האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

THE DISTRIBUTION OF REVEALED PREFERENCES UNDER SOCIAL PRESSURE

By

MOTI MICHAELI and DANIEL SPIRO

Discussion Paper # 609 May 2012

מרכז לחקר הרציונליות

CENTER FOR THE STUDY
OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

The Distribution of Revealed Preferences under Social Pressure[*]

Moti Michaeli[†]& Daniel Spiro[‡]

Abstract

This paper studies theoretically the aggregate distribution of revealed preferences when heterogeneous individuals make the trade off between being true to their real opinions and conforming to a social norm. We show that in orthodox societies, individuals will tend to either conform fully or ignore the social norm while individuals in liberal societies will tend to compromise between the two extremes. The model sheds light on phenomena such as polarization, alienation and hypocrisy. We also show that societies with orthodox individuals will be liberal on aggregate unless the social norm is upheld by an authority. This suggests that orthodoxy cannot be maintained under pluralism.

1 Introduction

The question addressed by this paper is very simple. Think of a society where every individual has a private opinion with respect to some issue, and those opinions are uniformly distributed. Suppose also that individuals dislike lying about their opinions. Then, the revealed preferences, i.e. what people declare openly, should be uniform as well. But now we

[*] We wish to thank Florian Biermann, Sergiu Hart, John Hassler, Arie Kacowicz, Assar Lindbeck, Arie Michaeli, Sten Nyberg, Ignacio Palacios-Huerta, Torsten Persson, Jörgen Weibull, Robert Östling and seminar participants at Hebrew University, Stockholm University and Stockholm School of Economics for valuable comments.

[†] Department of Economics, and Center for the Study of Rationality, The Hebrew University. Email: motimich@gmail.com. Moti receives funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° [249159].

[‡] Corresponding author, Institute for International Economic Studies, Stockholm University, daniel.spiro@iies.su.se, +46 70 5849860.

introduce a pressure to conform to some “social norm”. How will that change the distribution of revealed preferences?

We study this question analytically under a variety of societal traits. Although the problem faced by one individual is quite simple, it turns out that the outcomes at the aggregate level are quite diverse and largely depend on the characteristics of society. To help fix ideas, we label an orthodox society as one where the social pressure is concave, so that small deviations from the norm are punished heavily but large deviations only somewhat more. As an opposite label, liberal societies are those with a convex pressure function, so that society is tolerant to deviations as long as these are not too extreme. Similarly, at an individual level, we label opinions as rigid if the displeasure from lying is concave and lax if it is convex.

In the basic case, where individuals are punished for deviations from one common social norm, we find that liberal societies will induce individuals to compromise between conforming and stating their true opinions. Depending on how lax individuals’ personal opinions are, as compared to how liberal is society, the total outcome will either be bimodal polarization or unimodal concentration.

Meanwhile, orthodox societies will exhibit no compromise. An individual will either completely conform or completely speak her mind. This is due to the fact that when social pressure is concave, one essentially needs to fully conform in order to alleviate the pressure. Although all orthodox societies will display this lack of compromise, the further traits of individuals will determine who conforms and who ignores the norm. Depending on how rigid individuals’ personal opinions are, as compared to how orthodox society is, we will either see “alienation”, where extremists (those with opinions far from the norm) follow their hearts, or “hypocrisy”, i.e. a case where the norm is maintained by those opposing it the most.

An overarching analytical result that we find is that the relative curvature of social pressure compared to inner preferences determines who in society is most affected by social pressure. More precisely, if the relative concavity of social pressure (arising from deviations from the norm) is higher than that of the cognitive dissonance (from deviating from one’s bliss point), then individuals with inner preferences close to the social norm will concede relatively more than those with inner preferences far from it, and vice versa. The intuition for this is fairly simple – when social pressure is relatively concave, it affects small deviations from the norm relatively more, i.e. those with preferences close to the norm.

Another outcome that clearly separates orthodox and liberal societies is that if the norm represents the average declared opinion, then in liberal societies it will also represent the average true (hidden) opinion in society[1]. In contrast, in orthodox societies, we may well obtain a social norm centred on a point which is far from what people really think. In a sense, this highlights the prospects for a democracy which is representative of the true opinions in various societies.

In an extension of the basic model, we look at how individuals' pressure on each other aggregates up to a pluralistic societal pressure, and how this will affect the distribution of revealed preferences. A surprising result is that societies where the sources of pressure are heterogenous, but where each single pressure is concave, will, after the aggregation of all single pressure functions, have a convex societal pressure. Our interpretation of this is that a society with orthodox individuals with heterogenous opinions will, on aggregate, be liberal. In order to maintain an orthodox society, there needs to be a central authority that sanctions individuals, otherwise, the individual true opinions need to be very homogenous. This also alludes to the (im-)possibility of keeping an orthodox society once norms are shaped in a pluralistic way. We also show that in these cases – of pluralistic societies but with orthodox and rigid individuals – "hypocrisy" will arise, i.e. extremists will claim to be more moderate than many moderates.

This subject can be relevant at a few different levels. First, non-economic incentives are obviously an important part in decision making. In that sense, this paper contributes to the large literature on social norms by analyzing how the aggregate effects play out. This may have a bearing on how efficient and how broad we can expect different policy instruments to be – depending on whether they affect the deviations from desired behavior in general, or different types of deviations differently. Second, we provide an explanation for why we observe such a diversity of distributions of behavior and opinions in reality – other than the simplest explanation saying that whatever we observe is exactly the distribution of true inner preferences. Third, the paper indirectly highlights a problem in survey data and how we should interpret the observed distributions. For example, Manski (1993) has pointed out the problem of separating revealed and inner preferences empirically. We align with his view by showing that one cannot simply deflate the effects of social pressure across the board in some straightforward manner. Rather, social pressure can have non-linear and – perhaps more gravely – also ordinal effects on the revealed distribution. Finally, it clarifies that the commonly used assumption in many economic theories, of a linear-quadratic combination of utilities, is far from innocuous.

In this paper, social norms may represent consensual political opin-

[1] This is true at least if the distribution of true opinions is uniform.

ions, work ethics, or any other unwritten rules of conduct. This way social pressure does not only cover what is often referred to as cheap talk, but also more action oriented conduct, such as buying ethically produced goods, being unemployed, cheating on taxes, or choosing which type of car to buy. One can also interpret the sanctioning system as being a judicial punishment and thus, the model may represent how law-obedience is distributed in a society (particularly when there are several levels of disobedience, e.g. in the case of breaking speed limits).

The literature on social norms is vast and has been applied to various economic subjects such as choices of neighborhood (e.g. Schelling, 1971), herd behavior (e.g. Granowetter, 1978) and unemployment (e.g. Lindbeck et al, 2003). The most common formal approach is to let the stances of individuals be binary. This naturally limits any investigation of distributions (e.g. Brock & Durlauf, 2001; Lopez-Pintado & Watts, 2006). An exception is Bernheim's (1994) work on conformity. Just like in the current paper, he does not only investigate a continuum of inner blisspoints, but also a continuum of stances from which individuals may choose. Perