there is convergence to a stable steady state with a single norm* $\bar{s}$ *followed by a share* $x_{ss}(|\bar{s}|, K) > x_{conv}(|\bar{s}|, K)$*. Otherwise, if* $0 \leq x_i \leq x_{conv}(|\bar{s}|, K)$*, there is convergence to a stable steady state where each type follows her heart.*[17]

4. $x_{conv}(|\bar{s}|, K)$ *is increasing in* $|\bar{s}|$ *and decreasing in* K*.*

5. *The welfare distribution under* $|\bar{s}|$ *first-order stochastically dominates the welfare distribution under* $|\bar{s}'|$ *if and only if* $|\bar{s}| \leq |\bar{s}'|$*.*

To understand these results, recall that Lemma 1 shows that alienation is a distribution of actions that recreates itself. That is, if in period i there is a cutoff distance from the norm, beyond which types follow their hearts and within which they follow the norm, then there will exist such a cutoff also in period $i + 1$. This implies that, for a given $\bar{s}$, the dynamics of the model can be described by analyzing the dynamics of the proportion of norm conformers, $x_{i+1} = f(x_i)$. This function is the main building block for proving Proposition 3. We demonstrate a prototypical case in Figure 3 for $\alpha < 1$. The figure depicts a phase diagram with x_i on the horizontal axis and x_{i+1} on the vertical axis. The 45-degree diagonal depicts the steady state values, where $x_{i+1} = x_i$. As can be seen in the figure, $f(0) = 0$, and then $f(x_i)$ starts below the 45-degree line, but afterwards it increases and crosses the 45-degree line and stays above it. Hence, $x = 1$ and $x = 0$ are stable steady states in this case, while there is an interior unstable steady state between them. The value of x in this inner state (x_{conv} in the proposition) also forms the boundary between

[17] For brevity, we treat the unstable steady states (x_{uss}) as ones where if $x_i = x_{uss}$ then $x_{i+1} < x_{uss}$.

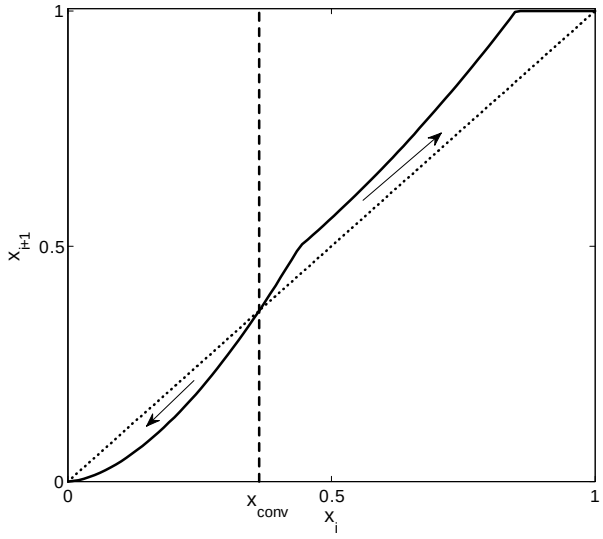

Figure 3: A phase diagram showing the dynamics for $\bar{s} = -0.5$, p being a step function, $\alpha = 0.6$ and $K = 1.5$. The dotted line depicts the diagonal where $x_{i+1} = x_i$, the solid line depicts the intertemporal dynamics $x_{i+1} = f(x_i)$. The vertical line depicts x_{conv}, i.e., the boundary between the zone of convergence to a single-norm equilibrium ($x = 1$) and to "pluralism" ($x = 0$).

the zone of convergence to a stable single norm (with $x_{ss} = 1$) and the zone of divergence toward a state of *pluralism* ($x = 0$). The figure also highlights that the steady state in which a norm exists is stable not only with respect to small perturbations: there is convergence to it from a rather broad range of initial conditions (depending on the value of K). In the specific example depicted in the figure, the stable single norm steady state is degenerate, in the sense that everyone in society adheres to the norm ($x_{ss} = 1$), but more generally there can be non-degenerate stable steady states ($x_{ss} < 1$), i.e., where part of the population is alienated.

Apart from convergence, the proposition also highlights the effect of the bias of the norm. Parts (1) and (3) of the proposition imply that a biased norm can persist also in a dynamic setting. This means that societies may be history dependent in the following sense. Suppose a group of individuals at some point choose the same action. Then, provided that they are sufficiently many ($x_i >$

$x_{conv}(|\bar{s}|)$), this mode of behavior may be established as a norm and may persist also after those individuals are gone, even if it does not represent the average private taste in society. Note also that if that initial group is only slightly larger than x_{conv}, the norm will gain more followers over time, thus becoming stronger. The fourth part of the proposition states that the minimum amount of conformity (x_{conv}) necessary for the norm to be sustainable in the long run is decreasing in the weight of the pressure and increasing in the bias of the norm. This can be demonstrated using Figure 3. By increasing K, the function $f(x_i)$ tilts upwards, which implies that x_{conv} decreases and so the zone of convergence to the single-norm equilibrium increases. In contrast, by increasing $|\bar{s}|$, the function $f(x_i)$ tilts downwards, implying a smaller zone of convergence. Hence, increasing K and increasing $|\bar{s}|$ work in opposite directions. This means that, while a biased norm can persist in this dynamic setting, the more biased it is, the less magnetic it is, unless it is compensated for by a larger K. Hence, biased norms are less sustainable than central norms in two ways. Firstly, they require people to care more about social pressure ($K_{\min}$ is higher). Secondly, they require more conformity in the first period (x_{conv} is higher).

Part 2 of the proposition implies that public cohesion in society – i.e., the extent of norm conformity – is falling with biasness. This has further implications for the sustaining and collapse of norms. To see why, suppose that a long time ago a steady state with a norm $\bar{s}$ was established based on the type distribution of that time. Now suppose that the type distribution, throughout history, has gradually shifted away from the norm due to a change of private sentiments in society. Part 2 implies that this shift will be accompanied by a decline in norm conformity. Eventually, once the type distribution has shifted sufficiently, the norm $\bar{s}$ will no longer constitute a stable steady state (this happens when $f(x_i)$ shifts below the 45-degree line in Figure 3) and the norm will collapse.

Part 5 of the proposition addresses the issue of welfare, stating that the welfare distribution under a biased norm is stochastically dominated by that of a more central norm. First order stochastic dominance between the welfare distributions under $\bar{s}$ and $\bar{s}'$ means that, if we rank individuals according to

their welfare under each norm, then an individual at rank r under $\bar{s}$ is at least as well off as an individual at rank r under $\bar{s}'$, and this holds for all ranks. Quite intuitively, the ranking of individuals under a given norm follows from their distance to that norm. That is, type t has a higher welfare than t' whenever $|t - \bar{s}| < |t' - \bar{s}|$. This implies that the types far from the norm who follow their hearts are at the bottom of the ranking. Also note that, when p is a step function, a person who follows her heart under $\bar{s}$ has the same welfare as a person who follows her heart under $\bar{s}'$. Thus, given that under a biased norm there are more people who follow their hearts, a biased norm implies there are more people with this lowest welfare. Moreover, those who conform under both norms are better off under the central norm, as there are less non conformers who pressure them.

However, while the welfare distribution under a central norm stochastically dominates all other welfare distributions, it is not Pareto dominant.[18] Our interpretation of these welfare results is that, while there will always be disagreement between specific individuals about what the best norm is, the alienating society is more likely to establish a central norm, as it will imply a higher welfare for more people. Hence, to the extent that the alienating society sustains biased norms, it is probably due to a shift of private sentiments away from what used to be, historically, a central norm. Generally, the proposition paints a coherent picture of biased norms being weaker than central norms in alienating societies as they imply less cohesion, lower welfare and a smaller zone of convergence, and require a harsher punishment to be sustained.

3 Inverting societies

The purpose of this section is to further examine the case represented by point 3 of Proposition 1 and by the lower left region in Figure 1. The fundamental characteristic of this case is that D is concave and, in particular, more concave than p. To capture this very concave cognitive dissonance, we let D be a step

[18] This is somewhat easy to see when considering the case of $\alpha < 1$ and noting that types sufficiently close to the norm would rather have the norm exactly at their bliss point rather than at 0, due to the concavity of D, but it can be shown that this result applies more broadly to any combination of K and α.

function,

$$D\left(|t-s|\right)=\begin{cases}1 \text{ if } |t-s|\neq 0\\ 0 \text{ if } |t-s|=0\end{cases} \tag{12}$$

while $p=K\left|s-s'\right|^{\beta}$ for some $\beta\leq 1$. A first useful result is the following.

Lemma 3 *Suppose D is according to (12) and $P\left(s;\cdot\right)$ has a unique min point at $\bar{s}$ and is increasing in the distance from $\bar{s}$ on each side. Then on each side of $\bar{s}$ there exists a cutoff value such that types within the cutoff follow their hearts while types beyond the cutoff choose $s^{*}(t)=\bar{s}$.*

The intuition for this result is straightforward. When D is a step function, an individual will either follow her heart or, once she deviates from her private taste, choose the action that lowers social pressure the most – this action is $\bar{s}$ in the lemma. This is so because she does not distinguish between actions that are not exactly her private taste. Hence, the lemma essentially says that a minimum point of social pressure $\bar{s}$ may function as a norm by inducing full conformity by some types. The question then is which individuals will be the full conformers and which individuals will follow their hearts. When social pressure is increasing with the distance from the norm, types far from the norm will find it the hardest to follow their hearts. Meanwhile, the dissonance of deviation from one's bliss point is independent of type. Hence, there will be a cutoff distance from the norm such that types beyond it fully conform, while types within it will follow their hearts. On the aggregate level this can be interpreted as an *inversion of preferences,* as those who despise the norm the most are the ones following it in public. Furthermore, the fact that those who nearly agree with the norm follow their hearts openly can be interpreted as existence of mild critique. This pattern of conformity thus echoes the result for the more general case of $\alpha<\beta\leq 1$, as stated in part 3 of Proposition 1.

Now, the previous lemma was a form of partial equilibrium since it assumed that P monotonically increases in the distance from a unique minimum point $\bar{s}$. The question then is whether the individual choices implied by Lemma 3 induce such properties of P. In the upcoming analysis we will again use y (with some abuse of notation) to denote the distance between the norm and

the type who is indifferent between following her heart and following the norm.

Lemma 4 *Suppose that $\beta \leq 1$ and that there exist a norm $\bar{s} \in [-1,1]$ and a cutoff value $y \in [0, 1+|\bar{s}|]$ such that all types with $|t-\bar{s}| \leq y$ choose $s^*(t) = t$ while the rest choose $s^*(t) = \bar{s}$. Then there exists a value $y_{\max}(\bar{s}) \geq 1$ such that $P(s;\cdot)$ has a unique min point at $\bar{s}$ and is increasing on each side of $\bar{s}$ if and only if $y \leq y_{\max}(\bar{s})$.*

While the previous lemma described what individuals choose given social pressure, this lemma describes the properties of social pressure given the choices of individuals. The bottom line of Lemma 4 is that if there is inversion preferences, then P will be increasing in the distance from the norm, as long as there are sufficiently many norm followers. This is the same as requiring that the most deviant action in society (at distance y from the norm) is not too deviant. $y_{\max}(\bar{s})$ then measures how deviant this behavior can be while still ensuring that P is everywhere increasing in the distance from $\bar{s}$.

Put together, Lemmas 3 and 4 allude to the existence of an equilibrium, since the first says that inversion of preferences will arise if P is increasing in the distance from $\bar{s}$ and the second says that given inversion, P will be increasing in the distance from $\bar{s}$. The conditions for the existence of such an equilibrium are presented in the following proposition.

Proposition 4 *Suppose D is according to (12) and p is according to (5) with $\beta \leq 1$. Then:*

1. *For each value of $\bar{s} \in [-1,1]$ there exists a lower bound for K, denoted by $K_{\min}(|\bar{s}|)$, such that a single-norm equilibrium with a norm $\bar{s}$ exists if and only if $K \geq K_{\min}(|\bar{s}|)$.*

2. *$K_{\min}(|\bar{s}|)$ is weakly decreasing in $|\bar{s}|$.*

The existence of single-norm equilibria when $\alpha < \beta < 1$ was stated already in Prop 1 (part 3), but the current proposition adds that any norm $\bar{s}$ in the

range $[-1, 1]$ can be sustained in equilibrium as long as $K \geq K_{\min}(|\bar{s}|)$.[19] For the norm to exist, it has to constitute an attractive mode of behavior, relative to other modes of behavior an individual can adopt. For this to be achieved, enough individuals need to fully conform, thereby lowering pressure at the norm. This requires that individuals care sufficiently about coordinating with others – K has to be sufficiently large. Unlike the alienating society (see Proposition 2), here the pattern of individual choice is that of inversion of preferences. In the appendix (Lemma 26) we show that inversion is the *only* pattern of individual choice consistent with a single-norm equilibrium when D is a step function.

The second part of the proposition implies that a biased norm not only may exist, but also the conditions for its existence are less demanding the more biased it is – individuals can care less about social pressure. The crude intuition for this is that inversion implies that types far from the norm uphold it. Hence, a norm that is far from people's tastes generates more conformity, which makes the norm stronger.

For a more detailed explanation, consider the distribution of actions under a central norm, as depicted in the upper left panel of Figure 4. Suppose now that we move the norm slightly toward the left edge. The conformity of types at the edges of the type distribution then implies that the "distribution package" will move together with the norm without changing appearance – those beyond $\bar{s} \pm y$ will fully conform, while those within this range will follow their hearts. This illustrates that biased norms may exist. Now, if we continue moving $\bar{s}$ leftward, at some point the type $t = \bar{s} - y$ will equal -1. When moving $\bar{s}$ beyond this point, the left wing of the uniform part will be truncated (as in the upper right panel of Figure 4). This truncation of the left wing further implies a narrowing of the right wing. The reason for this is that the effect of truncation of the left wing is similar to inducing conformity by people on the left side of the norm. Then, as explained in Section 1, the concavity of peer pressure implies that the conformity of leftists will inspire more conformity of

[19] Note that this value is not necessarily equal to the $K_{\min}(|\bar{s}|)$ under alienation in Proposition 2.

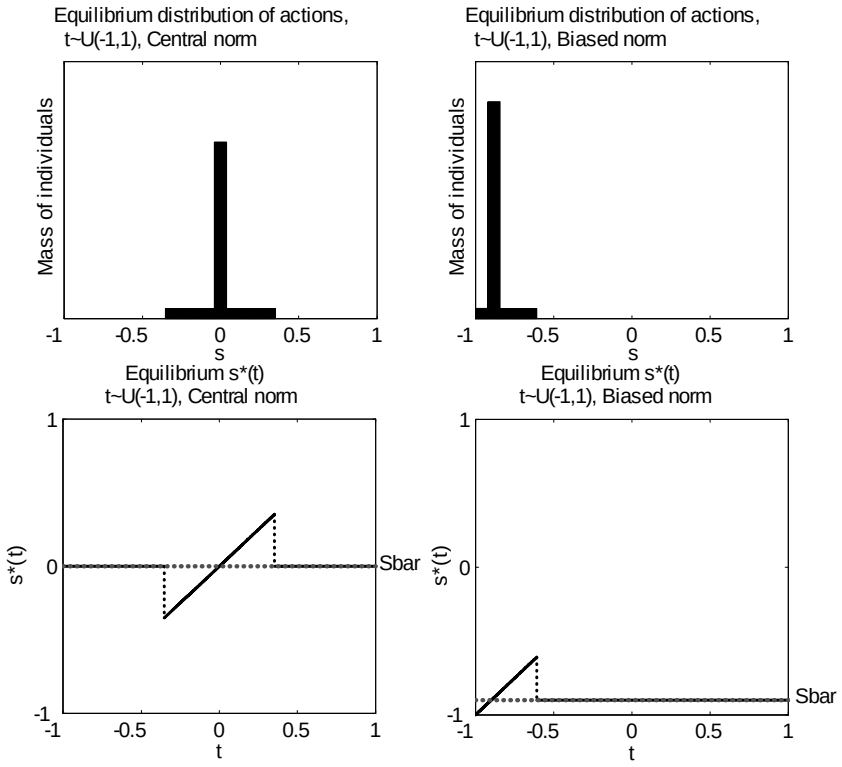

Figure 4: The left graphs show the distribution of actions (top) and $s^*(t)$ in equilibrium (bottom) with a central norm ($\bar{s} = 0$). The right graphs show the distribution of actions and $s^*(t)$ in equilibrium with a biased norm ($\bar{s} = 0.9$). In all figures $\beta = 0.6$, $\alpha = 0.1$ and $K = 1.6$.

rightists, making the norm a stronger focal point. Consequently, a lower K is needed in order to sustain the norm in equilibrium. All in all, biasness of the norm thereby compensates for weakness of social pressure, making biased norms more sustainable than central norms. Put differently, biasness facilitates coordination.

We will now analyze the dynamic stability of these equilibria and the properties of the stable ones. For this purpose we add the same dynamic structure to the model as we did in the previous section (see equation 11). Here we will perturb the cutoff y in a single-norm equilibrium and examine whether there

is convergence back to this equilibrium.[20]

Proposition 5 *Consider the dynamic model in (11) with D being a step function as in (12) and p as given in (5) with $\beta \leq 1$. Then:*

1. *For any $\bar{s} \in [-1, 1]$, there exists a single norm stable steady state if and only if $K > K_{\min}(|\bar{s}|)$.*
2. *$x_{ss}(|\bar{s}|, K) \leq x_{ss}(|\bar{s}'|, K)$ if and only if $|\bar{s}| \leq |\bar{s}'|$.*
3. *Consider a norm $\bar{s}$ and suppose $K > K_{\min}(|\bar{s}|)$. Let y_i denote a cutoff value in period i, such that all types with $t \in [\bar{s} - y_i, \bar{s} + y_i]$ follow their hearts while the rest follow the norm. Then there exists a value $y_{conv}(|\bar{s}|)$, such that there is convergence to a stable steady state with a single norm $\bar{s}$ if $y_i < y_{conv}(|\bar{s}|)$.*
4. *$y_{conv}(|\bar{s}|)$ is increasing in $|\bar{s}|$.*

To understand these results, first note that Lemmas 3 and 4 together imply that inversion of preferences in period i recreates inversion in period $i+1$ with a new cutoff value of conformity. This implies that the dynamic process can be described solely by the dynamics of the cutoff y_i. Figure 5 shows a phase diagram that depicts y_{i+1} (vertical axis) as a function of y_i (horizontal axis). As can be seen from the figure, there is a stable steady state with a norm when $y_i = y_{ss}$. The existence of such a steady state for a given $|\bar{s}|$ hinges on K being greater than $K_{\min}(|\bar{s}|)$, as defined in the static Proposition 4. It may be interesting to note that the steady state is never degenerate – there is always a share of the population (those close to the norm) who follow their hearts. In the proof of the proposition we show that an increased $|\bar{s}|$ pushes the function y_{i+1} downward, which implies that y_{ss} decreases with biasness, so that the most deviant behavior in the steady state becomes less deviant. This has the

[20] Since for given $\bar{s}$ and K there can exist more than one stable single-norm state steady, the comparison of the share of norm followers (x_{ss}) in statement 2 in the following proposition is similar to the comparison in Proposition 3. See the proof of statement 2 of Proposition 5 for details.

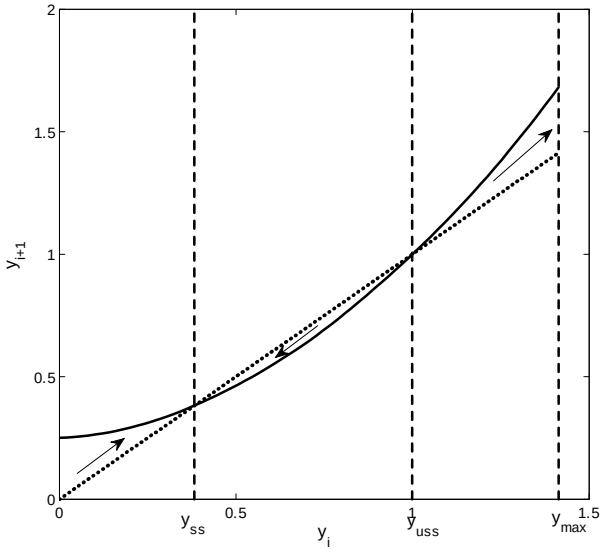

Figure 5: A phase diagram showing convergence to a stable single-norm equilibrium when $\bar{s} = -1$, for D being a step function, $\beta = 0.5$ and $K = 2$. The dotted line depicts the diagonal where $y_{i+1} = y_i$, the solid line depicts the intertemporal dynamics $y_{i+1} = f(y_i)$. The vertical lines depict the upper bounds for convergence, y_{uss} and $y_{\max}$ (with y_{uss} being the binding one in the case depicted here). The phase diagram is not defined for $y_i > y_{\max}$.

further consequence that the share of the population conforming increases with biasness (part 2 of the proposition). This means that cohesiveness increases with the bias of the norm as the deviant behavior becomes less extreme and there are more norm conformers.

If $y_i < y_{ss}$, society will converge to this stable steady state. Furthermore, there may be another, unstable, steady state at y_{uss}, which marks the border between the convergence zones. The existence of y_{uss} hinges on $f(y_i)$ intersecting the 45-degree line twice to the left of $y_{\max}$, as depicted in the diagram. Beyond $y_{\max}$, P is non-monotonic and hence the phase diagram is not applicable. If there exists such $y_{uss} < y_{\max}$, then $y_i < y_{uss}$ is a necessary and sufficient condition for convergence to the stable steady state y_{ss}. However, if there does not exist such a y_{uss}, there is convergence to y_{ss} starting from any $y_i < y_{\max}$. Hence, the sufficient conditions for convergence are that $y_i < y_{\max}$ and that

$y_i < y_{uss}$ whenever y_{uss} exists. The last point of the proposition states that the range of convergence, $[0, y_{conv} \equiv \min\{y_{uss}, y_{\max}\}]$, increases with biasness.[21]

y_{conv} may be interpreted as the maximum level of initial public deviance. If initially a norm exists and the most deviant behavior is less deviant than y_{conv}, then this norm will stay stable over time. It should be noted, as is exemplified in Figure 5, that y_{conv} is often much larger than y_{ss}. Hence, we can start with a norm that is to a non-trivial degree weaker than in the steady state and still converge to the steady state. What point (4) of the proposition suggests is that the most deviant behavior in the first period can be more deviant the more biased the norm is.[22] Furthermore, Lemma 3 tells us that, in the first period, the norm need not necessarily be established through inversion of preferences – it is sufficient that one focal mode of behavior exists and then inversion will ensue in later periods. So an inverting steady state may be attained from a non-inverting initial condition. We thus get history dependence: if a group of individuals, possibly a long time ago, had established together one focal mode of behavior, this mode of behavior could become an endogenous norm, upheld by those who despise it the most.

One may note that Proposition 5 is silent about welfare. The reason for this is that, unlike in the alienating society, here it is not possible to get a consistent ranking of norm locations according to first-order stochastic dominance of welfare distributions.[23] The rough intuition for this is that on the one hand,

[21] This is so because an increased $|\bar{s}|$ not only tilts the function y_{i+1} downwards, which implies an increase in y_{uss}, but also because $y_{\max}$ increases with biasness.

[22] We say "suggests" since the proposition only establishes *sufficient* conditions for convergence ($y_0 < \min\{y_{uss}, y_{\max}\}$) as beyond $y_{\max}$ pure inversion may not be maintained, which substantially complicates the analysis. To see what happens when y_0 is beyond $y_{\max}$ we have performed an extensive set of simulations of the model for different combinations of α, β and K. They consistently show the same results: there is in practice a maximum value of y_0 below which there is convergence to a steady state with inversion, and above which society converges to pluralism. Importantly, this numerical cutoff of convergence is increasing in biasness.

[23] There is a handful of examples of pairs of norms whose corresponding welfare distributions do not stochastically dominate each other. Furthermore, it can be shown that a sufficient condition for guaranteeing that *there is no* stochastic dominance of the welfare distributions under $|\bar{s}| = 0$ and $|\bar{s}| = 1$ is that y_{ss} under $|\bar{s}| = 0$ will be smaller than 1.5 times y_{ss} under $|\bar{s}| = 1$. This sufficient condition holds for many parameter combinations that support stable equilibria at both $|\bar{s}| = 0$ and $|\bar{s}| = 1$, e.g. $\beta = 0.5$ and $K = 4$.

the maximal welfare under each distribution, experienced by the type exactly at the norm, is higher the more biased the norm is (because biasness implies more norm conformers and a narrower uniform part); but on the other hand, types at the far edge of the uniform part are worse off under a more biased norm, because they are further away from the norm compared to their equally ranked counterparts under the less biased norm and are more pressured given that the norm is stronger. One possible interpretation for this ambiguity with respect to welfare is that in the inverting society it is less clear-cut which norm will arise in equilibrium, whereas in the alienating society central norms seem unambiguously more plausible. However, once a norm *has* been established, the predictions for the inverting society are unambiguous: Biased norms are more stable, as they require lower social pressure ($K_{\min}$ is lower), imply more cohesion (x_{ss} is higher) and maintain their dynamic attraction in the presence of more deviant behavior. Furthermore, a norm will not collapse due to a shift of private sentiments away from it, as this will only increase cohesion.

4 Conclusion

This paper studies the existence, location and sustainability of endogenous social norms under peer pressure. In many situations characterized by peer pressure, individuals may truly disagree (on a private level) about the right ideology or best conduct. Hence, there will not exist a consensus behavior that can make for an exogenous norm. Nevertheless, we show that in these situations a clear norm (or point of coordination) may be endogenously sustained and will also be dynamically stable. That is, there may seem to exist a consensus about a certain mode of behavior, which many in society adopt, while in fact individual preferences are completely heterogeneous. Moreover, a norm that is biased with respect to private preferences will sometimes be more sustainable than a representative norm. This can shed light on the sustainability of biased norms, as observed for example in religious communities, racial attitudes and honor cultures.

The paper maps societies into a class that cannot maintain an endogenous norm and a class that can. Within the class that *can*, the paper highlights

a fundamental difference between two main subclasses of societies. Firstly, in societies where pairwise pressure is sufficiently concave, individuals with tastes that are very different from the norm may be *alienated* and act according to these tastes in public. For a norm to survive in this type of society, it has to be sufficiently representative of the tastes of individuals in society. If society is very heterogeneous, or the norm is biased, a norm can be sustained only under strong pressure to coordinate. In the other subclass of societies, where pairwise pressure is not sufficiently concave, preferences will be *inverted* – the ones following their hearts will be those with tastes that are only slightly different from the norm, while those who privately dislike the norm the most will fully conform. This means that in this kind of society we should observe only small deviations from the norm. Here biased norms are more sustainable and more magnetic than representative norms.

We believe the model in this paper represents an essential element in human interaction. Namely, that coordination problems arise in between multiple individuals with heterogeneous tastes. Analytically proving outcomes in this setting is not a trivial matter and we have not exhausted the possible equilibria that can arise. However, our results of the dynamic model strongly indicate that the single-norm equilibrium, which has been the focus of this paper, is not just a technical possibility – outcomes will tend to gravitate toward these equilibria from a broad set of initial conditions.

A Appendix: Non-uniform distribution of types

The results of when alienation and when inversion arises and that these equilibria cannot exist for the same set of parameters hold for any continuous distribution of types. The results and logic of norm location, however, have to be refined.

When $\beta < \alpha$ (and $\beta \leq 1$), single-norm equilibria will be characterized by alienation. This has implications for the location of the norm and for the level of cohesion in society. It implies that unless K is very large, the norm can be sustained only if it is located such that many in society largely agree with it privately. This is since otherwise there would be a large portion of opposers to the norm, who, by opposing, would make conforming unattractive even to those who object the norm less. Figure 6 shows steady states under some other distributions of types. Under a normal distribution, the norm has to be located within the bell of the normal distribution (as represented by the mass point in the upper left panel of Figure 6). Alternatively, if the whole distribution is skewed, the norm needs to be located on the same side as the mass of types (upper right).

When $\alpha < \beta \leq 1$, single-norm equilibria are characterized by inversion. Hence, for a norm to be sustainable and have a high degree of cohesion it has to be located *away* from any mass of private opinions. Otherwise, if there is a mass of people with tastes close to the norm, these people will choose to follow their hearts, and by doing so will make the norm less attractive even to those whose tastes are further away (and are therefore subject to more pressure when following their hearts). This can be seen in Figure 6 (bottom left), where we illustrate a case with a normal distribution of types. The norm cannot be sustained within the bell-shape but only at the tails. On the bottom right of Figure 6 we see that if the type distribution is bimodal, a norm can be sustained virtually anywhere except close to the peaks.[24]

B Appendix: Initial analytical results

For ease of notation, throughout all the upcoming appendices, we will use $P(s)$ instead of $P(s; s')$ to denote the aggregate pressure felt by choosing action s under a predefined distribution of actions in society.

[24] Under a skewed distribution (e.g. an exponential distribution) it may be possible to support a norm also close to the peak of private preferences. However, this norm will only be followed by very few types with private tastes in the tail, while the vast majority will follow their hearts. Hence, the simulation- and intuition-based conjecture is that it will be hard to sustain a norm with *high degree of cohesion* if the norm is located such many nearly agree with it but not fully.

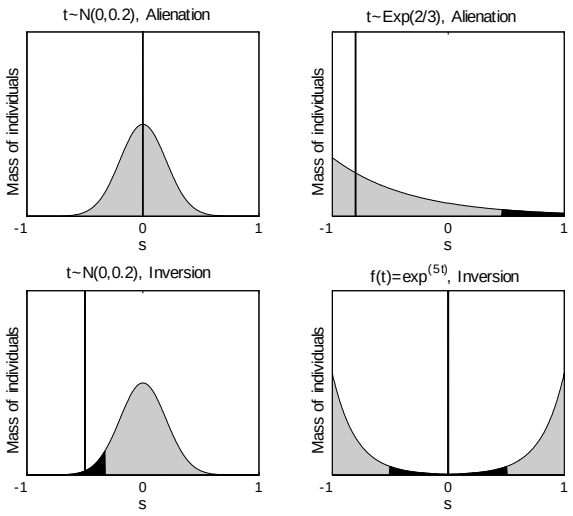

Figure 6: Histograms with single norm steady states in the dynamic model. In each histogram, the black surface represents the steady state distribution of actions while the grey surface represents the underlying distribution of types. The distribution of actions in the zeroth generation is such that all take the same action. Note that the y-axes have been truncated for visibility and that the distributions of types have been truncated (where applicable) to be between -1 and 1. Upper left: $\alpha = 0.5$, $\beta = 0.01$, $K = 1.2$, $\bar{s} = 0$. Upper right: $\alpha = 0.5$, $\beta = 0.01$, $K = 1.2$, $\bar{s} = -0.8$. Lower left: $\alpha = 0.01$, $\beta = 0.5$, $K = 2.5$, $\bar{s} = -0.5$. Lower right: $\alpha = 0.01$, $\beta = 0.5$, $K = 1.5$, $\bar{s} = 0$.

Lemma 5 *Let there be a range of types that follow their hearts. Then the aggregate pressure that results is strictly increasing in the distance from the middle of the range.*

Proof. *The range of types that follow their hearts form a uniform part in the distribution of actions, with pdf* $= \frac{1}{2}$ *within the range and* 0 *outside. Denote this uniform range by* $[a, b]$ *with* $a < b$*. Then*

$$P(s) = \frac{1}{2}K \int_a^b |s - \tau|^\beta \, d\tau$$

$$= \begin{cases} \frac{1}{2}K \frac{(b-s)^{\beta+1} - (a-s)^{\beta+1}}{\beta+1} & \text{if } s < a \\ \frac{1}{2}K \frac{(s-a)^{\beta+1} + (b-s)^{\beta+1}}{\beta+1} & \text{if } a \leq s \leq b \\ \frac{1}{2}K \frac{(s-a)^{\beta+1} - (s-b)^{\beta+1}}{\beta+1} & \text{if } s > b \end{cases}$$

$$P'(s) = \begin{cases} \frac{1}{2}K\left[-(b-s)^{\beta} + (a-s)^{\beta}\right] < 0 \text{ if } s < a \\ \frac{1}{2}K\left[(s-a)^{\beta} - (b-s)^{\beta}\right] \text{ if } a \leq s \leq b \\ \frac{1}{2}K\left[(s-a)^{\beta} - (s-b)^{\beta}\right] > 0 \text{ if } s > b \end{cases}$$

It is easy to see that $P'(s) > 0$ *if* $s > \frac{a+b}{2}$ *and* $P'(s) < 0$ *if* $s < \frac{a+b}{2}$*, implying that* $P(s)$ *is strictly increasing in the distance from the middle of the range.* ■

C Appendix: Proof of Proposition 1

C.1 Part 1

First we note that there cannot be a norm at one of the distribution edges, i.e., at $\bar{s} = -1$ or at $\bar{s} = 1$. To see this, note that a norm at, say, $\bar{s} = 1$ implies that the slope of the aggregate pressure P at the norm is positive (because deviation to the left decreases the pressure from all actions besides $s = 1$ while, when $\beta > 1$, not affecting the pressure stemming from the mass of people at the norm), and so everyone would like to deviate to the left, contradicting the existence of a norm there.[25] Next we consider norms at the interior of $[-1, 1]$. Here, note that when $\beta > 1$ then p and p' are continuous everywhere, which implies that $P = \int p$ and $P' = \int p'$ must be continuous everywhere as well. In particular at $s = \bar{s}$. Hence, $P' \mid_{s=\bar{s}}$is well defined, and so either $P' \mid_{s=\bar{s}} = 0$ or $P' \mid_{s=\bar{s}} \neq 0$.

If $P' \mid_{s=\bar{s}} = 0$, then it must be that $s^*(t) \neq \bar{s}$ for any $t \neq \bar{s}$, because for $t \neq \bar{s}$, a small enough deviation from $\bar{s}$ toward t decreases D without increasing P. Thus there is no positive mass of individuals at $\bar{s}$, so it cannot be the norm.

If $P' \mid_{s=\bar{s}} \neq 0$ then either $P' \mid_{s=\bar{s}} > 0$ or $P' \mid_{s=\bar{s}} < 0$. If $P' \mid_{s=\bar{s}} > 0$, then (1) no type with $t < \bar{s}$ will state the norm, as deviating in the left direction from $\bar{s}$ reduces both P and D, and (2) at most one type with $t > \bar{s}$ can have $|D'(\bar{s}; t)| = |P'(\bar{s})|$ when D is strictly concave or strictly convex (i.e., when $\alpha \neq 1$), and so only this one type can have a local min point of L at $\bar{s}$. This means that when $\alpha \neq 1$ there can be no positive mass at $\bar{s}$, which violates the definition of a norm. Now suppose $\alpha = 1$ so that $D = |t - \bar{s}|$. Then each type either follows her heart or states a statement s such that $|P'(s)| = 1$. Then there can potentially be multiple types choosing the same action $\bar{s}$ such that $|P'(\bar{s})| = 1$, which implies $\bar{s}$ can be a norm. Suppose this holds and that $\bar{s}$ is the unique norm. Then the fact that no type with $t < \bar{s}$ states $\bar{s}$, together with (i) the uniqueness of the norm $\bar{s}$ and (ii) the fact that a type who does

[25] This holds also if $s = 1 \ \forall t$. In this case the slope of the aggregate pressure P at the norm is 0, but deviation to the left (i.e. towards one's bliss point) is still profitable as it reduces D.

not state an s such that $|P'(s)| = 1$ necessarily follows her heart, imply a uniform distribution of actions to the left of $\bar{s}$, stemming from the choices of types at this range to follow their hearts (this is necessarily so since (a) there must be a finite number of points with $s < \bar{s}$ and $|P'(s)| = 1$ implying that if $s \neq t$ for a positive mass of types with $t < \bar{s}$ then uniqueness of the norm is violated, and (b) all types with $t > \bar{s}$ state $s \geq \bar{s}$ because they either follow their hearts or choose the unique norm). The shape of the pressure imposed by the uniform part at $s = [-1, \bar{s}]$ is symmetric around its center, creating the same slope at both edges of this part, $s = -1$ and $s = \bar{s}$. On top of it, there is the pressure stemming from actions $s \geq \bar{s}$. As $\beta > 1$, each of these sources of pressure implies a steeper slope at $s = -1$ than at $s = \bar{s}$, which altogether means that $|P'(-1)| > |P'(\bar{s})| = 1$. This implies that types close to $t = -1$ will gain by deviating to the right from their bliss points, in contradiction to the assumption that they follow their hearts The same argument applies when $P' |_{s=\bar{s}} < 0$.■

C.2 Part 2

C.2.1 Sentence 1: Existence

Suppose that indeed $s(t) = \bar{s} = 0 \quad \forall t \in [-1, 1]$. To show that this is an equilibrium, and using symmetry, we need to show that, for a sufficiently large K, $L(s,t) > L(0,t) \quad \forall t \in (0,1]$ and $s \in (0,t]$. Given that all types conform, the pressure function at $s \in (0,1]$ is simply given by $P(s) = Ks^{\beta}$. We thus need to show that, for a sufficiently large K, $(t-s)^{\alpha} + Ks^{\beta} > t^{\alpha} \quad \forall t \in (0,1]$ and $s \in (0,t]$. Let $f(s,t) \equiv \frac{t^{\alpha} - (t-s)^{\alpha}}{s^{\beta}}$ be a function defined for $s \in (0,t]$. We will show that $f(s,t)$ is finite. If $\alpha = 1$ then $f(t,s) = s^{1-\beta}$, which is finite. If $\alpha > 1$, the numerator $t^{\alpha} - (t-s)^{\alpha}$ is increasing in t hence reaches its maximum at $t = 1$ where it equals $1 - (1-s)^{\alpha}$. This means that in this case $f(s,t)$ is bounded from above by $\frac{1-(1-s)^{\alpha}}{s^{\beta}}$. It is easy to see that the numerator is finite, and when $s \to 0$ we get by L'Hôpital's rule that, if $\beta < 1$, $\lim_{s \to 0} \frac{1-(1-s)^{\alpha}}{s^{\beta}} = \lim_{s \to 0} \frac{\alpha(1-s)^{\alpha-1}}{\beta s^{\beta-1}} = 0$, hence $\frac{1-(1-s)^{\alpha}}{s^{\beta}}$ is finite, implying that $f(s,t)$ is finite (if $\beta = 1$ then $\lim_{s \to 0} \frac{1-(1-s)^{\alpha}}{s} = \lim_{s \to 0} \frac{\alpha(1-s)^{\alpha-1}}{1} = \alpha$). Finally, if $\alpha < 1$, the numerator $t^{\alpha} - (t-s)^{\alpha}$ is decreasing in t hence reaches its maximum when $s = t$, where it equals s^{α}. This means that in this case $f(s,t)$ is bounded from above by $s^{\alpha-\beta}$, which is itself bounded (even when $s \to 0$, given that $\beta \leq \alpha$). Overall, we thus get that for any $\alpha \geq \beta$ the function $f(s,t)$ is bounded from above by some finite value $\sup f(t,s)$. Hence, for any $K > \sup f(s,t)$, there exists an equilibrium in which $s = \bar{s} = 0 \quad \forall t \in [-1,1]$. In particular, this implies that types closest to the norm fully conform. For showing the existence

of multiple equilibria, a similar proof can be constructed for $\bar{s} \neq 0$.■

C.2.2 Sentence 2: Properties

We will now prove the second statement. We will prove the statement for types to the right of the norm, that is, that there exists a $t_{\max}$ such that the set of types $]\bar{s}, t_{\max}]$ fully conform. A necessary and sufficient condition for full conformity of t is that $\bar{s} = \arg\min L$. A sufficient condition for this to hold is that, for any $s \in [\bar{s}, t]$, $P'(s) > -D'(t-s)$. For types sufficiently close to the norm this is equivalent to showing that $\lim_{s\to 0^+} P'(s) > \lim_{s\to t^-} \alpha (t-s)^{\alpha-1}$. Since the pressure imposed by those stating the norm, $\lim_{s\to 0^+} xK\beta (s-\bar{s})^{\beta-1}$, is strictly greater than $\lim_{s\to t^-} \alpha (t-s)^{\alpha-1}$ for any $x > 0$, it is sufficient to show that the pressure imposed by the individuals who do not conform cannot cancel out, completely or partly, the pressure imposed by the norm followers. Since $K\beta (s-s')^{\beta-1}$ is finite for any strictly positive $s - s'$, while $\lim_{s\to 0^+} xK\beta (s-\bar{s})^{\beta-1} = \infty$, it is sufficient to study the possibility that the non-conforming distribution of stances has a point of singularity exactly at $\bar{s}$. Hence, we will show that $\lim_{s\to 0^+} P'(s) > \lim_{s\to t^-} \alpha (t-s)^{\alpha-1}$ holds even if the distribution of stances of the non-conforming individuals has a point of singularity with a slope of minus infinity exactly at the norm.[26]

We perform the proof for $\bar{s} = 0$ but equivalent statements hold for any $\bar{s} \neq 0$. Suppose a norm $\bar{s} = 0$ exists and let $q(s')$ denote the distribution of stances outside the norm. Then

$$P(s \geq 0) = Kxs^{\beta} + K\left[\int_{-1}^{s} (s-s')^{\beta} q(s')\, ds' + \int_{s}^{1} (s'-s)^{\beta} q(s')\, ds'\right]$$

where $x > 0$ since a norm exists.

$$\frac{d}{ds}\left\{\int_{-1}^{s} (s-s')^{\beta} q(s')\, ds' + \int_{s}^{1} (s'-s)^{\beta} q(s')\, ds'\right\}$$
$$= \int_{-1}^{s} \beta (s-s')^{\beta-1} q(s') ds' - \int_{s}^{1} \beta (s'-s)^{\beta-1} q(s') ds'$$

[26] The case of a point of singularity with a slope of plus infinity is not interesting here because it will only increase $\lim_{s\to 0^+} P'(s)$ relative to $\lim_{s\to t^-} \alpha (t-s)^{\alpha-1}$.

$$\lim_{s\to 0^+} P'(s) = xK\beta \lim_{s\to 0^+} s^{\beta-1} + K\beta \lim_{s\to 0^+} \left\{ \int_{-1}^{s} (s-s')^{\beta-1} q(s')ds' - \int_{s}^{1} (s'-s)^{\beta-1} q(s')ds' \right\}$$

If $q(s')$ does not have a point of singularity at $s = 0$ then it is immediate that

$$\beta < \alpha \Rightarrow \lim_{s\to 0^+} P'(s) = xK\beta \lim_{s\to 0^+} s^{\beta-1} > \lim_{s\to t^-} \alpha (t-s)^{\alpha-1}$$

However, if $q(s')$ does have a point of singularity at $s' = 0$, then, given that $q(s')$ is integrable, $q(s')$ must have an integrable singularity at $s' = 0$, hence $\lim_{s'\to 0^+} q(s')s' = 0$.[27] This means that for any $\varepsilon > 0$, however small, there is a $\delta > 0$ such that, for $s' < \delta$, $q(s') < \varepsilon/s'$. The integral

$$I \equiv \int_{s}^{1} (s'-s)^{\beta-1} q(s')ds'$$

can then be split into two, $I = I_1 + I_2$, where

$$I_1 = \int_{s}^{\delta} (s'-s)^{\beta-1} q(s')ds'$$

and

$$I_2 = \int_{\delta}^{1} (s'-s)^{\beta-1} q(s')ds'$$

I_2 is finite for any $\delta > s$, so that $\lim_{s\to 0^+} s^{1-\beta} I_2 = 0$. As for I_1, changing the integration variable to $z = s'/s$ we have

$$\begin{aligned} I_1 &= s^{\beta} \int_{1}^{\delta/s} (z-1)^{\beta-1} q(zs)dz < s^{\beta} \int_{1}^{\delta/s} (z-1)^{\beta-1} \frac{\varepsilon}{zs} dz \\ &= s^{\beta-1}\varepsilon \int_{1}^{\delta/s} (z-1)^{\beta-1} z^{-1} dz \end{aligned}$$

[27] This is so because the function $1/s'$ is not integrable and so if $\lim_{s'\to 0^+} q(s')s' = \lim_{s'\to 0^+} \frac{q(s')}{1/s'} \neq 0$ it would imply that $q(s')$ is not integrable too, hence cannot be a valid distribution of stances.

where the inequality follows from the fact that $q(s') < \varepsilon/s'$. It follows that $\lim_{s\to 0^+} s^{1-\beta} I_1 < \varepsilon A$, where $A \equiv \int_1^\infty (z-1)^{\beta-1} z^{-1} dz$ is a positive and finite constant that is independent of ε. Then, since ε can be made arbitrarily small, it follows that $\lim_{s\to 0^+} s^{1-\beta} I_1 = 0$ hence $\lim_{s\to 0^+} s^{1-\beta} I = 0$, which means that I diverges at $s = 0$ slower than $s^{\beta-1}$ and so again

$$\beta < \alpha \Rightarrow \lim_{s\to 0^+} P'(s) = xK\beta \lim_{s\to 0^+} s^{\beta-1} > \lim_{s\to t^-} \alpha (t-s)^{\alpha-1}$$

Since we just established that types sufficiently close to the norm fully conform in any single-norm equilibrium, it follows that, if somebody does not fully conform, it has to be someone sufficiently far from the norm.■

C.3 Part 3

C.3.1 Sentence 1: Existence

Suppose a single norm exists at $\bar{s} \leq 0$ and

$$s'(t) = \begin{cases} t \text{ if } t \in [\bar{s}-x, \bar{s}+x] \\ \bar{s} \text{ if } t \in [-1, \bar{s}-x[\text{ or if } t \in]\bar{s}+x, 1] \end{cases}$$

where $x < \bar{s} + 1 \leq 1$ (i.e., the type $\bar{s} - x$ exists). This is an equilibrium if:[28]

- type $t = \bar{s} + x$ has no inner solution and is indifferent between choosing $s = \bar{s} + x$ and $s = \bar{s}$.
- types $t \in [\bar{s}, \bar{s}+x[$ choose $s^*(t) = t$,
- types $t \in]\bar{s}+x, 1]$ choose $s^*(t) = \bar{s}$,

which we will show holds, in three corresponding lemmas, for some x if K is sufficiently large. First, however, some auxiliary calculations.

$$\begin{aligned} P(s) &= K\frac{1}{2}\int_{\bar{s}-x}^{\bar{s}+x} (|s-t|)^\beta dt + (1-x) K (s-\bar{s})^\beta \qquad (13) \\ &= \begin{cases} K\left[(1-x)(s-\bar{s})^\beta + \frac{1}{2}\frac{(x+(s-\bar{s}))^{\beta+1} + (x-(s-\bar{s}))^{\beta+1}}{\beta+1}\right] & \text{for } s-\bar{s} \leq x \\ K\left[(1-x)(s-\bar{s})^\beta + \frac{1}{2}\frac{(x+(s-\bar{s}))^{\beta+1} - ((s-\bar{s})-x)^{\beta+1}}{\beta+1}\right] & \text{for } s-\bar{s} > x \end{cases}. \end{aligned}$$

[28] We perform the whole analysis for $t \geq \bar{s}$, as any statement that holds at some distance d to the right of the norm, holds also at distance d to the left of it, due to the symmetry of the distribution of actions around $\bar{s}$ in the equilibrium we check.

Letting $\sigma \equiv s - \bar{s}$, we can rewrite $P(\sigma)$ as

$$P(\sigma) = \begin{cases} K\left[(1-x)\,\sigma^{\beta} + \frac{1}{2}\frac{(x+\sigma)^{\beta+1}+(x-\sigma)^{\beta+1}}{\beta+1}\right] & \text{for } \sigma \leq x \\ K\left[(1-x)\,\sigma^{\beta} + \frac{1}{2}\frac{(x+\sigma)^{\beta+1}-(\sigma-x)^{\beta+1}}{\beta+1}\right] & \text{for } \sigma > x \end{cases}. \tag{14}$$

When type t chooses action s she feels the loss

$$L = (t-s)^{\alpha} + P(\sigma). \tag{15}$$

The difference in loss between fully conforming to the norm ($s = \bar{s}$) and following her heart ($s = t$) is thus

$$\begin{aligned} \Delta L &\equiv L(s=\bar{s}) - L(s=t) \\ &= (t-\bar{s})^{\alpha} + K\left[\frac{x^{\beta+1}}{\beta+1} - (1-x)\,(t-\bar{s})^{\beta} - \frac{1}{2}\frac{(x+(t-\bar{s}))^{\beta+1} + (x-(t-\bar{s}))^{\beta+1}}{\beta+1}\right], \end{aligned} \tag{16}$$

Lemma 6 *If K is sufficiently large, there exists a type $t = \bar{s} + x$, with $x \leq 1$, who*

1. *is indifferent between choosing $s = \bar{s} + x$ and $s = \bar{s}$.*
2. *has no inner solution.*

Proof. *1) Let $\tau \equiv t - \bar{s}$. Type t with $\tau = x$ is indifferent between $s = \bar{s}$ and $s = t$ if ΔL, as given in (16), equals zero, i.e., if*

$$x^{\alpha} = K\left[-\frac{x^{\beta+1}}{\beta+1} + (1-x)\,x^{\beta} + \frac{1}{2}\frac{(2x)^{\beta+1}}{\beta+1}\right],$$

hence

$$K = \frac{x^{\alpha-\beta}}{1 - \frac{2+\beta-2^{\beta}}{\beta+1}x}. \tag{17}$$

Investigating $K(x)$, we get that $\lim_{x\to 0} K = \infty$ and $\lim_{x\to 1} K = \frac{\beta+1}{2^{\beta}-1}\,(>0)$. Differentiating K with respect to x we get that

$$\begin{aligned} \frac{dK}{dx} &= \frac{x^{\alpha-\beta}}{\left(1 - \frac{2+\beta-2^{\beta}}{\beta+1}x\right)^2}\left[\left(1 - \frac{2+\beta-2^{\beta}}{\beta+1}x\right)\frac{\alpha-\beta}{x} + \frac{2+\beta-2^{\beta}}{\beta+1}\right] \\ &= \frac{x^{\alpha-\beta-1}}{(\beta+1)\left(1 - \frac{2+\beta-2^{\beta}}{\beta+1}x\right)^2}\left[(\alpha-\beta)(\beta+1) + \left(2+\beta-2^{\beta}\right)(1-\alpha+\beta)\,x\right], \end{aligned}$$

and so $\frac{dK}{dx}=0$ *only at one point,* $x_0 \equiv \frac{(\beta-\alpha)(1+\beta)}{(1+\beta-\alpha)\left(2+\beta-2^{\beta}\right)}$.

The sign of $\frac{dK}{dx}$ *is the same as that of*

$$F(\alpha,\beta,x) \equiv (\alpha-\beta)(\beta+1)+\left(2+\beta-2^{\beta}\right)(1-\alpha+\beta)\,x.$$

$F(\alpha,\beta,0)<0$ *since* $\alpha<\beta$, *and* $F(\alpha,\beta,1)=\left(2-2^{\beta}\right)(1+\beta)+\alpha(2^{\beta}-1)$, *which is positive for any* $\alpha\in(0,1)$. *It follows that* $F(\alpha,\beta,x)$ *changes sign as* x *goes from* 0 *to* 1, *and so does* $\frac{dK}{dx}$. *Thus* $x_0\in\,]0,1[$. *Hence,* K *is U-shaped in* x*: it starts at* ∞ *and is decreasing until* x_0 *and increases thereafter. Substituting* x_0 *into* K, *we get* $K(x_0)=(1+\beta-\alpha)(x_0)^{\alpha-\beta}$. *Thus, if* $K\geq(1+\beta-\alpha)(x_0)^{\alpha-\beta}$, *there exists a type* t *with* $\tau=x$ *who is indifferent between choosing* $s=\bar{s}+x$ *and* $s=\bar{s}$, *where* $x(K)$ *is implicitly given by equation (17).*

2) We showed that if $K\geq(1+\beta-\alpha)(x_0)^{\alpha-\beta}$, *then there exists a type* t *with* $\tau=x$, *s.t.* x *satisfies (17), who is indifferent between choosing* $s=t$ $(=\bar{s}+x)$ *and* $s=\bar{s}$. *In order to show that this type has no inner solution, we will now show that* $L''<0$ *for any action in the range* $]\bar{s},\bar{s}+x[$.

Differentiating L *from equation (15) twice with respect to* σ *(after substituting* $t-s=\tau-\sigma$*) we get*

$$L'=-\alpha(\tau-\sigma)^{\alpha-1}+\begin{cases}K\left[\beta(1-x)\sigma^{\beta-1}+\frac{1}{2}\left[(x+\sigma)^{\beta}-(x-\sigma)^{\beta}\right]\right] & \text{for } \sigma\leq x\\ K\left[\beta(1-x)\sigma^{\beta-1}+\frac{1}{2}\left[(x+\sigma)^{\beta}-(\sigma-x)^{\beta}\right]\right] & \text{for } \sigma\geq x\end{cases} \tag{18}$$

$$L''=\alpha(\alpha-1)(\tau-\sigma)^{\alpha-2}+\begin{cases}K\left[\beta(\beta-1)(1-x)\sigma^{\beta-2}+\frac{\beta}{2}\left[(x+\sigma)^{\beta-1}+(x-\sigma)^{\beta-1}\right]\right] & \text{for } \sigma\leq x\\ K\left[\beta(\beta-1)(1-x)\sigma^{\beta-2}+\frac{\beta}{2}\left[(x+\sigma)^{\beta-1}-(\sigma-x)^{\beta-1}\right]\right] & \text{for } \sigma\geq x\end{cases} \tag{19}$$

Plugging in $\tau=x$, *and considering naturally only actions with* $\sigma\leq\tau=x$, *we get that*

$$\begin{aligned}L''(\tau=x)&<\alpha(\alpha-1)(x-\sigma)^{\alpha-2}+K\left[\beta(\beta-1)(1-x)\sigma^{\beta-2}+\beta(x-\sigma)^{\beta-1}\right]\\&<\alpha(\alpha-1)(x-\sigma)^{\alpha-2}+K\beta(x-\sigma)^{\beta-1}\equiv G(\sigma).\end{aligned}$$

Next, $G(\sigma)<0$ *if and only if*

$$K\beta(x-\sigma)^{\beta-\alpha+1}<\alpha(1-\alpha).$$

The LHS is decreasing in σ *hence it is thus sufficient to show that* $G(0)<0$. *Using the connection between* K *and* x *as given in (17), for* $G(0)<0$ *one*

needs that

$$\frac{\beta x}{1-\frac{2+\beta-2^{\beta}}{\beta+1}x}<\alpha\left(1-\alpha\right).$$

Noting that $1-\frac{2+\beta-2^{\beta}}{\beta+1}x>0$, *we get that* $G(0)<0$ *if*$x<\frac{1}{\frac{\beta}{\alpha(1-\alpha)}+\frac{2+\beta-2^{\beta}}{\beta+1}}$.

Hence, if K *is sufficiently large so that* $x<\frac{1}{\frac{\beta}{\alpha(1-\alpha)}+\frac{2+\beta-2^{\beta}}{\beta+1}}$ *(recall that* $\lim_{x\to 0}K=\infty$*), then type* t *with* $\tau=x$ *has no inner solution.* ■

Lemma 7 $s^{*}\left(t\right)=t$ *for all* t *with* $\tau<x$.

Proof. *Sufficient conditions for this statement are 1)* $L\left(s=t\right)<L\left(s=\bar{s}\right)$ *and 2) no inner solution for* t *with* $\tau\in[0,x[$.

1) When $\tau<x$, $\Delta L=L\left(s=\bar{s}\right)-L\left(s=t\right)$ *is given by equation (16). We will show that* $\Delta L>0$ *by showing that* $\frac{\Delta L}{x^{\alpha}}>0$. *Define* $c\equiv\tau/x\in[0,1]$. *We then have*

$$f(x,c)\equiv\frac{\Delta L}{x^{\alpha}}=c^{\alpha}+\frac{K}{x^{\alpha-\beta}}\left[\frac{x}{\beta+1}-\left(1-x\right)c^{\beta}-\frac{x}{2}\frac{\left(1+c\right)^{\beta+1}+\left(1-c\right)^{\beta+1}}{\beta+1}\right]$$

$$=c^{\alpha}+\left(1-\frac{2+\beta-2^{\beta}}{\beta+1}x\right)^{-1}\left[\frac{x}{\beta+1}-\left(1-x\right)c^{\beta}-\frac{x}{2}\frac{\left(1+c\right)^{\beta+1}+\left(1-c\right)^{\beta+1}}{\beta+1}\right]$$

$$=\left(1-\frac{2+\beta-2^{\beta}}{\beta+1}x\right)^{-1}\left(\left[\frac{1}{\beta+1}+c^{\beta}-\frac{1}{2}\frac{\left(1+c\right)^{\beta+1}+\left(1-c\right)^{\beta+1}}{\beta+1}-\frac{2+\beta-2^{\beta}}{\beta+1}c^{\alpha}\right]x+c^{\alpha}-c^{\beta}\right).$$

We will show that $f(x,c)>0$ *for every* x *and any* $c\in[0,1]$. *Define the part in the squared brackets as*

$$g(c)\equiv1+\left(\beta+1\right)c^{\beta}-\frac{1}{2}\left[\left(1+c\right)^{\beta+1}+\left(1-c\right)^{\beta+1}\right]-\left(2+\beta-2^{\beta}\right)c^{\alpha}.$$

Then

$$f(x,c)=\left(1-\frac{2+\beta-2^{\beta}}{\beta+1}x\right)^{-1}\left(\frac{x}{\beta+1}g(c)+c^{\alpha}-c^{\beta}\right)$$

$$=-\frac{g(c)}{2+\beta-2^{\beta}}+\left(1-\frac{2+\beta-2^{\beta}}{\beta+1}x\right)^{-1}\left(2+\beta-2^{\beta}\right)^{-1}\left[g(c)+\left(c^{\alpha}-c^{\beta}\right)\left(2+\beta-2^{\beta}\right)\right].$$

Finally, let

$$h(c)\equiv g(c)+\left(c^{\alpha}-c^{\beta}\right)\left(2+\beta-2^{\beta}\right).$$

We then have

$$h(0) = g(0) = 0,$$
$$h(1) = g(1) = 0,$$

and by differentiating $h(c)$ twice with respect to c we get

$$\begin{aligned}\frac{h''(c)}{\beta+1} &= \beta(\beta-1)c^{\beta-2} - \frac{1}{2}\beta\left((1+c)^{\beta-1} + (1-c)^{\beta-1}\right) + -\beta(\beta-1)c^{\beta-2}\frac{2+\beta-2^{\beta}}{\beta+1} \\ &= \beta(\beta-1)c^{\beta-2}\left[1 - \frac{2+\beta-2^{\beta}}{\beta+1}\right] - \frac{1}{2}\beta\left((1+c)^{\beta-1} + (1-c)^{\beta-1}\right) < 0,\end{aligned}$$

and so $h(c) \geq 0$ in the range $c \in [0,1]$. Consequentially, $f(x,c)$ increases in x for every given $c \in [0,1]$, and so $f(x,c) \geq f(0,c)$, where

$$f(0,c) = c^{\alpha} - c^{\beta} > 0, \ \forall c \in]0,1[\,.$$

This proves that $\Delta L = f(x,c)x^{\alpha}$ is positive, i.e., $L(s=t) < L(s=\bar{s})$ for all t with $\tau < x$.

2) From equation (19) it is easy to see that for any given σ, the function L'' increases in τ, and so reaches it maximum for $\tau = x$. In the second part of the proof of Lemma 6 we saw that L'' is negative (for any σ) whenever $x < \frac{1}{\frac{\beta}{\alpha(1-\alpha)} + \frac{2+\beta-2^{\beta}}{\beta+1}}$. This implies that there is no inner solution for t with $\tau \in [0,x[$. ■

Lemma 8 *$s^*(t) = \bar{s}$ for all t with $\tau > x$.*

Proof. *To prove the lemma we show that the following sufficient conditions hold for all t with $\tau > x$: 1) $L(s=t) > L(s=\bar{s})$ .2) No inner solution in the range $\sigma \in [x,\tau]$. 3) No inner solution in the range $\sigma \in (0,x)$.*

1) $L(s=\bar{s}) = \tau^{\alpha} + K\frac{x^{\beta+1}}{\beta+1}$ as before, but for $\sigma \geq x$ we have:

$$L = (t-s)^{\alpha} + K\left[(1-x)\sigma^{\beta} + \frac{1}{2}\frac{(x+\sigma)^{\beta+1} - (\sigma-x)^{\beta+1}}{\beta+1}\right], \qquad (20)$$

and so

$$\begin{aligned}\Delta L &\equiv L(s=\bar{s}) - L(s=t) \\ &= \tau^{\alpha} + K\left[\frac{x^{\beta+1}}{\beta+1} - (1-x)\tau^{\beta} - \frac{1}{2}\frac{(x+\tau)^{\beta+1} - (\tau-x)^{\beta+1}}{\beta+1}\right].\end{aligned}$$

Define $c \equiv \tau/x$ (>1). *We then have*

$$f(x,c) \equiv \frac{\Delta L}{x^{\alpha}} = c^{\alpha} + \frac{K}{x^{\alpha-\beta}}\left[\frac{x}{\beta+1} - (1-x)\,c^{\beta} - \frac{x}{2}\frac{(1+c)^{\beta+1} - (c-1)^{\beta+1}}{\beta+1}\right]$$

$$= c^{\alpha} + \left(1 - \frac{2+\beta-2^{\beta}}{\beta+1}x\right)^{-1}\left[\frac{x}{\beta+1} - (1-x)\,c^{\beta} - \frac{x}{2}\frac{(1+c)^{\beta+1} - (c-1)^{\beta+1}}{\beta+1}\right]$$

$$= \left(1 - \frac{2+\beta-2^{\beta}}{\beta+1}x\right)^{-1}\left(\left[\frac{1}{\beta+1} + c^{\beta} - \frac{1}{2}\frac{(1+c)^{\beta+1} - (c-1)^{\beta+1}}{\beta+1} - \frac{2+\beta-2^{\beta}}{\beta+1}c^{\alpha}\right]x + c^{\alpha} - c^{\beta}\right).$$

Now let

$$g(c) \equiv 1 + (\beta+1)\,c^{\beta} - \frac{1}{2}\left[(1+c)^{\beta+1} - (c-1)^{\beta+1}\right] - \left(2+\beta-2^{\beta}\right)c^{\alpha}.$$

Then

$$f(x,c) = \left(1 - \frac{2+\beta-2^{\beta}}{\beta+1}x\right)^{-1}\left(\frac{x}{\beta+1}g(c) + c^{\alpha} - c^{\beta}\right)$$

$$= -\frac{g(c)}{2+\beta-2^{\beta}} + \left(1 - \frac{2+\beta-2^{\beta}}{\beta+1}x\right)^{-1}\left(2+\beta-2^{\beta}\right)^{-1}\left[g(c) + \left(c^{\alpha} - c^{\beta}\right)\left(2+\beta-2^{\beta}\right)\right].$$

We will show now that

$$h(c) \equiv g(c) + \left(c^{\alpha} - c^{\beta}\right)\left(2+\beta-2^{\beta}\right)$$

$$= 1 - \frac{1}{2}(1+c)^{\beta+1} + \frac{1}{2}(c-1)^{\beta+1} + c^{\beta}\left(2^{\beta}-1\right).$$

is negative, and so $f(x,c)$ *decreases in* x. *Define* $r \equiv \frac{1}{c}$, *so that* $r \in (0,1)$. *We then can define a new function* $z(r)$ *such that*

$$z(r) \equiv r^{\beta+1}h = r^{\beta+1} - \frac{1}{2}(1+r)^{\beta+1} + \frac{1}{2}(1-r)^{\beta+1} + r\left(2^{\beta}-1\right).$$

We have $z(0) = z(1) = 0$, *and*

$$z''(r) = (\beta+1)\,\beta\left[r^{\beta-1} - \frac{1}{2}(1+r)^{\beta-1} + \frac{1}{2}(1-r)^{\beta-1}\right]$$

$$> \frac{1}{2}(\beta+1)\,\beta\left[(1-r)^{\beta-1} - (1+r)^{\beta-1}\right] > 0,$$

and so $z(r) < 0$ *in the range* $r \in (0,1)$*, i.e.,* $h(c) < 0$ *in the range* $c > 1$*. Consequentially,* $f(x,c)$ *decreases in* x *at this range, and so* $f(x,c) < f(0,c)$*, where*

$$f(0,c) = c^{\alpha} - c^{\beta} < 0, \ \forall c > 1.$$

This proves that $\Delta L = f(x,c)x^{\alpha}$ *is negative, i.e.,* $L(s=t) > L(s=\bar{s})$ *for all* t *with* $\tau > x$*.*

2) From equation (19) for $\sigma \geq x$ *we get that* L'' *is a sum of three negative elements, hence* $L'' < 0$*, i.e., no inner solution in that range.*

3) Type t *with* $\tau > x$ *has no inner solution in the range* $\sigma \in (0,x)$ *if* $L(\sigma,\tau) - L(0,\tau) > 0$ *for any* $\sigma \in [0,x]$*. From equation (18) for* $\sigma \leq x$ *we have*

$$L' = -\alpha(\tau-\sigma)^{\alpha-1} + K\left[\beta(1-x)\sigma^{\beta-1} + \frac{1}{2}\left[(x+\sigma)^{\beta} - (x-\sigma)^{\beta}\right]\right],$$

and so $\frac{\partial L'}{\partial \tau} = -\alpha(\alpha-1)(\tau-\sigma)^{\alpha-1} > 0$*. This implies that, when* $\tau > x$*,* $L(\sigma,\tau) - L(\sigma,\tau) > L(\sigma,x) - L(0,x) \geq 0$ *for any* $\sigma \in [0,x]$*, where the second inequality follows from Lemma 6.* ■

C.3.2 Sentence 2: Properties

We will prove the statement for types to the right of the norm, showing that there exists a $t_{\max}$ such that the set of types $]\bar{s}, t_{\max}]$ do not fully conform. We rewrite the variables to $\sigma \equiv s - \bar{s}$ and $\tau \equiv t - \bar{s}$. First note that a necessary condition for full conformity of type t is that she prefers it over speaking her mind: $L(0;\tau) = P(\sigma = 0) + D(\tau) \leq L(\tau;\tau) = P(\tau)$. Suppose $\tau_{\max}$ is sufficiently small. Then this condition is equivalent to $\lim_{\sigma \to 0^+} P'(\sigma) \geq \lim_{\sigma \to \tau^-} D'(\tau - \sigma) = \lim_{\sigma \to \tau^-} \alpha(\tau-\sigma)^{\alpha-1}$. Hence, for the types closest to the norm to fully conform, $P'(\sigma)$ has to approach infinity faster than $\alpha(\tau-\sigma)^{\alpha-1}$. We will now show that this cannot be the case. The fastest way in which $P'(\sigma)$ may approach infinity when $\sigma \to 0^+$ is when all individuals choose $\sigma = 0$, because then $\lim_{\sigma \to 0^+} p'(\sigma)$ is maximized for each pairwise pressure p. Suppose this is indeed the case, i.e., all types fully conform. Then $\lim_{\sigma \to 0^+} P'(\sigma) = \lim_{\sigma \to 0^+} \beta K \sigma^{\beta-1}$, which approaches infinity slower than $\alpha(\tau-\sigma)^{\alpha-1}$ when $\alpha < \beta$. This shows that the types closest to the norm will not fully conform. Hence, since in any single-norm equilibrium there must exist a positive mass of individuals who do fully conform, these have to be some types sufficiently far from the norm.

D Appendix: Alienating societies

D.1 Proof of Lemma 1

The minimization problem of the individual is

$$\min_{s} L(s;t;s') = P(s;s') + |s-t|^{\alpha}. \tag{21}$$

Suppose a single norm exists with a share x stating it. Then

$$P(s) = \begin{cases} K \text{ if } s \neq \bar{s} \\ (1-x)K \text{ if } s = \bar{s} \end{cases}. \tag{22}$$

Therefore $L(s;t;s')$ is increasing in $|s-t|$ except potentially at $s=\bar{s}$, where $P(s) < K$. Thus it is immediate that for each type t, $s^*(t)$ will be either t or $\bar{s}$. Moreover, it is immediate that $s^*(t) = t$ if and only if xK, the difference between $P(t)$ and $P(\bar{s})$, falls below $|t-\bar{s}|^{\alpha}$, thus follows the lemma.

D.2 Proof of Lemma 2

If $y \leq 1-|\bar{s}|$, the norm is sufficiently centered so that y types on each side follow the norm, which implies $x = y$. When $1-|\bar{s}| < y \leq 1+|\bar{s}|$, the norm is sufficiently biased, say to the left, so that there are no longer y types to the left of the norm following the norm. Then, the total number of individuals declaring the norm is the distance from -1 to $\bar{s}$ on the left and y types on the right. It then follows that the share of norm followers is $x = (y+1-|\bar{s}|)/2$. Finally, when $y > 1+|\bar{s}|$, we get that even the type who is the furthest away from the norm (i.e. at distance $1+|\bar{s}|$ from it) follows it, implying that all types follow the norm.

D.3 Proof of Proposition 2

Since Lemma 1 implies that, given a single norm with a share x of followers, $s^*(t)$ is according to (8), a necessary and sufficient condition for this $s^*(t)$ to be the distribution of actions in a single-norm equilibrium is that $x(y)$ that is obtained from this distribution of actions in Lemma 2 would equal the value of x that was initially assumed in Lemma 1 for creating this particular $s^*(t)$. This is more conveniently written as a dynamic process, where the requirement is to have $x_{i+1}(y_{i+1}(x_i)) = x_i$. Using (7) and (10) we can write

$$x_{i+1} = f(x_i;K,|\bar{s}|) \equiv \begin{cases} (x_iK)^{1/\alpha} \text{ if } (x_iK)^{1/\alpha} \leq 1-|\bar{s}| \\ \frac{(x_iK)^{1/\alpha}+1-|\bar{s}|}{2} \text{ if } 1-|\bar{s}| < (x_iK)^{1/\alpha} < 1+|\bar{s}| \\ 1 \text{ if } (x_iK)^{1/\alpha} \geq 1+|\bar{s}| \end{cases}. \tag{23}$$

We start by proving parts (1) and (2) of the proposition for the case of $\alpha \geq 1$. If one of the following holds: (1) $\alpha > 1$; (2) $\alpha = 1$, $|\bar{s}| < 1$ and $K \geq 1$; or (3) $\alpha = 1$, $|\bar{s}| = 1$ and $K \geq 2$; then $\lim_{x_i \to +0} f'(x_i; K, |\bar{s}|) \geq 1$, so that $f(x_i; K, |\bar{s}|)$ starts (weakly) above the 45 degree line. In this case, the continuity of $f(x_i; K, |\bar{s}|)$ and the fact that $f(x_i = 1) \leq 1$ imply that $f(x_i; K, |\bar{s}|)$ crosses the 45 degree line at least once in the range $x_i \in (0, 1]$, with the crossing point(s) constituting single norm EQ. Alternatively, if $\alpha = 1$ and either (1) $|\bar{s}| < 1$ and $K < 1$; or (2) $|\bar{s}| = 1$ and $K < 2$; then $f(x_i; K, |\bar{s}|)$ is linear in parts and $\lim_{x_i \to +0} f'(x_i; K, |\bar{s}|) < 1$, so that the first linear part is below the 45 degree line. As the slope of $f(x_i; K, |\bar{s}|)$ only decreases when moving from the first linear part to the second and from the second to the third, we get that $f(x_i; K, |\bar{s}|)$ is below the 45 degree line in $x_i \in (0, 1]$, in which case there is no single-norm equilibrium with a strictly positive share x of norm followers. All in all we get that when $\alpha > 1$, $K_{\min}(|\bar{s}|) = 0$; when $\alpha = 1$ and $|\bar{s}| < 1$, $K_{\min}(|\bar{s}|) = 1$; and when $\alpha = 1$ and $|\bar{s}| = 1$, $K_{\min}(|\bar{s}|) = 2$. These values of $K_{\min}(|\bar{s}|)$ are independent of $|\bar{s}|$, which concludes the proof of parts (1) and (2) of the proposition for the case of $\alpha \geq 1$.

We now proceed to proving parts (1) and (2) of the proposition for the case of $\alpha < 1$. To do so, we will now assume that a single-norm equilibrium exists at $|\bar{s}|$ and prove the existence of a value $K_{\min}(|\bar{s}|)$ such that the assumption holds if and only if $K \geq K_{\min}(|\bar{s}|)$, and that $K_{\min}(|\bar{s}|)$ is increasing in $|\bar{s}|$.

Looking at the borders between regions in equation (23), we get that if $K \geq (1 + |\bar{s}|)^{\alpha}$ then at $x_i = 1$ we are in the third region, implying that $x_{i+1}(x_i) = x_i$ at $x_i = 1$, hence a single-norm equilibrium exists (with full compliance to the norm). Otherwise, $(x_i K)^{1/\alpha} \leq K^{1/\alpha} < 1 + |\bar{s}|$, and the third region is irrelevant. Moreover, x_{i+1} in the second region is strictly smaller than 1 and so $x_i = 1$ is not an equilibrium.

Define now

$$\begin{aligned} G(x_i; K, |\bar{s}|) &\equiv x_{i+1}(x_i) - x_i = f(x_i; K, |\bar{s}|) - x_i, \\ &= \begin{cases} (x_i K)^{1/\alpha} - x_i \text{ if } (x_i K)^{1/\alpha} \leq 1 - |\bar{s}| \\ \frac{(x_i K)^{1/\alpha} + 1 - |\bar{s}|}{2} - x_i \text{ if } 1 - |\bar{s}| < (x_i K)^{1/\alpha} < 1 + |\bar{s}| \\ 1 - x_i \text{ if } (x_i K)^{1/\alpha} \geq 1 + |\bar{s}| \end{cases} \end{aligned} \tag{24}$$

which in a single-norm equilibrium equals zero for some $x_i \neq 0$. G is continuous in x_i, K and $|\bar{s}|$, with $G(0; K, |\bar{s}|) = 0$ and $G'(0; K, |\bar{s}|) < 0$, and when $K^{1/\alpha} < 1 + |\bar{s}|$ we also get that $G(1; K, |\bar{s}|) < 0$. Differentiation of G with respect to

x_i yields

$$G'(x_i;K,|\bar{s}|)=\begin{cases}\frac{1}{\alpha}K^{1/\alpha}(x_i)^{1/\alpha-1}-1 \text{ if } (x_iK)^{1/\alpha}<1-|\bar{s}| \\ \frac{1}{2\alpha}K^{1/\alpha}(x_i)^{1/\alpha-1}-1 \text{ if } 1-|\bar{s}|<(x_iK)^{1/\alpha}<1+|\bar{s}| \\ -1 \text{ if } (x_iK)^{1/\alpha}>1+|\bar{s}|\end{cases} \quad (25)$$

and

$$G''(x_i;K,|\bar{s}|)=\begin{cases}\frac{1}{\alpha}\left(\frac{1}{\alpha}-1\right)K^{1/\alpha}(x_i)^{1/\alpha-2} \text{ if } (x_iK)^{1/\alpha}<1-|\bar{s}| \\ \frac{1}{2\alpha}\left(\frac{1}{\alpha}-1\right)K^{1/\alpha}(x_i)^{1/\alpha-2} \text{ if } 1-|\bar{s}|<(x_iK)^{1/\alpha}<1+|\bar{s}| \\ 0 \text{ if } (x_iK)^{1/\alpha}>1+|\bar{s}|\end{cases} \quad (26)$$

which immediately shows G is strictly convex in the first two regions. It thus follows that when $K^{1/\alpha}<1+|\bar{s}|$, G can get a local max only at the border between these two regions, where $x_i=(1-|\bar{s}|)^{\alpha}/K$. Therefore, when $K^{1/\alpha}<1+|\bar{s}|$, there exists a single-norm equilibrium if and only if the borderline point falls within the range $[0,1]$ and G at this point is weakly positive.[29] Substituting $x_i=(1-|\bar{s}|)^{\alpha}/K$ in equation (24) yields $G=(1-|\bar{s}|)-(1-|\bar{s}|)^{\alpha}/K$, which equals 0 when $K=(1-|\bar{s}|)^{\alpha-1}$. Substituting this value of K back in x_i we get that $x_i=1-|\bar{s}|$, thus falls within the range $[0,1]$, and so there exists a single-norm equilibrium for $K=(1-|\bar{s}|)^{\alpha-1}$. If K is larger, then the value of x_i at the border between the regions is smaller (hence falls within the range $[0,1]$ too), and the value of G at this point is larger, i.e., positive.

As a result, if we let

$$K_{\min}(|\bar{s}|)\equiv\min\left\{(1-|\bar{s}|)^{\alpha-1},(1+|\bar{s}|)^{\alpha}\right\}, \quad (27)$$

then for $K<K_{\min}(|\bar{s}|)$ no single-norm equilibrium exists, while for any $K\geq K_{\min}(|\bar{s}|)$ there exists a single-norm equilibrium at $|\bar{s}|$. It is also worth noting that if $K=K_{\min}(|\bar{s}|)$, the analysis above implies that $\max_{x_i}G(x_i)=0$ (and reached either at the border between the two regions, if $K_{\min}(|\bar{s}|)=(1-|\bar{s}|)^{\alpha-1}$, or at $x_i=1$, if $K_{\min}(|\bar{s}|)=(1+|\bar{s}|)^{\alpha}$); while if $K>K_{\min}(|\bar{s}|)$, then $G(x_i)>0$ either at the borderline point or at $x_i=1$.

Finally, the fact that $K_{\min}(|\bar{s}|)$ is increasing in $|\bar{s}|$ follows directly from the fact that $(1-|\bar{s}|)^{\alpha-1}$ and $(1+|\bar{s}|)^{\alpha}$ are both increasing in $|\bar{s}|$.■

[29] Note that if the borderline point falls outside the range $[0,1]$, it means that only the first region applies, and then the convexity of G means that $G(1,K,|\bar{s}|)<0\Rightarrow G(x_i,K,|\bar{s}|)<0$ $\forall x_i\in\,]0,1]$, hence no single norm equilibrium exists (we know that $G(1,K,|\bar{s}|)<0$ because $K^{1/\alpha}<1+|\bar{s}|$).

D.4 Proof of Proposition 3

We first remind, that in the proposition we treat the unstable steady states (x_{uss}) as ones where if $x_i = x_{uss}$ then $x_{i+1} < x_{uss}$. This includes the cases where $x_{uss} = x_{conv}$. In the proof we do not make this shortcut.

Comparison method: Next, we remind how we perform the comparison of $x_{ss}(|\bar{s}|, K)$ in statement (2) of the proposition and the comparison of welfare distributions in statement (5). For given values of K and $\bar{s}$ there can be at most one stable single norm steady state if $\alpha \geq 1$,[30] and up to two different stable single norm steady states, corresponding to two different values of $x_{ss}(|\bar{s}|, K)$, if $\alpha < 1$. When there are two such steady states, Lemma 14 says that one is always "degenerate" ($x_{ss}(|\bar{s}|, K) = 1$) and one is always "non-degenerate" ($x_{ss}(|\bar{s}|, K) \in]0, 1[$) – these are $\tilde{x}$ and x_{end} respectively in Section D.4.2 below. The comparison of $x_{ss}(|\bar{s}|, K)$ with $x_{ss}(|\bar{s}'|, K)$ in statement (2) of the proposition is thus applied as follows: When there exist two $x_{ss}(|\bar{s}|, K)$ and two $x_{ss}(|\bar{s}'|, K)$, the proposition compares the non-degenerate with each other and the degenerate with each other; when there exist two steady states for $\bar{s}$ and one for $\bar{s}'$ (or vice versa), the proposition compares $\max\{x_{ss}(|\bar{s}|, K)\}$ with $\max\{x_{ss}(|\bar{s}'|, K)\}$; finally, when there is only one steady state for each norm, the proposition compares the unique $x_{ss}(|\bar{s}|, K)$ with the unique $x_{ss}(|\bar{s}'|, K)$.

Moreover, whenever there exist two stable steady states for a given norm, the welfare distribution in the degenerate steady state first-order stochastically dominates the welfare distribution in the non-degenerate steady state (see Lemma 19 below). Hence, we apply the same comparison rule to the comparison of welfare distributions in statement (5) of the proposition.

Proof for the case $\alpha \geq 1$: Statements (1)-(4) are proved in Section D.4.1 and statement (5) is proved by Lemma 20.

Proof for the case $\alpha < 1$:

- **Statement 1)**: The 'if' part follows from Lemma 14. As for the 'only if' part, we show in the proof of Proposition 2 that the function G is strictly positive at some point iff $K > K_{\min}$. Hence, if $K \leq K_{\min}$, then $\forall x_i$ we have $x_{i+1} \leq x_i$, which means that there can be no convergence from the left to any steady state, implying that a stable steady state with a single norm cannot exist.

- **Statement 2)**: Lemma 14 shows that at most two stable single norm steady states may exist: a degenerate and a non-degenerate (the notation of x with various ornaments is defined in equation (28)). Lemma 15

[30] When $\alpha > 1$ this follows from the fact that in this case $f'(x_i; K, |\bar{s}|)$ is decreasing at every x_i (see equation 23). The case $\alpha = 1$ is analyzed separately under Section D.4.1 below.

implies that when comparing x_{ss} of two norms $\bar{s}$ and $\bar{s}'$ (where $|\bar{s}| \leq |\bar{s}'|$) in statement (2) of the proposition, we need to compare only the non-degenerate stable steady states. To see why, note that one of the following two scenarios must hold: (i) there exists a degenerate stable steady state for $\bar{s}'$, in which case (by Lemma 15) the maximal share of x_{ss} under both norms is 1, and so we have to compare only the non-degenerate stable steady states, if both exist; (ii) there does not exist a degenerate stable steady state for $\bar{s}'$, in which case either there exists a degenerate stable steady state for $\bar{s}$, hence it is immediate that $\max\{x_{ss}(|\bar{s}|, K)\} = 1 > \max\{x_{ss}(|\bar{s}'|, K)\}$, or there does not exist a degenerate stable steady state for $\bar{s}$, in which case we again have to compare only the non-degenerate stable steady states. Thus, it is sufficient to compare $x_{ss}(|\bar{s}|, K)$ in the non-degenerate stable steady states, if they exist. Statement 2 then follows from part (3) of Lemma 11.

- **Statement 3)**: Follows from parts (1)-(3) of Lemma 13 (note that $\dot{x}$, $\hat{x}$ and $\check{x}$ are defined in equation (28)).
- **Statement 4)**: Follows from parts (4) and (5) of Lemma 13.
- **Statement 5)**: Follows from Lemma 20.

D.4.1 The case of $\alpha \geq 1$

We here prove statements (1)-(4) of Proposition 3 for the case $\alpha \geq 1$. We prove $\alpha > 1$ and $\alpha = 1$ separately.

$\boldsymbol{\alpha > 1}$

When $\alpha > 1$ we get by (23) that $\lim_{x_i \to +0} f'(x_i; K, |\bar{s}|) = K/\alpha \lim_{x_i \to +0} (x_i K)^{1/\alpha - 1} = \infty$, implying that there is convergence to the single-norm equilibrium (whose existence for every $K > K_{\min}(|\bar{s}|) = 0$ was shown in the proof of Proposition 2) from every $x_i > 0$ (this proves statement 1). It thus follows that in this case $x_{conv}(|\bar{s}|) = 0$ and so (i) $x_{conv}(|\bar{s}|)$ is independent of $|\bar{s}|$ and K; (ii) $x_{ss}(|\bar{s}|) > x_{conv}(|\bar{s}|)$; and (iii) if $0 \leq x_i \leq x_{conv}(|\bar{s}|)$, then it must be the case that $x_i = 0$ and so $x_{i+1} = x_i = 0$, i.e., there is convergence to a stable steady state where each type follows her heart ($x_{ss}(|\bar{s}|) = 0$). This proves statements 3 and 4. Now note that since $f'(x_i; K, |\bar{s}|)$ is decreasing at every x_i (see equation 23) there can exist at most one stable steady state for each combination of $\bar{s}$ and K. Increasing $|\bar{s}|$ has the effect of decreasing $x_{ss}(|\bar{s}|)$, as it everywhere weakly decreases x_{i+1} as a function of x_i (by increasing region (2) in equation (23), and since the function f in this region is smaller the larger is $|\bar{s}|$). This proves statement 2.

$\alpha = 1$

When $\alpha = 1$ we have two separate cases to consider. The first one is when $|\bar{s}| < 1$. Here $K_{\min}(|\bar{s}|)$ was shown to equal 1 (see the proof of Proposition 2). Here the function f is piecewise linear, where for $K < 1$ it stays below the 45 degree line (see the proof of Proposition 2) and so there is no single-norm equilibrium; and for $K > 1$ it stays above the 45 degree line until it reaches 1 and stays there (see equation 23), implying a single-norm equilibrium at $x_i = 1$. Thus $x_{ss}(|\bar{s}|) = 1$ (hence independent of $|\bar{s}|$, which proves statement (2)), and there exists a stable steady state if and only if $K > K_{\min}(|\bar{s}|) = 1$ with a share of followers $x_{ss}(|\bar{s}|)$ (if $K = 1$ the function f lies on the 45 line degree in the first region, and so there is a continuum of steady states but none is stable). It follows that if and only if $K > 1$ then there is convergence to $x_{ss}(|\bar{s}|)$ from any $x_i > 0$, which proves statements (1) and (3). Hence, $x_{conv}(|\bar{s}|) = 0$, which is independent of $|\bar{s}|$ and K, which proves statement (4). The second case is when $|\bar{s}| = 1$ (and it was shown in the proof of Proposition 2 that $K_{\min}(|\bar{s}|) = 2$). Here the function f (see equation 23) starts immediately in region (2), and is above the 45 degree line if and only if $K > 2$. The same arguments used in proving the previous case apply here, with $x_{ss}(|\bar{s}|) = 1$ and $x_{conv}(|\bar{s}|) = 0$.

D.4.2 The case of $\alpha < 1$

The proof of statements (1)-(4) of Proposition 3 for the case $\alpha < 1$ builds on a few preliminary results and auxiliary lemmas which are presented here.

Note first that Lemmas 1 and 2 show that alienation recreates alienation. Hence, the full dynamics can be described by the dynamics of x, the share of norm followers, as given in equation (24). Following equation (26), it is straightforward to see that $x_{i+1} = f(x_i; K, |\bar{s}|)$ is convex within each of the first two regions and has a kink at the border between the regions. Together with $G'(0; K, |\bar{s}|) < 0$ (see equation (25)), this means we can define the following values of x_{i+1} (see Figure 7) that exhaust the possible fix points, and which will be used throughout the upcoming lemmas.

$$\hat{x} \equiv \{x_i : x_{i+1} = x_i \text{ and } x_i \text{ is in the first region}\} \tag{28}$$

$$\check{x} \equiv \{x_i : x_{i+1} = x_i \text{ and } x_i \text{ is in the second region and } G' > 0\}$$

$$\tilde{x} \equiv \{x_i : x_{i+1} = x_i \text{ and } x_i \text{ is in the second region and } G' < 0\}$$

$$\ddot{x} \equiv \left\{x_i : (x_i K)^{1/\alpha} = 1 - |\bar{s}|\right\} \text{ (i.e., at the border between regions (1) and (2))}$$

$$\dot{x} \equiv \left\{x_i : (x_i K)^{1/\alpha} = 1 - |\bar{s}| \text{ and } G(x_i) = 0 \text{ and } G_2'(x_i) < 0\right\}$$

$$x_{end} \equiv \{x_i : x_{i+1} = x_i = 1\} \text{ (i.e., at the endpoint)}$$

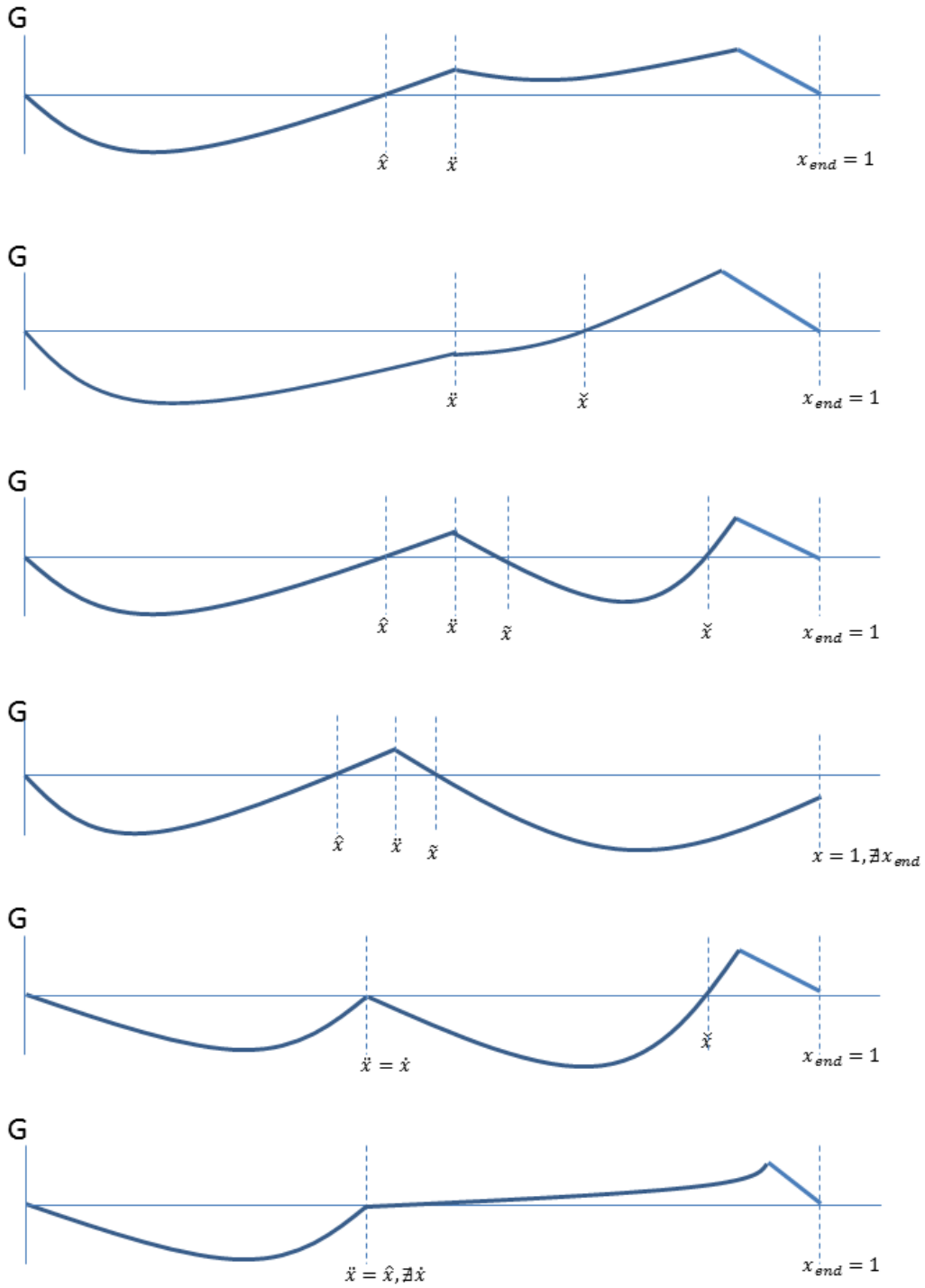

Figure 7: Some variations of the G function of equation (24), depicting the potential fix points defined in equation (28). Note that these variations of G are not exhaustive but are only meant to complement the proof.

Note that when $G(\ddot{x}) = 0$ then either $G_2'(x_i) < 0$, in which case $\ddot{x} = \dot{x}$, or $G_2'(x_i) > 0$.

Lemma 9 *Consider a given x_i. Then $G'(x_i : x_i < \ddot{x}) > G'(x_i : x_i > \ddot{x})$.*
Proof. *Let G_1, G_2 and G_3 denote the values of G in regions (1), (2) and (3) respectively. When $x_i < \ddot{x}$, G_1 applies, and when $x_i > \ddot{x}$, G_2 applies. Then for a given x_i, $G_1' = \frac{1}{\alpha}K^{1/\alpha}(x_i)^{1/\alpha-1} - 1 > \frac{1}{2\alpha}K^{1/\alpha}(x_i)^{1/\alpha-1} - 1 = G_2'$.* ■

Lemma 10 *G' is weakly falling in $|\bar{s}|$ for any $x_i < (1+|\bar{s}|)^{\alpha}/K$.*
Proof. *When $x_i < (1+|\bar{s}|)^{\alpha}/K$ we are in region (1) or region (2) of equation (25). Here, $\frac{dG_1'}{d|\bar{s}|} = \frac{dG_2'}{d|\bar{s}|} = 0$. Moreover, $\ddot{x} = (1-|\bar{s}|)^{\alpha}/K$ decreases in $|\bar{s}|$. This implies that if $|\bar{s}|$ increases, region (2) expands at the expense of region (1). Then, by Lemma 9, we get that G' is weakly falling in $|\bar{s}|$.* ■

Lemma 11 *1) If $\hat{x}$ exists then it is independent of $|\bar{s}|$. 2) If $\check{x}$ exists then it is weakly increasing in $|\bar{s}|$. 3) If $\tilde{x}$ exists it is weakly decreasing in $|\bar{s}|$.*
Proof. *1) By definition $\hat{x}$ is in region 1. Hence G_1 applies. Since G_1 is independent of $|\bar{s}|$ so must $\hat{x}$ be. 2) By definition $\check{x}$ is in region 2. Lemma 10 together with $G(0) = 0$ imply that G is weakly falling in $|\bar{s}|$ in region 1 and 2. Combined with the fact that $G'(\check{x}) > 0$ (by definition) this implies $\check{x}$ (if it exists) is weakly increasing in $|\bar{s}|$. 3) Same logic as part 2 but now with $G'(\tilde{x}) < 0$.* ■

Lemma 12 *If $\exists\hat{x}$ for some $|\bar{s}|$ then $\exists\hat{x}$ for any $|\bar{s}'| < |\bar{s}|$.*
Proof. *G_1 is independent of $|\bar{s}|$. Then the fact that $|\bar{s}'| < |\bar{s}|$ implies that region (1) is broader under $|\bar{s}'|$, so if $\exists\hat{x}$ for some $|\bar{s}|$ then $\exists\hat{x}$ for any $|\bar{s}'| < |\bar{s}|$.*
■

Lemma 13 *Suppose $K > K_{\min}$. Let $x_{conv} \equiv \{x_i : x_i = \min\{\hat{x}, \check{x}\}\}$ (when $\hat{x}$ or $\check{x}$ or both exist). Then:*

1. *If $x_i > x_{conv}(|\bar{s}|)$ there is convergence to a stable single norm steady state where a share $x_{ss}(|\bar{s}|) > x_{conv}(|\bar{s}|)$ of the population state $\bar{s}$.*
2. *Otherwise, provided that $\nexists\dot{x}$, if $0 \leq x_i < x_{conv}(|\bar{s}|)$, there is convergence to a stable steady state where each type follows her heart ($x_{ss}(|\bar{s}|) = 0$).*
3. *Furthermore, if $\exists\dot{x}$, then when $0 < x_i < \dot{x}$ there is convergence to a stable steady state where each type follows her heart ($x_{ss}(|\bar{s}|) = 0$), and when $\dot{x} \leq x_i < x_{conv}$ there is convergence to an unstable single norm steady state where a share $\dot{x}$ state the norm.*[31]

[31] In line with our general treatment of unstable steady states as converging to less conformity, statement (3) in the proposition treats this special case as one where x_i, upon reaching $\dot{x}$, only passes through it and continues to $x_{ss} = 0$.

4. x_{conv} *increases in* $|\bar{s}|$.

5. x_{conv} *decreases in* K.

Proof. *We start with statement 2)* $G'(\hat{x}) > 0$ *since* $G_1(0) = 0$, $G_1'(0) < 0$ *and* G_1 *is convex.* $G'(\check{x}) > 0$ *by definition. This implies* $\hat{x}$ *and* $\check{x}$ *are unstable steady states. Furthermore, they are the only unstable states.*[32] *Hence, if* $\hat{x}$ *exists, it must be the smallest strictly positive steady state, and so* $G_1(0) = 0$ *and* $G_1'(0) < 0$ *imply that* $\forall x_i < \hat{x} = x_{conv}$ *we have* $G(x_i) < 0$*, i.e.,* $x_{i+1} < x_i$*. Otherwise there is no steady state in the first region, in which case* $\check{x}$ *must be the smallest strictly positive steady state. Then again* $G_1(0) = 0$ *and* $G_1'(0) < 0$ *imply that* $x_{i+1} < x_i \quad \forall x_i < x_{conv}$*. Thus, the instability of* x_{conv} *implies that* $x_{ss}(|\bar{s}|) = 0$.

1) In the proof of Proposition 2 we showed that $G > 0$ *for some* x_i *iff* $K > K_{\min}$*. This implies* $\hat{x}$ *or* $\check{x}$ *or both exist. Since* $G' > 0$ *at both, this implies* $x_{i+1} > x_i$ *in a neighborhood of* $x_i > x_{conv}$*, which implies convergence to a stable steady state.*

3) When $\exists \dot{x}$*, we know by convexity of* G_1 *(and since the definition of* $\hat{x}$ *requires that* $G' > 0$ *at* $\hat{x}$*) that* $\hat{x}$ *does not exist. Hence, the only possible fix points are* $\dot{x}, \check{x}$ *and* x_{end}*. Note that by the definition of* $\dot{x}$ *it must be stable in a neighborhood above* $\dot{x}$*. By convexity of* G_1*,* $\dot{x}$ *must be unstable from below. Since there are no other fix points below* $\dot{x}$*,* $x_i < \dot{x}$ *implies convergence to* $x_{ss} = 0$*. This concludes the first subsentence. Furthermore, by instability of* $\check{x}$ *and stability of* $\dot{x}$ *from above we know that if* $x_i \in]\dot{x}, x_{conv}[$*, then there will be convergence to* $\dot{x}$ *which implies the second subsentence.*

4) $x_{conv} \equiv \hat{x}$ *whenever* $\exists \hat{x}$*. From Lemma 11 we know that* $\hat{x}$ *is independent of* $|\bar{s}|$ *and from Lemma 12 we know it exists iff* $|\bar{s}|$ *is sufficiently small. Hence, as* $|\bar{s}|$ *is increased,* x_{conv} *is either constant, or it makes a discrete jump to equal* $\check{x}$ *(which we know exists since* $K > K_{\min}$ *while in this scenario* $\hat{x}$ *seizes to exist). Furthermore, by Lemma 11 we know* $\check{x}$ *is increasing in* $|\bar{s}|$*. Put together, this implies that* x_{conv} *is either constant or increasing in* $|\bar{s}|$.

5) By definition of $\hat{x}$ *we get* $\hat{x} = K^{1/(\alpha-1)}$*, which decreases in* K*. By definition of* $\check{x}$ *and using equation (24) we get an implicit expression* $H = (\check{x}K)^{1/\alpha} + 1 - |\bar{s}| - 2\check{x} = 0$ *defining* $\check{x}$*. Using the implicit function theorem we get* $d\check{x}/dK = -(\check{x})^{1/\alpha} K^{1/\alpha - 1}/\alpha / \left(K^{1/\alpha}(\check{x})^{1/\alpha - 1}/\alpha - 2\right) < 0 \Leftrightarrow K^{1/\alpha}(\check{x})^{1/\alpha - 1} > 2\alpha$*. From equation (25) this condition corresponds to the condition for* $G_2' > 0$,

[32] To see this note that $\tilde{x}$ must be stable by $G'(\tilde{x}) < 0$. Furthermore, recall that $\nexists \dot{x}$. Hence, the only way for $\ddot{x}$ to be a steady state is if $G(\ddot{x}) = 0$ and $G'(\ddot{x}) > 0$, which implies $\ddot{x} = \hat{x}$ (see above). Finally, if x_{end} exists in region 3 it must be stable since $G_3' < 0$ and if x_{end} exists in region 1 or 2 then it must be that either $x_{end} = \check{x}$ or $x_{end} = \hat{x}$.

which holds by the definition of $\check{x}$. Hence x_{conv} is locally decreasing in K. Note now that, by equation (24), G_1 and G_2 are increasing in K. Hence, as K increases, we cannot switch from $x_{conv} = \hat{x}$ to $x_{conv} = \check{x}$. This implies that x_{conv} is decreasing in K also globally. ■

Lemma 14 *Suppose $K > K_{\min}$. Then there exists a stable steady state with a single norm at $x_{ss} = \tilde{x}$ or at $x_{ss} = 1$ or at both. No other stable steady state with a single norm exists.*

Proof. *When $K > K_{\min}$, a stable steady state must exist (see the proof of Proposition 2). All the steady states except for $\tilde{x}$ and 1 must be unstable since they all imply $G' > 0$ on at least one side of the steady state. Hence, when $K > K_{\min}$ there exists a stable steady state at $x_{ss} = \tilde{x}$ or at $x_{ss} = 1$ or at both, and since $x_{ss} \neq 0$, the steady state contains a single norm.* ■

Lemma 15 *Let p be a step function as in (6) and D as given in (4) with $\alpha > 0$. Consider two norms $\bar{s}$ and $\bar{s}'$ where $|\bar{s}| \leq |\bar{s}'|$. If there exists a degenerate stable steady state for $\bar{s}'$ then there exists a degenerate stable steady state for $\bar{s}$ too.*

Proof. *The condition for existence of a degenerate stable steady state for $\bar{s}'$ is $K^{1/\alpha} \geq 1 + |\bar{s}'|$ (see the definition of region 3 in equation 23). Since $|\bar{s}| \leq |\bar{s}'|$ we immediately get that $K^{1/\alpha} \geq 1 + |\bar{s}|$, hence there exists a degenerate stable steady state for $\bar{s}$ too.* ■

D.4.3 Welfare results

Statement (5) of Proposition 3 is proved by Lemma 20 below. But first we present some auxiliary results.

Lemma 16 *Let p be a step function as in (6) and D as given in (4) with $\alpha > 0$. Then, in any single norm stable steady state, welfare decreases in $|t - \bar{s}|$.*

Proof. *Consider first a degenerate stable steady state. In a degenerate stable steady state the welfare of each individual is solely determined by her dissonance from going to the norm, hence it is immediate that welfare decreases in the distance from the norm. Now consider a non-degenerate stable steady state. There are two groups of types to consider: the types close to the norm, who fully conform, and the types far from the norm, who follow their hearts. Within the group who fully conforms, welfare differs only due to the differences in D. As D increases in the distance to the norm, welfare within this group is strictly decreasing in $|t - \bar{s}|$. Furthermore, the type furthest away from the norm among them is indifferent between conforming and following her heart.*

The types far from the norm follow their hearts and suffer only the loss from social pressure, which, for p being a step function, is fixed at K. Among them we have the indifferent type, implying that these types are ranked at the bottom of the welfare distribution in society, and so welfare (weakly) decreases in the distance from the norm within this group, hence decreases in $|t-\bar{s}|$ globally. ■

Definition 3 *We call $r(t,\bar{s}) \in [0,1]$ the welfare ranking of an individual of type t in a given equilibrium with a norm at $\bar{s}$, if the fraction of people in society whose welfare is higher than that of t equals $r(t,\bar{s})$.*

Lemma 17 *Let p be a step function as in (6) and D as given in (4) with $\alpha > 0$. Consider two norms $\bar{s}$ and $\bar{s}'$ where $|\bar{s}| \leq |\bar{s}'|$, such that for each norm there exists a degenerate stable steady state ($x_{ss}(|\bar{s}|, K) = x_{ss}(|\bar{s}'|, K) = 1$). Then the welfare distribution in the steady state corresponding to $\bar{s}$ first-order stochastically dominates the welfare distribution in the steady state corresponding to $\bar{s}'$.*

Proof. *In a degenerate stable steady state the welfare of all individuals is solely determined by their dissonance from going to the norm, and equals $-D(|t-\bar{s}|)$. Given Lemma 16, a type with ranking $r \leq 1-|\bar{s}'|$ is at distance $d = r$ from the norm in both steady states and has the same ranking under $\bar{s}$ and under $\bar{s}'$, where in both cases her welfare equals $-D(d)$. A type with ranking $1-|\bar{s}'| < r < 1-|\bar{s}|$ has welfare of $-D(r)$ under $\bar{s}$ and $-D(2r-1+|\bar{s}'|) < -D(r)$ under $\bar{s}'$, because, under $\bar{s}'$ this type is at distance $2r-1+|\bar{s}'| > r$ from the norm (as the types at one side of the norm were already exhausted). Finally, a type with ranking $r \geq 1-|\bar{s}|$ has welfare of $-D(2r-1+|\bar{s}|)$ under $\bar{s}$ and $-D(2r-1+|\bar{s}'|) \leq -D(2r-1+|\bar{s}|)$ under $\bar{s}'$.* ■

Lemma 18 *Let p be a step function as in (6) and D as given in (4) with $\alpha > 0$. Consider two norms $\bar{s}$ and $\bar{s}'$ where $|\bar{s}| \leq |\bar{s}'|$, such that for each norm there exists a non-degenerate stable steady state ($x_{ss}(|\bar{s}|, K) \neq 1$ and $x_{ss}(|\bar{s}'|, K) \neq 1$). Then the welfare distribution in the steady state corresponding to $\bar{s}$ first-order stochastically dominates the welfare distribution in the steady state corresponding to $\bar{s}'$.*

Proof. *From Lemma 16 we know that in both steady states welfare decreases in the distance from the respective norm. Hence the welfare of the conformers, who are types close to the norm, is higher than that of the non conformers, who are types far from the norm. From Lemma 11 part (3) and the proof of Lemma 14, we know that $x_{ss}(|\bar{s}|, K) \geq x_{ss}(|\bar{s}'|, K)$. Thus, the bottom $1-x_{ss}(|\bar{s}|, K)$*

in both welfare distributions are people who follow their heart and their welfare, which is solely determined by the social pressure imposed on them, is $-K$. Right above them, in terms of welfare, there are the $x_{ss}(|\bar{s}|, K) - x_{ss}(|\bar{s}'|, K)$ types who follow their heart under $\bar{s}'$ but conform under $\bar{s}$. Their welfare equals $-K$ under $\bar{s}'$ and is higher under $\bar{s}$ (otherwise they could follow their heart under $\bar{s}$ too and get $-K$). Finally, the top $1 - x_{ss}(|\bar{s}'|, K)$ in both welfare distributions are people who conform in both cases. If, in the steady state under $\bar{s}'$, the conformers are on both sides of the norm (as can happen for $\alpha > 1$), then they are on both sides of the norm also under $\bar{s}$, which implies the same number of conformers under both norms. Hence, in both welfare distributions a ranking $r \leq 1 - x_{ss}(|\bar{s}'|, K)$ corresponds to distance $d = r$ from the norm, implying the same welfare under both distributions for the type ranked r. Otherwise, the conformers under $\bar{s}'$ are only on one side of the norm, and a similar reasoning to that of the proof to Lemma 17 applies: types with ranking $r \leq 1 - |\bar{s}'|$ have the same dissonance $D(r)$ in both steady states but are better off under $\bar{s}$ because there are less non-conformers who pressure them, and types with ranking $1 - |\bar{s}'| < r < 1 - |\bar{s}|$ are strictly better off under $\bar{s}$ both in terms of dissonance and social pressure. ■

Lemma 19 *Let p be a step function as in (6) and D as given in (4) with $\alpha > 0$. Consider a norm $\bar{s}$ such that there exist two stable steady states for this norm, one degenerate ($x_{ss}(|\bar{s}|, K) = 1$) and one non-degenerate ($x_{ss}(|\bar{s}|, K) \neq 1$). Then the welfare distribution in the degenerate stable steady state first-order stochastically dominates the welfare distribution in the non-degenerate stable steady state.*

Proof. *From Lemma 16 we get that the same welfare ranking applies to the two welfare distributions, where at the top of the ranking are those conforming in the two steady states, ordered by their distance from the norm, and after them the types who do not conform in the non-degenerate stable steady state, again ordered by their distance to the norm. The types at the top of the ranking, who conform in the two steady states, are all better-off in the degenerate stable steady state, because, while their dissonance is the same in both cases, they suffer from an additional social pressure in the non-degenerate stable steady state, from the actions chosen by the non conformers. The types at the bottom of the ranking, who conform only in the degenerate stable steady state, are also all better-off in that case, because their choice to conform implies that they are better-off by conforming, so for them $L = D(|t - \bar{s}|) < P(s; s') = K$, while their loss in the non-degenerate stable steady state equals K (which is the social pressure on the non conformers).* ■

Lemma 20 *Consider the dynamic model in (11) with p being a step function as in (6) and D as given in (4) with $\alpha > 0$. Then the welfare distribution under $\bar{s}$ first-order stochastically dominates the welfare distribution under $\bar{s}'$ if and only if $|\bar{s}| \leq |\bar{s}'|$.*

Proof. *Let $|\bar{s}| \leq |\bar{s}'|$. If $|\bar{s}| = |\bar{s}'|$ then symmetry implies that the welfare distributions are the same under both norms and the proposition holds in the weak sense. Otherwise there are two separate cases to consider. Case (i): there exists a degenerate stable steady state for $\bar{s}'$. By Lemma 15 this implies that there exists a degenerate stable steady state for $\bar{s}$ too, and by Lemma 17 we get that the welfare distribution in the degenerate stable steady state corresponding to $\bar{s}$ first-order stochastically dominates the welfare distribution in the degenerate stable steady state corresponding to $\bar{s}'$. If in addition there exists also a non-degenerate stable steady state for one (and only one) of the norms, then 19 implies that no further comparisons are needed (recall that in this case we compare $\max\{x_{ss}(|\bar{s}|, K)\}$ with $\max\{x_{ss}(|\bar{s}'|, K)\}$). Finally, if there exist a non-degenerate stable steady state for each of the two norms, then Lemma 18 implies that also the welfare distribution in the non-degenerate stable steady state corresponding to $\bar{s}$ first-order stochastically dominates the welfare distribution in the non-degenerate stable steady state corresponding to $\bar{s}'$. Case (ii): there does not exist a degenerate stable steady state for $\bar{s}'$, i.e., there exists only a non-degenerate one, in which case there are three sub-cases: (a) For $\bar{s}$ there exist two stable steady states, one degenerate and one non-degenerate, and then by Lemmas 18 and 19 we get that both welfare distributions under $\bar{s}$ first-order stochastically dominate the one welfare distribution under $\bar{s}'$. (b) For $\bar{s}$ there exists only a non-degenerate stable steady state, and then by Lemma 18 we get that the welfare distribution under $\bar{s}$ first-order stochastically dominates the welfare distribution under $\bar{s}'$. (c) For $\bar{s}$ there exists only a degenerate stable steady state. Then the ranking r that corresponds to distance d from the norm is at least as low under $\bar{s}'$ as it is under $\bar{s}$ (where, recall, the lower is r the better is the ranking), yet the welfare at this distance is lower under $\bar{s}'$. This is so because either the type at distance d conforms under $\bar{s}'$, in which case her dissonance is the same in both steady states, but she suffers from an additional social pressure in the non-degenerate stable steady state of $\bar{s}'$, from the actions chosen by the non conformers; or the type at distance d follows her heart under $\bar{s}'$, in which case her welfare is $-K$, whereas in the degenerate stable steady state of $\bar{s}$ all types have welfare higher than $-K$ (otherwise they would follow their hearts and get $-K$).* ■

E Appendix: Inverting societies

Let

$$s_l \equiv \bar{s} + 1 \text{ and}$$

$$\sigma \equiv s - \bar{s}.$$

These notations will be useful for proofs that deal with the case in which $\bar{s} < 0$ and $y > \bar{s} + 1$, where the distribution of actions is asymmetric around $\bar{s}$, and s_l then denotes the size of the uniform part to the left of $\bar{s}$, which equals the distance of $\bar{s}$ from the left corner of the types distribution, -1.

E.1 Proof of Lemma 3

When D is a step function taking the value of 0 or 1 and $P(s)$ has a unique min point at $\bar{s}$, we immediately have

$$s^*(t) = \begin{cases} \bar{s} \text{ if } 1 + P(\bar{s}) \leq P(t) \\ t \text{ if } 1 + P(\bar{s}) > P(t) \end{cases}. \tag{29}$$

Since, by assumption, P is increasing on each side of the norm, we get that types sufficiently far from the norm will state the norm and types sufficiently close to the norm will state their type.■

E.2 Proof of Lemma 4

If $[\bar{s} - y, \bar{s} + y] \cap [-1, 1] = [\bar{s} - y, \bar{s} + y]$, the distribution of actions is composed of a mass of individuals at $\bar{s}$ and a uniform part that is symmetric around $\bar{s}$. The pressure that results from each of the two parts of this distribution of actions increases in the distance from $\bar{s}$ (see Lemma 5 regarding the contribution from the uniform part), and so the lemma holds. Otherwise, assume without loss of generality that $\bar{s} < 0$ and that all types at $[-1, \bar{s} + y]$ follow their hearts, with $y > \bar{s} + 1$. The aggregate $P(s)$ that results from this distribution of actions can be written as

$$P(s) = \begin{cases} Kx\left|s - \bar{s}\right|^{\beta} + K\frac{1}{2}\frac{(s+1)^{\beta+1} + (\bar{s}+y-s)^{\beta+1}}{\beta+1} \text{ if } s \leq \bar{s} + y \\ Kx\left|s - \bar{s}\right|^{\beta} + K\frac{1}{2}\frac{(s+1)^{\beta+1} - (s-\bar{s}-y)^{\beta+1}}{\beta+1} \text{ if } s > \bar{s} + y \end{cases} \tag{30}$$

with

$$x = \left(1 - \frac{y}{2} - \frac{\bar{s}+1}{2}\right).$$

From the following expression of $P'(s)$

$$P'(s) = \begin{cases} -K\left(1-\frac{y}{2}-\frac{\bar{s}+1}{2}\right)\beta\left(\bar{s}-s\right)^{\beta-1}+K\frac{1}{2}\left(s+1\right)^{\beta}-K\frac{1}{2}\left(\bar{s}+y-s\right)^{\beta} & \text{if } s<\bar{s} \\ K\left(1-\frac{y}{2}-\frac{\bar{s}+1}{2}\right)\beta\left(s-\bar{s}\right)^{\beta-1}+K\frac{1}{2}\left(s+1\right)^{\beta}-K\frac{1}{2}\left(\bar{s}+y-s\right)^{\beta} & \text{if } \bar{s}<s\leq\bar{s}+y \\ K\left(1-\frac{y}{2}-\frac{\bar{s}+1}{2}\right)\beta\left(s-\bar{s}\right)^{\beta-1}+K\frac{1}{2}\left(s+1\right)^{\beta}-K\frac{1}{2}\left(s-\bar{s}-y\right)^{\beta} & \text{if } s>\bar{s}+y \end{cases} \tag{31}$$

it is clear that (a) $P'(s)\to-\infty$ as $s\to^{-}\bar{s}$ and $P'(s)\to\infty$ as $s\to^{+}\bar{s}$; and (b) $P(s)$ is decreasing in s for $s<\bar{s}$ (recall that $y>\bar{s}+1$) and is increasing in s for $s>\bar{s}+y$. Moreover, when $\frac{-1+\bar{s}+y}{2}<s\leq\bar{s}+y$ (i.e., s in the right half of the uniform part), we get that $(s+1)>(\bar{s}+y-s)$, hence $P'(s)$ is positive too (this comes from the fact that the part of $P(s)$ that originates in the uniform part is increasing in the distance from $\frac{-1+\bar{s}+y}{2}$, the center of this part). Therefore, the global min can only be found at $s\in\left[\bar{s},\frac{-1+\bar{s}+y}{2}\right]$. In this range we have

$$P'(s)=K\left(1-\frac{y}{2}-\frac{\bar{s}+1}{2}\right)\beta\left(s-\bar{s}\right)^{\beta-1}+K\frac{1}{2}\left(s+1\right)^{\beta}-K\frac{1}{2}\left(\bar{s}+y-s\right)^{\beta}.$$

Note first that (i) if $y=\bar{s}+1$, the distribution of actions is symmetric around $\bar{s}$, and so $P'(s)\geq0$ at the range $s\in\left[\bar{s},\frac{-1+\bar{s}+y}{2}\right]$; and (ii) if $y=1-\bar{s}$ (this is the distance from $\bar{s}$ to the furthest edge), then $P'(s)<0$ at the range $s\in\left[\bar{s},\frac{-1+\bar{s}+y}{2}\right]$, since Lemma 5 implies that $P(s)$ is increasing in the distance from $0>\frac{-1+\bar{s}+y}{2}$. Differentiating with respect to y we get

$$\frac{dP'(s)}{dy}=\frac{1}{2}K\left[-\beta\left(s-\bar{s}\right)^{\beta-1}-\beta\left(\bar{s}+y-s\right)^{\beta-1}\right]<0 \tag{32}$$

This inequality, together with i) and ii), then implies that $\exists y\in\left]\bar{s}+1,1-\bar{s}\right[$, denoted by $y_{\max}(\bar{s})$, such that $P'(s)\geq0$ at the whole range $s\in\left[\bar{s},\frac{-1+\bar{s}+y}{2}\right]$ if and only if $y\leq y_{\max}(\bar{s})$.[33] We will now show that $y_{\max}(\bar{s})\geq1$, by showing that for $y=1$ and every given $\bar{s}$, $P'(s)\geq0$ at the whole range $s\in\left[\bar{s},\frac{-1+\bar{s}+y}{2}\right]$.

Rewriting the expression for $P'(s)$ we get

$$P'(s)=\frac{1}{2}K\left[\left(2-y-s_l\right)\beta\sigma^{\beta-1}+\left(s_l+\sigma\right)^{\beta}-\left(y-\sigma\right)^{\beta}\right]. \tag{33}$$

[33] This already takes into account the fact that the range $\left[\bar{s},\frac{-1+\bar{s}+y}{2}\right]$ is itself increasing in y.

Differentiating with respect to s_l we get

$$\frac{dP'(s)}{ds_l} = \frac{1}{2}K\left[-\beta\sigma^{\beta-1} + \beta\left(s_l+\sigma\right)^{\beta-1}\right] \leq 0 \tag{34}$$

This inequality suggests that $P'(s)$ is minimal when s_l is maximal (i.e., equals $1-\varepsilon$, where $\bar{s} = -\varepsilon \to 0$). Note that in this case $\sigma \to 0$, as the range of s shrinks to be $s \in \left[-\varepsilon, \frac{-\varepsilon}{2}\right]$. Plugging $s = -\lambda\varepsilon$ into (33), and letting $\lambda \in [0.5, 1]$, we then have

$$\begin{aligned} P'(s) &= \frac{\varepsilon}{2}\beta\left(-\lambda\varepsilon+\varepsilon\right)^{\beta-1} + \frac{1}{2}\left(-\lambda\varepsilon+1\right)^{\beta} - \frac{1}{2}\left(-\varepsilon+1+\lambda\varepsilon\right)^{\beta} \\ &= \frac{\varepsilon^{\beta}}{2}\beta\left[\left(1-\lambda\right)\right]^{\beta-1} + \frac{1}{2}\left(1-\lambda\varepsilon\right)^{\beta} - \frac{1}{2}\left[1-\left(1-\lambda\right)\varepsilon\right]^{\beta}, \end{aligned}$$

we get[34]

$$P'(s) = \frac{\varepsilon^{\beta}}{2}\beta\left[\left(1-\lambda\right)\right]^{\beta-1} + \frac{1}{2}\left[\beta\left(1-2\lambda\right)\varepsilon + O(\varepsilon^2)\right]$$

and so, if $\beta < 1$

$$\lim_{\varepsilon\to 0} P'(s) = \lim_{\varepsilon\to 0}\frac{\varepsilon^{\beta}}{2}\beta\left[\left(1-\lambda\right)\right]^{\beta-1} = 0^{+}$$

and if $\beta = 1$

$$\lim_{\varepsilon\to 0} P'(s) = \frac{\varepsilon}{2}\left[1+1-2\lambda\right] = 0^{+}.$$

This means that even for the maximal s_l, $P'(s)$ is positive everywhere when $y = 1$, implying that $y_{\max}(\bar{s}) \geq 1$.■

E.3 Proof of Proposition 4

The proof of the proposition builds on a few auxiliary lemmas that are outlined first. The actual proof of the proposition follows after the lemmas.

Lemma 21 *If $\beta = 1$ then $y_{\max}(\bar{s}) = 1 \quad \forall\bar{s}$.*

Proof. *Lemma 4 implies that $y_{\max}(\bar{s}) \geq 1$. Plugging in $\beta = 1$ and letting $s \to^{+} \bar{s}$ in equation 31 yields $P'(s) = K(1-y)$. This expression is negative for $y > 1$, which, by the definition of $y_{\max}(\bar{s})$ implies that $y_{\max}(\bar{s}) \leq 1$. Thus $y_{\max}(\bar{s}) = 1 \quad \forall\bar{s}$.* ■

[34] In the following expression, $O(\varepsilon^2)$ is the standard mathematical notation for an element in the order of ε^2.

Lemma 22 *Suppose that* $\beta < 1$ *and* $s_l \in [0,1]$. *Then* $(1-s_l)\beta - 2 + (s_l+1)^\beta < 0$.

Proof. $(1-s_l)\beta \geq 0$ *and* $2-(s_l+1)^\beta \geq 0$. *However, we have* $(s_l+1)^\beta < s_l + 1 = 2-(1-s_l) < 2-(1-s_l)\beta$, *and so* $(1-s_l)\beta - \left[2-(s_l+1)^\beta\right] < 0$. ■

Lemma 23 *Suppose* $\beta \leq 1$. *Let* $\bar{s} \leq 0$ *and* $y \leq y_{\max}(\bar{s})$, *and suppose that all types* $t \in [\bar{s}-y, \bar{s}+y] \cap [-1,1]$ *follow their hearts and the rest state* $\bar{s}$. *If type* $t = \bar{s}+y$ *is indifferent between the two corner solutions* $s^*(t) = \bar{s}$ *and* $s^*(t) = t$, *then for any type the best response is*

$$s^*(t) = \begin{cases} \bar{s} \text{ if } |t-\bar{s}| > y \\ t \text{ otherwise} \end{cases}.$$

Proof. *For types with* $t > \bar{s}$ *the result follows from Lemmas 3 and 4. As for types* $t < \bar{s}$, *if* $[\bar{s}-y, \bar{s}+y] \cap [-1,1] = [\bar{s}-y, \bar{s}+y]$ *then the distribution of actions is symmetric around* $\bar{s}$ *and the result follows from* P *then being symmetric and monotonically increasing in* $|s-\bar{s}|$. *Otherwise, by construction all types at* $[-1, \bar{s}+y]$ *follow their hearts, where* $y > \bar{s}+1$. *We need to show that indeed all types with* $t < \bar{s}$ *have strict preference for the solution* $s^*(t) = t$. *Since we know from Lemma 4 that* P *is strictly increasing in the distance from* $\bar{s}$ *while* D *is fixed, it is sufficient to show that* $s^*(t) = t$ *for the type* $t = -1$. *Looking at* $t = -1$, *the fact that* P *gets its global min point at* $\bar{s}$ *and equation (29) imply that it is sufficient to show that* $1 + P(\bar{s}) - P(-1) \geq 0$. *Furthermore, note that the indifference of type* $t = \bar{s}+y$ *implies that* $1 + P(\bar{s}) - P(\bar{s}+y) = 0$. *Therefore, it is sufficient to show that* $P(\bar{s}+y) \geq P(-1)$:

$$P(\bar{s}+y) = Kxy^\beta + K\frac{1}{2}\frac{(\bar{s}+y+1)^{\beta+1}}{\beta+1},$$
$$P(-1) = Kx|-1-\bar{s}|^\beta + K\frac{1}{2}\frac{(\bar{s}+y+1)^{\beta+1}}{\beta+1},$$

and so $P(\bar{s}+y) \geq P(-1)$ *if and only if* $y \geq \bar{s}+1$, *which holds by assumption.* ■

Lemma 24 *Let* $\bar{s} \in [-1,1]$ *and let* D *be given by (12), and suppose that* $\beta \leq 1$. *For every* $y \leq y_{\max}(\bar{s})$, *let* $S(y)$ *denote a distribution of actions in society such that all types* $t \in [\bar{s}-y, \bar{s}+y] \cap [-1,1]$ *follow their hearts while the rest choose* $\bar{s}$. *Denote by* $K(y)$ *the value of* K *that, given the pressure function* $P(s)$ *that results from* $S(y)$, *implies indeed* $s^*(t) = \bar{s}$ *for all types with* $|t-\bar{s}| > y$ *and*

$s^*(t) = t$ for all types with $|t - \bar{s}| \leq y$. Then, when $\beta < 1$, $K(y)$ has either a U-shape or a W-shape, and when $\beta = 1$, $K(y)$ is monotonically decreasing.

Proof. *Without loss of generality, let $\bar{s} \leq 0$. The given distribution of actions and the fact that $y \leq y_{\max}(\bar{s})$ imply by Lemma 4 that P is increasing in $|\sigma|$ (recall $\sigma \equiv s - \bar{s}$). Moreover, from Lemma 3 we know that*

$$s^*(t) = \begin{cases} \bar{s} \text{ if } 1 + P(\bar{s}) \leq P(t) \\ t \text{ if } 1 + P(\bar{s}) > P(t) \end{cases}$$

which implies types sufficiently far from the norm will state the norm and types sufficiently close to the norm will state their type. We are looking for the value of K for which the type who is indifferent between the two options is at distance y from $\bar{s}$. I.e., $1 + P(\bar{s}) = P(\bar{s} + y)$. Lemma 23 implies that this distance y applies to both sides. However, as y grows from 0, we move from a region where the uniform part is symmetric around $\bar{s}$ (when $y \leq s_l$) to a region where it is asymmetric (when $y \in [s_l, 2 - s_l]$). Therefore the analysis will be first performed separately for each region, and then the two analyses will be combined.

Region (1): $y \leq s_l$

In this region the uniform part of S is symmetric around the norm and so the share of individuals following the norm is $x = 1 - y$ and $P(\sigma)$ is given by:

$$P(\sigma) = \begin{cases} Kx|\sigma|^\beta + K\frac{1}{2}\frac{(|\sigma|+y)^{\beta+1}+(y-|\sigma|)^{\beta+1}}{\beta+1} \text{ if } |\sigma| \leq y \\ Kx|\sigma|^\beta + K\frac{1}{2}\frac{(|\sigma|+y)^{\beta+1}-(|\sigma|-y)^{\beta+1}}{\beta+1} \text{ if } |\sigma| > y \end{cases} \tag{35}$$

The type who is indifferent between the two options is at distance y from $\bar{s}$, i.e., $1 + P(0) = P(y)$, if

$$\begin{aligned} 1/K + \frac{1}{2}\frac{2y^{\beta+1}}{\beta+1} &= (1-y)y^\beta + \frac{1}{2}\frac{(2y)^{\beta+1}}{\beta+1} \\ \Rightarrow 1/K &= (1-y)y^\beta + \left(2^\beta - 1\right)\frac{y^{\beta+1}}{\beta+1}. \end{aligned} \tag{36}$$

Region (2): $y \in [s_l, 2 - s_l]$

In this region the uniform part of S is asymmetric around the norm, and the share of individuals following the norm is $x = \left(1 - \frac{y}{2} - \frac{s_l}{2}\right)$. Rewriting (30)

we get that $P(\sigma)$ *is given by:*

$$P(\sigma) = \begin{cases} Kx|\sigma|^{\beta} + K\frac{1}{2}\frac{(s_l+\sigma)^{\beta+1}+(y-\sigma)^{\beta+1}}{\beta+1} & \text{if } \sigma \leq y \\ Kx|\sigma|^{\beta} + K\frac{1}{2}\frac{(s_l+\sigma)^{\beta+1}-(\sigma-y)^{\beta+1}}{\beta+1} & \text{if } \sigma \geq y \end{cases}.$$

The type who is indifferent between the two options is at distance y *from* $\bar{s}$*, i.e.,* $1 + P(0) = P(y)$*, if*

$$1/K + \frac{1}{2}\frac{(s_l)^{\beta+1} + (y)^{\beta+1}}{\beta+1} = \left(1 - \frac{y}{2} - \frac{s_l}{2}\right) y^{\beta} + \frac{1}{2}\frac{(s_l+y)^{\beta+1}}{\beta+1} \Rightarrow$$
$$1/K = \left(1 - \frac{y}{2} - \frac{s_l}{2}\right) y^{\beta} + \frac{1}{2}\frac{(s_l+y)^{\beta+1} - (s_l)^{\beta+1} - y^{\beta+1}}{\beta+1} \quad (37)$$

Joining the two regions:

Following equations 36 and 37, we can get the following expression for $\frac{1}{K}$ *as a function of* y*.*

$$\frac{1}{K}(y) = \begin{cases} (1-y)\, y^{\beta} + \left(2^{\beta} - 1\right)\frac{y^{\beta+1}}{\beta+1} & \text{if } y \leq s_l \\ \left(1 - \frac{y}{2} - \frac{s_l}{2}\right) y^{\beta} + \frac{1}{2}\frac{(s_l+y)^{\beta+1}-(s_l)^{\beta+1}-y^{\beta+1}}{\beta+1} & \text{if } y \in [s_l, 2-s_l] \end{cases} \quad (38)$$

Differentiating in both regions yields

$$\frac{d(1/K)}{dy} = \begin{cases} (1-y)\, y^{\beta-1}\beta - y^{\beta}\left(2 - 2^{\beta}\right) & \text{if } y \leq s_l \\ \left(1 - \frac{y}{2} - \frac{s_l}{2}\right)\beta y^{\beta-1} - y^{\beta} + \frac{1}{2}(s_l+y)^{\beta} & \text{if } y \in [s_l, 2-s_l] \end{cases}. \quad (39)$$

When $\beta = 1$ *we get that* $\frac{d(1/K)}{dy} = 1 - y$ *in both regions, hence* $1/K$ *is a strictly increasing function of* y *in the range* $[0,1]$ *(and* $K(y)$ *is strictly decreasing in* $y \in [0,1]$*). Since in this case* $y_{\max} = 1$ *(see Lemma 21), we get that the lemma holds for* $\beta = 1$*. We continue now with the case of* $\beta < 1$*. Differentiating once more*

$$\frac{d^2(1/K)}{dy^2}$$
$$= \begin{cases} -y^{\beta}\beta + (1-y)\, y^{\beta-2}(\beta-1)\beta - \beta y^{\beta-1}\left(2 - 2^{\beta}\right) < 0 & \text{if } y \leq s_l \\ \left(1 - \frac{y}{2} - \frac{s_l}{2}\right)\beta(\beta-1)\, y^{\beta-2} - \frac{3}{2}\beta y^{\beta-1} + \frac{1}{2}\beta(s_l+y)^{\beta-1} < 0 & \text{if } y \in [s_l, 2-s_l] \end{cases} \quad (40)$$

so that $1/K$ *is concave in* y *in both regions. Moreover, it is easy to verify that* $\frac{1}{K}(y)$ *is continuous at* $y = s_l$*, the border between the two regions. If* $\bar{s} = 0$ *(*$s_l = 1$*), then only the first region applies. It is easy to verify that in the first*

region we get the following

$$\begin{cases} \frac{d(1/K)}{dy} > 0 \text{ as } y \to 0 \\ \frac{d(1/K)}{dy} < 0 \text{ as } y \to 1 \end{cases},$$

and so in this case $\frac{1}{K}(y)$ *is hill-shaped. Otherwise* $\bar{s} < 0$ *(*$s_l < 1$*). For the applicability of* $\frac{1}{K}(y)$ *in this lemma we require that* $y \leq y_{\max}(\bar{s})$*. When* $\bar{s} < 0$ *we still have* $\frac{d(1/K)}{dy} > 0$ *as* $y \to 0$*, but* $s_l < 1 \leq \bar{y} \equiv \min\{y_{\max}(\bar{s}), 2 - s_l\}$ *(recall that from Lemma 4 we know that* $y_{\max}(\bar{s}) \geq 1$*), and so region (2) applies to large enough values of* y*. Moreover,* $\frac{d^2(1/K)}{dy^2} < 0$ *implies that* $\frac{d(1/K)}{dy}$ *is strictly decreasing in* y*. Hence,* $\bar{y} \geq 1$ *implies that* $\frac{d(1/K)}{dy}|_{y=\bar{y}} \leq \frac{d(1/K)}{dy}|_{y=1} = \frac{1}{2}\left[(1-s_l)\beta - 2 + (s_l+1)^{\beta}\right]$*, which by Lemma 22 is strictly negative. Hence we know that* $\frac{1}{K}(y)$ *has a positive slope at* $y \to 0$ *and a negative slope at* $y = \min\{y_{\max}(\bar{s}), 2 - s_l\}$*, and in between it is concave in each of the regions. It thus follows that* $\frac{1}{K}(y)$ *has at least one and at most two max points and that these max points are internal, i.e.* $\frac{1}{K}(y)$ *is either hill-shaped or* M*-shaped, and so* $K(y)$ *is either* U*-shaped or* W*-shaped.* ■

Lemma 25 *Let* D *be given by (12) and let* $\beta < 1$*. Suppose there exists a value of* K *such that a single-norm equilibrium at* $\bar{s} \in [-1, 1]$*, where all types* $t \in [\bar{s} - y, \bar{s} + y] \cap [-1, 1]$ *follow their hearts while the rest choose* $\bar{s}$*, exists for some* $y > y_{\max}(\bar{s})$*. Then* $K \geq K_{\min}(|\bar{s}|)$*.*

Proof. *Without loss of generality, let* $\bar{s} \leq 0$*. Since the existence of the equilibrium that is described in the lemma requires that* $t = \bar{s} + y$ *will be indifferent between following her heart and choosing* $\bar{s}$*, and since* $y > y_{\max}(\bar{s}) \geq 1 \geq s_l$*, the value of* K *that may allow such an equilibrium (if it indeed exists) is given by equation (37), with first and second derivatives as in the second lines of equations (39) and (40) respectively. Then, the fact that* $\frac{d^2(1/K)}{dy^2} < 0$ *implies that the value of* $\frac{d(1/K)}{dy}$ *at any* $y > y_{\max}(\bar{s})$ *is strictly smaller than* $\frac{d(1/K)}{dy}|_{y=1} = \frac{1}{2}\left[(1-s_l)\beta - 2 + (s_l+1)^{\beta}\right]$*, which by Lemma 22 is negative. Hence,* $\frac{1}{K}(y)$ *is decreasing when* $y > y_{\max}(\bar{s})$*, implying that for any* $y > y_{\max}(\bar{s})$*, an equilibrium as described in the lemma requires* $K(y) > K(y_{\max}(\bar{s})) \geq K_{\min}(|\bar{s}|)$*.* ■

Lemma 26 *Let* D *be given by (12) and suppose that* $\beta \leq 1$*. Then the only possible distribution of actions in a single-norm equilibrium at* $\bar{s} \in [-1, 1]$ *is one where all types* $t \in [\bar{s} - y, \bar{s} + y] \cap [-1, 1]$ *for some* $y > 0$ *follow their hearts while the rest choose* $\bar{s}$*.*

Proof. *First note that if D is a step function as in (12), then for any $t \in [-1,1]$, either $s^*(t) = t$ or $s^*(t) \in \arg\min(P)$. Then, the existence of a single-norm equilibrium at $\bar{s}$ implies that (i) $\bar{s} \in \arg\min(P)$ and (ii) $s^*(t) = t$ for every t for whom $s^*(t) \neq \bar{s}$. Together with the uniform distribution of types, this implies that the distribution of actions can contain only uniform parts apart from the peak at $\bar{s}$.*

Moreover, the continuity of $P(s)$ implies that for types sufficiently close to $\bar{s}$, $1 + P(\bar{s}) > P(t)$ (since then $P(t) \to P(\bar{s})$), and so the distribution of actions must necessarily contain a uniform part that is attached to $\bar{s}$. We will now show that there can be no other uniform parts in the distribution of actions. Without loss of generality, let $\bar{s} \leq 0$, and suppose that there exist (one or more) uniform parts that are detached from $\bar{s}$. Consider the rightmost uniform part. Since P is continuous, at the left edge of this specific uniform part there must be a type t who is indifferent between $s^*(t) = t$ and $s^*(t) = \bar{s}$, i.e., for whom $1 + P(\bar{s}) = P(t)$. Note also that the sources of the pressure $P(s)$ can be divided into two sections – those that compose the rightmost uniform part, and those that lie to the left of this uniform part. The sources of the first section impose the same pressure on the type at the left edge of the rightmost uniform part and on the type at the right edge of this uniform part (due to symmetry). The sources of the second section impose more pressure on the latter, because this type is farther away from the norm. Together with the fact that D is the same for both types, this implies that $1 + P(\bar{s}) < P(t)$ for this latter type, in contradiction to the assumption that this type chooses $s^*(t) = t$. Since a rightmost and detached uniform part cannot exist this implies that no detached uniform part can exist to the right of $\bar{s}$. A similar argument applies to the left of $\bar{s}$ and hence we have shown that the only uniform part that can exist is attached to $\bar{s}$.

Finally, we need to show that this uniform part can be written as $[\bar{s} - y, \bar{s} + y] \cap [-1,1]$ for some y, which boils down to showing that it cannot be asymmetric if it does not touch any of the edges of the type distribution. I.e., this part cannot be $[\bar{s} - y_1, \bar{s} + y_2] \subset [-1,1]$ where $y_1, y_2 > 0$ and $y_1 \neq y_2$. Suppose to the contrary that this case holds. Then the aggregate pressure $P(s)$ is given by:

$$P(\sigma) = \begin{cases} Kx|\sigma|^{\beta} + K\frac{1}{2}\frac{(\sigma+y_1)^{\beta+1}+(y_2-\sigma)^{\beta+1}}{\beta+1} & \text{if } -y_1 \leq \sigma \leq y_2 \\ Kx|\sigma|^{\beta} + K\frac{1}{2}\frac{(y_2-\sigma)^{\beta+1}-(-y_1-\sigma)^{\beta+1}}{\beta+1} & \text{if } \sigma < -y_1 \\ Kx|\sigma|^{\beta} + K\frac{1}{2}\frac{(\sigma+y_1)^{\beta+1}-(\sigma-y_2)^{\beta+1}}{\beta+1} & \text{if } \sigma > y_2 \end{cases} \tag{41}$$

where $x = \frac{y_1+y_2}{2}$. Moreover, both the type $t_1 = \bar{s} - y_1$ and the type $t_2 = \bar{s} - y_2$ are indifferent between $s^*(t) = t$ and $s^*(t) = \bar{s}$. Hence it must hold simultaneously

that $1+P(0)=P(-y_1)$ *and* $1+P(0)=P(y_2)$*, i.e.,* $P(-y_1)=P(y_2)$*. Substituting* $\sigma=-y_1$ *and* $\sigma=y_2$ *in equation (41) we get*

$$Kxy_1^{\beta}+K\frac{1}{2}\frac{(y_2+y_1)^{\beta+1}}{\beta+1}=Kxy_2^{\beta}+K\frac{1}{2}\frac{(y_2+y_1)^{\beta+1}}{\beta+1}$$
$$\Rightarrow y_1^{\beta}=y_2^{\beta}$$

which contradicts $y_1\neq y_2$*.* ■

Lemma 27 *Suppose* $\beta\leq 1$*.* $K_{\min}(|\bar{s}|)$ *is weakly decreasing in* $|\bar{s}|$*.*

Proof. *We start with the case* $\beta<1$*. First note that* $K_{\min}$ *is never found on the border between the regions (1) and (2),*[35] *since* $\frac{d(1/K)}{dy}|_{y\to^+ s_l}$ *is strictly greater (unless* $s_l=0$*) than* $\frac{d(1/K)}{dy}|_{y\to^- s_l}$*. We can therefore rewrite equation (38) as a function of* $\bar{s}$ *for the two regions and differentiate* $1/K$ *w.r.t.* $\bar{s}$*. This yields*

$$\frac{d(1/K)}{d\bar{s}}=\begin{cases}0 \text{ if } y\leq \bar{s}+1\\ -\frac{y^{\beta}}{2}+\frac{1}{2}(\bar{s}+y+1)^{\beta}-\frac{1}{2}(\bar{s}+1)^{\beta} \text{ if } y\in[\bar{s}+1,\min\{y_{\max}(\bar{s}),1-\bar{s}\}]\end{cases} \quad (42)$$

$$\frac{d^2(1/K)}{d\bar{s}^2}=\begin{cases}0 \text{ if } y\leq \bar{s}+1\\ \frac{1}{2}\beta(\bar{s}+y+1)^{\beta-1}-\frac{1}{2}\beta(\bar{s}+1)^{\beta-1} \text{ if } y\in[\bar{s}+1,\min\{y_{\max}(\bar{s}),1-\bar{s}\}]\end{cases} \quad (43)$$

Note that $\frac{d(1/K)}{d\bar{s}}|_{y\to^+\bar{s}+1}=\left(2^{\beta-1}-1\right)(\bar{s}+1)^{\beta}<0$ *and* $\frac{d^2(1/K)}{d\bar{s}^2}\leq 0$*. These results imply that* $\frac{1}{K}(y)$ *is constant in* $\bar{s}$ *in the first region and is strictly decreasing in* $\bar{s}$ *in region (2) (note that this does not violate the continuity of* $\frac{1}{K}(y)$ *as can be verified by plugging* $y=s_l$ *in equation (38)). Hence, since we have been analyzing the case of* $\bar{s}\leq 0$*, more generally* $K(y)$ *is weakly decreasing in* $|\bar{s}|$*. In particular* $K_{\min}$ *is weakly decreasing in* $|\bar{s}|$ *– it stays constant if* $K_{\min}$ *is achieved in region (1) both before and after the change in* $|\bar{s}|$*, and is strictly decreasing if* $K_{\min}$ *is achieved in region (2) after the change in* $|\bar{s}|$*.*

Now for the case $\beta=1$*. Plugging* $\beta=1$ *into equation (38) we get that both regions are independent of* $\bar{s}$*. Hence,* $K_{\min}$ *is independent of* $\bar{s}$*.* ■

Proof of Proposition 4

Lemma 24 implies that for any $\bar{s}\in[-1,1]$, one can construct a distribution of stances, denoted by $S(y)$, such that all types $t\in[\bar{s}-y,\bar{s}+y]\cap[-1,1]$ for some $y\leq y_{\max}(\bar{s})$ follow their hearts while the rest choose $\bar{s}$, if a suitable value of K is chosen. This means $S(y)$ forms a single-norm equilibrium. Moreover, this lemma says that $K(y)$, the value for which this single-norm equilibrium

[35] These regions are defined in the proof of Lemma 24.

exists for a given y, is either U-shaped or W-shaped as a function of y when $\beta < 1$; and $K(y)$ is strictly decreasing in y with a min point at $y = y_{\max}$ when $\beta = 1$. When $y \to 0$ we have

$$\lim_{y\to 0} 1/K = \lim_{y\to 0} \left\{ (1-y)\, y^{\beta} + \left(2^{\beta} - 1\right) \frac{y^{\beta+1}}{\beta+1} \right\} = 0,$$

so that $K(y) \to \infty$. Let $K_{\min}(|\bar{s}|)$ denote the minimal value of $K(y)$. It thus immediately follows that for $K \geq K_{\min}(|\bar{s}|)$ there exists a fix point y while for $K < K_{\min}(|\bar{s}|)$ there does not. This proves the if part of statement (1). As for the only if part of the statement, note that Lemma 26 implies that in any single-norm equilibrium, all types $t \in [\bar{s} - y, \bar{s} + y] \cap [-1, 1]$ for some $y < 1 + |\bar{s}|$ follow their hearts while the rest choose $\bar{s}$. It thus suffices to show that if such an equilibrium exists for some $y > y_{\max}(\bar{s})$, then still $K \geq K_{\min}(|\bar{s}|)$. For $\beta < 1$ this is proved in Lemma 25. For $\beta = 1$ we know from Lemma 21 that $y_{\max} = 1$. Then, when $y > y_{\max} = 1$, no K can sustain a single-norm equilibrium at $\bar{s}$. This can be seen by setting $\beta = 1$ and letting $s \to^{+} \bar{s}$ in equation 31, and noting that, for $y > 1$, $\bar{s}$ is not the global min point of P and so cannot be the norm given that D is a step function. As for statement (2) of the proposition, the fact that $K_{\min}$ is weakly decreasing in $|\bar{s}|$ follows directly from Lemma 27.■

E.4 Proof of Proposition 5

The proof of the proposition builds on a few auxiliary lemmas, and on expressions within these lemmas, that are outlined first. The actual proof of the proposition follows after the lemmas.

Lemma 28 *Suppose $\beta \leq 1$. Suppose in some generation i there exists a cutoff distance from the norm y_i, such that all types in that generation that fulfill $|t - \bar{s}| > y_i$ follow the norm and all types fulfilling $|t - \bar{s}| \leq y_i$ follow their hearts and that $y_i \leq y_{\max}(\bar{s})$. Then there exists a cutoff y_{i+1} in the next generation, such that all types that fulfill $|t - \bar{s}| > y_{i+1}$ follow the norm and all types that fulfill $|t - \bar{s}| \leq y_{i+1}$ follow their hearts. Furthermore y_{i+1} is an increasing function of y_i.*

Proof. *When $y_i \leq y_{\max}(\bar{s})$ then by Lemma 4 P is increasing with distance from $\bar{s}$. Since D is a fixed cost it implies that types sufficiently far from $\bar{s}$ follow $\bar{s}$ and types sufficiently close follow their t (note that this cutoff may be such that all types choose $s = t$). By Lemma 23 we know that if the cutoff type $t = \bar{s} + y_{i+1}$ is such that $\bar{s} - y_{i+1} < -1$ then type $t = -1$ strictly prefers stating her type. This implies that we only need to focus on the indifferent type $t > \bar{s}$.*

The indifferent type (which we define as $t_c \equiv \bar{s} + y_{i+1}$) is such that

$$L(t_c, t_c) = P_{i+1}(t_c) = P_{i+1}(\bar{s}) + D(t_c, \bar{s}) = L(\bar{s}, t_c).$$

Define

$$F \equiv D(\bar{s}, t_c)/K + P_{i+1}(\bar{s})/K - P_{i+1}(t_c)/K = 0.$$

Then $F = 0$ implicitly gives us y_{i+1} as a function of y_i). For a given y_i, F can take one of the following forms:

$$F = \tag{44}$$

$$\begin{cases} F_1 \equiv \frac{1}{K} + \frac{1}{2}\frac{(\bar{s}+1)^{\beta+1} + (y_i)^{\beta+1}}{\beta+1} - \frac{1}{2}\left[(1 - y_i - \bar{s})(y_{i+1})^{\beta} + \frac{(\bar{s}+y_{i+1}+1)^{\beta+1} + (y_i - y_{i+1})^{\beta+1}}{\beta+1}\right] \text{ if } y_i \geq y_{i+1}, \bar{s} - y_i < -1 \\ F_2 \equiv \frac{1}{K} + \frac{y_i^{\beta+1}}{\beta+1} - \left[(1 - y_i)(y_{i+1})^{\beta} + \frac{1}{2}\frac{(y_{i+1}+y_i)^{\beta+1} + (y_i - y_{i+1})^{\beta+1}}{\beta+1}\right] \text{ if } y_i \geq y_{i+1}, \bar{s} - y_i \geq -1 \\ F_3 \equiv \frac{1}{K} + \frac{1}{2}\frac{(\bar{s}+1)^{\beta+1} + (y_i)^{\beta+1}}{\beta+1} - \frac{1}{2}\left[(1 - y_i - \bar{s})(y_{i+1})^{\beta} + \frac{(\bar{s}+y_{i+1}+1)^{\beta+1} - (y_{i+1} - y_i)^{\beta+1}}{\beta+1}\right] \text{ if } y_i \leq y_{i+1}, \bar{s} - y_i < -1 \\ F_4 \equiv \frac{1}{K} + \frac{y_i^{\beta+1}}{\beta+1} - \left[(1 - y_i) y_{i+1}^{\beta} + \frac{1}{2}\frac{(y_{i+1}+y_i)^{\beta+1} - (y_{i+1} - y_i)^{\beta+1}}{\beta+1}\right] \text{ if } y_i \leq y_{i+1}, \bar{s} - y_i \geq -1 \end{cases}$$

Note that when $\bar{s} - y_t \to -1$ then $F_1 = F_2$ and $F_3 = F_4$; that when $y_{i+1} \to y_i$ then $F_1 = F_3$ and $F_2 = F_4$; and finally that when $\bar{s} - y_i \to -1$ and $y_{i+1} \to y_i$ then $F_1 = F_3 = F_2 = F_4$. Hence, since each of F_1, F_2, F_3 and F_4 is continuous then F is a continuous function and hence y_{i+1} is a continuous function of y_i. This implies that, if y_{i+1} is an increasing function y_i for each of F_1, F_2, F_3 and F_4, then y_{i+1} is an increasing function of y_i also globally. By the implicit function theorem we have

$$\frac{dy_{i+1}}{dy_i} = -\frac{F_{y_i}}{F_{y_{i+1}}}.$$

Note that the bracket in each F equals $P(s)|_{s=y_{i+1}}$, which implies that

$$F_{y_{i+1}} = -\frac{dP}{dy_{i+1}} = -\frac{dP}{ds}|_{s=y_{i+1}}, \tag{45}$$

which we know is negative by Lemma 4. Hence, if F_{y_i} is positive then $\frac{dy_{i+1}}{dy_i}$ is positive.

$$F_{y_i} = \begin{cases} \frac{1}{2}(y_i)^{\beta} + \frac{1}{2}(y_{i+1})^{\beta} - \frac{1}{2}(y_i - y_{i+1})^{\beta} & \text{if } y_i \geq y_{i+1}, \bar{s} - y_i < -1 \\ y_i^{\beta} + y_{i+1}^{\beta} - \frac{1}{2}(y_{i+1} + y_i)^{\beta} - \frac{1}{2}(y_i - y_{i+1})^{\beta} & \text{if } y_i \geq y_{i+1}, \bar{s} - y_i \geq -1 \\ \frac{1}{2}y_i^{\beta} + \frac{1}{2}y_{i+1}^{\beta} - \frac{1}{2}(y_{i+1} - y_i)^{\beta} & \text{if } y_i < y_{i+1}, \bar{s} - y_i < -1 \\ y_i^{\beta} + y_{i+1}^{\beta} - \frac{1}{2}(y_{i+1} + y_i)^{\beta} - \frac{1}{2}(y_{i+1} - y_i)^{\beta} & \text{if } y_i > y_{i+1}, \bar{s} - y_i > -1 \end{cases} \tag{46}$$

From this expression one can see that F_{y_i} is strictly positive on all rows: the first and third row trivially follow from $\frac{1}{2}(y_i)^\beta > \frac{1}{2}(y_i - y_{i+1})^\beta$ and the second and fourth row follow since $\frac{1}{2}y_i^\beta + \frac{1}{2}y_{i+1}^\beta \geq \frac{1}{2}(y_{i+1} + y_i)^\beta$ and $\frac{1}{2}y_i^\beta > \frac{1}{2}(y_i - y_{i+1})^\beta$. ■

Lemma 29 *Suppose $\beta \leq 1$. Then:*

1. *$y_{\max}(\bar{s})$ (from Lemma 4) is weakly increasing in $|\bar{s}|$.*
2. *Let $K(y)$ be implicitly given by equation (37) and let $\tilde{y}$ denote an implicit solution to this equation for a given value of K. Then if $K'(\tilde{y}) > 0$, $\tilde{y}$ is weakly increasing in $|\bar{s}|$, and if $K'(\tilde{y}) < 0$, $\tilde{y}$ is weakly decreasing in $|\bar{s}|$*

Proof. *$y_{\max}(\bar{s})$ is the maximum value of y such that $P(s)$ is monotonically increasing in $|s - \bar{s}|$. In Lemma 4 we show that it is unique for a given $\bar{s}$, such that $P(s)$ is monotonically increasing if and only if $y \leq y_{\max}(\bar{s})$. For $\beta = 1$ we know from Lemma 21 that $y_{\max} = 1 \quad \forall \bar{s}$. For $\beta < 1$ we will show that $y_{\max}(s_l)$ is decreasing in s_l (recall that $s_l \equiv \bar{s} + 1$), which is equivalent to the first statement in the lemma. Suppose that s_l is given, and that $y = y_{\max}(s_l)$. It follows then that $\exists s \in [-1, 1]$ such that $P'(s) = 0$. If we then increase s_l by some ϵ while keeping $y = y_{\max}(s_l)$, we get by equation (34) that $\exists s \in [-1, 1]$ such that $P'(s) < 0$, implying that $P(s)$ is not monotonically increasing in $|s - \bar{s}|$ for any $y \leq y_{\max}(s_l)$. This means that $y_{\max}(s_l + \epsilon) < y_{\max}(s_l)$, i.e., $y_{\max}(\bar{s})$ is increasing in $|\bar{s}|$ as in statement (1).*

2) Equation (37) depicts the function $K(y)$ in region (2) (as defined in Lemma 24). From the proof of Lemma 24 we know that if $\beta < 1$ then $K(y)$ is weakly decreasing in $|\bar{s}|$ and if $\beta = 1$ then $K(y)$ is constant in $|\bar{s}|$, and this holds in particular for region (2). It thus follows that, for a given value of K, any implicit solution $\tilde{y}$ for which $K'(\tilde{y}) > 0$ is weakly increasing in $|\bar{s}|$, and any implicit solution $\tilde{y}$ for which $K'(\tilde{y}) > 0$ is weakly decreasing in $|\bar{s}|$. ■

Proof of Proposition 5

Statement 1): Recalling that $F = 0$ in equation (44) implicitly gives us $y_{i+1}(y_i)$, we can see in that equation that when $y_i = 0$, the only way for F to equal zero is to have $F = F_4 = 1/K - y_{i+1}^\beta$, implying that $y_{i+1}(0) > 0$.[36] Lemma 28 further shows that y_{i+1} is an increasing function of y_i. If $K < K_{\min}(|\bar{s}|)$, we know from Lemma 24 that no steady state exists. Otherwise, if $K \geq K_{\min}(|\bar{s}|)$,

[36] To see this note that when $y_i = 0$, F_4 and F_2 are the only relevant cases and that if $F = F_2$ then by construction it must be that $y_{i+1} = 0$ implying $F = F_2 \equiv 1/K \neq 0$, which contradicts $F = 0$.

then by Lemma 24 we know that a steady state exists (at least one). Next, note that F in equation (44) is strictly decreasing in K (this applies to F_1, F_2, F_3 and F_4). This implies that $F_K < 0$, which together with $F_{y_{i+1}} < 0$ (see equation 45) implies that $\frac{dy_{i+1}}{dK} = -\frac{F_K}{F_{y_{i+1}}} < 0$, i.e., that the function $y_{i+1}(y_i)$ goes down when K increases. This means that when $K < K_{\min}(|\bar{s}|)$, the function $y_{i+1}(y_i)$ always stays above the 45 degree line (i.e. the line that implies $y_{i+1} = y_i$); when $K = K_{\min}(|\bar{s}|)$ the function $y_{i+1}(y_i)$ is tangent to the 45 degree line, and when $K > K_{\min}(|\bar{s}|)$ the function $y_{i+1}(y_i)$ crosses the 45 degree line at least once. It thus follows that when $K = K_{\min}(|\bar{s}|)$, any steady state would not be stable, as there can be no convergence to it from the right. Furthermore, if $K > K_{\min}(|\bar{s}|)$, it implies together with $y_{i+1}(0) > 0$ that there must be at least one stable steady state, as there is at least one point where the function $y_{i+1}(y_i)$ crosses the 45 degree line, starting above it and continuing below it. Denoting the leftmost stable steady state by y_{ss} and $\min\{y(K_{\min}(\bar{s}))\}$ by $y_{\min}(\bar{s})$ (note that $y(K_{\min}(\bar{s}))$ is unique if $K(y)$ is U-shaped and may have at most two solutions when it is W-shaped). Then we know that $y_{ss} \leq y_{\min}(\bar{s})$ because our analysis up till now implies that $y_{i+1}(y_{\min}(\bar{s})) < y_{\min}(\bar{s})$.[37] From $y_{i+1}(0) > 0$ we know that $y_{ss} \neq 0$, and since $y_{ss} \leq y_{\min}(\bar{s})$, it follows that $x_{ss} \in]0, 1[$.

Statement 2): Let $\bar{s}$ and $\bar{s}'$ be two norms such that $|\bar{s}| \leq |\bar{s}'|$. First note x is decreasing if y is increasing. Hence, to show that $x_{ss}(|\bar{s}|, K) \leq x_{ss}(|\bar{s}'|, K)$, it is sufficient to show that $y_{ss}(|\bar{s}|, K) \geq y_{ss}(|\bar{s}'|, K)$. In what follows, we will show that, for given $\bar{s} \leq 0$ and K, there exist at most two stable steady states with a norm and whenever two such steady states exist, one has a truncated left uniform part (when $\bar{s} - y_{ss} < -1$) and one has equally long uniform parts on each side of the norm ($\bar{s} - y_{ss} > -1$). Our comparison rule is: if each of $|\bar{s}|$ and $|\bar{s}'|$ has a unique stable steady state, then we compare these; if one of $|\bar{s}|$ and $|\bar{s}'|$ has a unique stable steady state and the other has two, then we compare $\min\{y_{ss}(|\bar{s}|)\}$ with $\min\{y_{ss}(|\bar{s}'|)\}$; and if $|\bar{s}|$ and $|\bar{s}'|$ have two stable steady states each, we compare the truncated steady states with each other and the non-truncated steady states with each other.

Let now $K > K_{\min}(|\bar{s}|)$ and take a steady state, be it stable or unstable. To verify stability we need to compute dy_{i+1}/dy_i at the steady state – it is stable from both sides if and only if the derivatives are smaller than 1. To simplify calculations, note first from (46) that in steady states, where $y_{i+1} = y_i$, $\frac{dF_1}{dy_i} = \frac{dF_3}{dy_i}$ and $\frac{dF_2}{dy_i} = \frac{dF_4}{dy_i}$. This means we can work solely with F_3 and F_4,

[37] Note that $y_{\min}(\bar{s})$ is a steady state when $K = K_{\min}(|\bar{s}|)$, in which case $y_{i+1}(y_{\min}(\bar{s})) = y_{\min}(\bar{s})$. As K is further increased, $y_{i+1}(y_{\min}(\bar{s}))$ goes down.

depending on the region of y, as defined in Lemma 24.[38] If the steady state falls in the first region, where $y < s_l$, then F_4 applies. There we have

$$\begin{aligned}\frac{dy_{i+1}}{dy_i} &= -\frac{F_{y_i}}{F_{y_{i+1}}} \\ &= -\frac{y_t^\beta + y_{t+1}^\beta - \frac{1}{2}\left(y_{t+1}+y_t\right)^\beta - \frac{1}{2}\left(y_{t+1}-y_t\right)^\beta}{-\left[\left(1-y_t\right)\beta y_{t+1}^{\beta-1} + \frac{1}{2}\left(y_{t+1}+y_t\right)^\beta - \frac{1}{2}\left(y_{t+1}-y_t\right)^\beta\right]} \\ &= \frac{2y_i^\beta - 2^{\beta-1}y_i^\beta}{\left[\left(1-y_i\right)\beta y_i^{\beta-1} + 2^{\beta-1}y_i^\beta\right]}\end{aligned} \tag{47}$$

which is strictly smaller than 1 iff

$$\begin{aligned}2y_i^\beta - 2^{\beta-1}y_i^\beta &< \left(1-y_i\right)\beta y_i^{\beta-1} + 2^{\beta-1}y_i^\beta \\ y_i &< \frac{\beta}{\left(2-2^\beta+\beta\right)}.\end{aligned}$$

One can verify that $\frac{\beta}{\left(2-2^\beta+\beta\right)}$ is the FOC solution in region (1) (to see this, one can equate the first part of equation (39) to 0 and solve for y). From Lemma 24 we know that this is the only local extremum in region (1) and that this is a minimum point. Hence, in this region, a steady state y_i is stable if and only if $\frac{dK}{dy}|_{y_i} < 0$. If instead the steady state falls in the second region, where $y > s_l$, then F_3 applies. There

$$\begin{aligned}\frac{dy_{i+1}}{dy_i} &= -\frac{F_{y_i}}{F_{y_{i+1}}} \\ &= -\frac{\frac{1}{2}y_t^\beta + \frac{1}{2}y_{t+1}^\beta - \frac{1}{2}\left(y_{t+1}-y_t\right)^\beta}{-\left[\left(1-\frac{y_t}{2}-\frac{(\bar{s}+1)}{2}\right)\beta y_{t+1}^{\beta-1} + \frac{1}{2}\left(\bar{s}+y_{t+1}+1\right)^\beta - \frac{1}{2}\left(y_{t+1}-y_t\right)^\beta\right]} \\ &= \frac{y_i^\beta}{\left[\left(1-\frac{y_i}{2}-\frac{(\bar{s}+1)}{2}\right)\beta y_i^{\beta-1} + \frac{1}{2}\left(\bar{s}+y_i+1\right)^\beta\right]}\end{aligned} \tag{48}$$

which is smaller than 1 iff

$$\left(1-y_i-\bar{s}\right)\beta y_i^{\beta-1} + \left(\bar{s}+y_i+1\right)^\beta - 2y_i^\beta > 0.$$

[38] Unless the steady state falls exactly at the border between the two regions, where $y = s_l$, in which case there is convergence to this steady state only from one side.

This inequality (short of a factor of 1/2) corresponds to $d(1/K)/dy$ being positive in the second region, as can be seen in the second region of equation (39). That is, in this region too, a steady state y_i is stable if and only if $\frac{dK}{dy}|_{y_i} < 0$. Finally, we know that in steady states, equation (38) holds. If the steady state is in region (1) of this equation (where y is relatively small), then it is independent of $\bar{s}$. Otherwise the steady state is in region (2) (where y is relatively large). Then part (2) of Lemma 29 along with stability under $\frac{dK}{dy}|_{y_i} < 0$ implies that in region (2) y_{ss} is decreasing in $|\bar{s}|$. Therefore, we we get the following observations: (I) if a norm $\bar{s}$ has two stable steady states, one non-truncated (region (1)) and one truncated (region (2)), then y is smaller in the non-truncated steady state. (II) if $\bar{s}'$ has a non-truncated steady state, then it must be that also $\bar{s}$ has a non-truncated steady state (since in region 1 y_{ss} is independent of $|\bar{s}|$). Together, these two observations imply that whenever $\bar{s}'$ has a non-truncated steady state, then either (i) both norms have only a non-truncated steady state, in which case $y_{ss}(|\bar{s}|) = y_{ss}(|\bar{s}'|)$; (ii) $\bar{s}'$ has only a non-truncated steady state while $\bar{s}$ has two steady states, in which case (I) implies we should compare only the non-truncated to each other thus again $y_{ss}(|\bar{s}|) = y_{ss}(|\bar{s}'|)$; (iii) both norms have two steady states, where the non-truncated have the same value of y_{ss} and the truncated are such that $y_{ss}(|\bar{s}'|) \leq y_{ss}(|\bar{s}|)$.

The only cases left to deal with are those where $\bar{s}'$ has only a truncated steady state. If this is the case, then either (i) $\bar{s}$ also has only a truncated steady state, implying that $y_{ss}(|\bar{s}'|) \leq y_{ss}(|\bar{s}|)$ (since we showed that y_{ss}, in truncated steady states, is falling in $|\bar{s}|$); or (ii) $\bar{s}$ has a non-truncated steady state, in which case (I) implies that we must compare this steady state to the unique (and truncated) steady state of $\bar{s}'$. To see that statement (2) of the proposition holds for this case too, take a third norm $\bar{s}''$ such that the distance of $\bar{s}''$ from -1 equals $y_{ss}(|\bar{s}|)$ in the non-truncated steady state. Then, since we know that y_{ss} is constant in region (1), we get that $\bar{s}''$ has a stable steady state with $y_{ss}(|\bar{s}''|) = y_{ss}(|\bar{s}|)$, and this steady state is at the border of region (2), hence $y_{ss}(|\bar{s}'|) \leq y_{ss}(|\bar{s}''|) = y_{ss}(|\bar{s}|)$.

Statement 3): Since $K > K_{\min}(|\bar{s}|)$ is given, we know from the proof of statement (1) that there exists a stable steady state with a single norm $\bar{s}$ such that there is convergence to it from any $y_i < y_{ss}$. To show convergence to a stable steady state from the right, let $y_{conv} \equiv \min\{y_{uss}, y_{\max}(|\bar{s}|)\}$, where y_{uss} is the rightmost steady state in $[0, y_{\max}(|\bar{s}|)]$ that is unstable from both sides, if such a one exists. Suppose y_{uss} does not exist, so that $y_{conv} = y_{\max}(|\bar{s}|)$. Then either there is a unique, and stable, steady state y_{ss}, and therefore $y_{i+1} < y_i$ $\forall y_i \in]y_{ss}, y_{\max}(|\bar{s}|)]$, implying convergence to y_{ss}; or, there may be steady

states in $]y_{ss}, y_{\max}(|\bar{s}|)]$ that are unstable only from one side, in which case $y_{i+1} < y_i$ in their neighborhood, implying once again convergence to y_{ss}. Otherwise $y_{conv} = y_{uss}$, and the complete instability of y_{uss} implies that when $y_i \overset{-}{\rightarrow} y_{uss}$, $y_{i+1} < y_i$, and so there is convergence to a stable steady state from any $y_i < y_{uss}$.[39]

Statement 4): Revisiting Lemma 29, part (1) of that lemma implies that $y_{conv}(|\bar{s}|)$ is increasing in $|\bar{s}|$ whenever $y_{conv} = y_{\max}(|\bar{s}|)$. If instead $y_{conv} = y_{uss}$, then it was shown in the proof to statement (2) of this proposition that $y_{conv}(|\bar{s}|)$ is weakly increasing in $|\bar{s}|$. This concludes the proof.■

References

[1] Acemoglu, D., & Jackson, M. O. (2014). "Social Norms and the Enforcement of Laws" Working paper No. w20369. National Bureau of Economic Research.

[2] BBC (2010): http://news.bbc.co.uk/2/hi/europe/8641070.stm. Accessed on 28/05/2015.

[3] Bernheim, D.B., (1994), "A Theory of Conformity", *Journal of Political Economy,* Vol. 102, No. 5, pp. 841-877.

[4] M. Blumenthal, C. Christian, and J. Slemrod. (2001) "Do Normative Appeals affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota". *National Tax J.*, 54(1):125-138.

[5] Borsari, B., & Carey, K. B. (2001). "Peer influences on college drinking: A review of the research". *J. of substance abuse*, 13(4), 391-424.

[6] Carvalho, J. P. (2012). "Veiling". *The Quarterly Journal of Economics*, Vol 128, Iss 1,pp. 337-370.

[7] Carvalho, J. P. (2014). "Coordination and Culture", mimeo Dept of Economics, University of California Irvine.

[8] Centola, D., Willer, R., & Macy, M. (2005). "The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms". *American J. of Sociology*, 110(4), 1009-1040.

[9] Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). "A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior". *Adv. in experimental social psych*, 24(20), 1-243.

[10] Cialdini, R. B. (2003). "Crafting normative messages to protect the environment". *Current directions in psychological science*, 12(4), 105-109.

[39] There may be two stable steady states to the left of y_{uss}, with convergence from small values of y_i to the first steady state and from large values of y_i to the second steady state, but this statement, and hence statement (3) of the proposition, holds in this case too.

[11] Clark, A. E., & Oswald, A. J. (1998). "Comparison-concave utility and following behaviour in social and economic settings." *J. of Public Economics*, 70, 133-155.
[12] Cohen, D. (2001). "Cultural variation: considerations and implications". *Psychological bulletin*, 127(4), 451.
[13] Cohen, D., Nisbett, R. E., Bowdle, B., & Schwarz, N. (1996). "Insult,aggression, and the southern culture of honor". *Journal of Personality and Social Psychology*, 70, 945-960.
[14] Colson, E. (1975) *Tradition and contract: The problem of order*. Chicago: Aldine.
[15] Fields, J. M., & Schuman, H. (1976). "Public beliefs about the beliefs of the public". *Public Opinion Quarterly*, 40(4), 427-448.
[16] Gladwell, M. (2000). *The tipping point*. Boston: Little, Brown.
[17] Glaeser, E. L., & Scheinkman, J. (2000). "Non-market interactions" (No. w8053). National Bureau of Economic Research.
[18] Granovetter, M., (1978), "Threshold Models of Collective Behavior", *The American J. of Sociology,* Vol. 83, No. 6, pp. 1420-1443.
[19] Jackson, M. O., & Zenou, Y. (2014). "Games on networks". *Handbook of game theory, 4*.
[20] Kandel E., Lazear, E. P., (1992), "Peer Pressure and Partnerships," *The J. of Political Economy*, Vol. 100, No. 4, pp. 801-817.
[21] Kitts, J. A. (2003). "Egocentric bias or information management? Selective disclosure and the social roots of norm misperception". *Social Psychology Quarterly*, 222-237.
[22] Kuran, T., (1995), "The Inevitability of Future Revolutionary Surprises," *The American J. of Sociology*, Vol. 100, No. 6, pp. 1528-1551.
[23] Kuran, T., & Sandholm, W. H. (2008). "Cultural integration and its discontents". *The Rev. of Economic Studies*, 75(1), 201-228.
[24] Lewis, D. K. (1969): "*Convention: a Philosophical Study*". Harvard University Press.
[25] Manski, C.F.,Mayshar, J. (2003)"Private Incentives and Social Interactions: Fertility Puzzles in Israel," *J. of the European Economic Association*, Vol. 1, No.1, pp. 181-211.
[26] Michaeli, M. & Spiro, D., (2015), "Norm conformity across societies," *J. of Public Economics*, Vol. 132, pp. 51-65.
[27] Milgram, S. (1992). "The experience of living in cities". In S. Milgram (Ed.), The individual in a social world: Essays and experiments (pp. 10-30). New York: McGraw-Hill.
[28] Miller, D., & Prentice, D. (1994). "Collective errors and errors about the collective," *Personality and Social Psychology Bulletin*, 20, 541-550.

[29] Neary, P. R. (2012). "Competing conventions". *Games and Economic Behavior*, 76(1), 301-328.
[30] O'Gorman, H. J. (1975). "Pluralistic ignorance and white estimates of white support for racial segregation". *Public Opinion Quarterly*, 39(3), 313-330.
[31] Ozgur, O. (2011), "Local interactions." In: Benhabib, J., Bisin, A., Jackson, M., (Eds.), Handbook of Social Economics, Volume 1, North Holland.
[32] Robinson, C. E. (1932). *Straw votes*. New York: Columbia University Press.
[33] Schanck, R. L. (1932). "A study of a community and its groups and institutions conceived of as behaviors of individuals". *Psychological Monographs*, 43(2), i.
[34] Schelling, T. C. (1960). *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.
[35] Vandello, J., & Cohen, D. (2000). "Endorsing, enforcing, or distorting? How southern norms about violence are perpetuated". Unpublished manuscript, Princeton University, Princeton, NJ.
[36] Wilson, J. Q., & Kelling, G. (1982). "Broken windows". *Atlantic*, 29-38.
[37] Young, H. P. (1993),"The Evolution of Conventions," Econometrica, 61(1), 57-84.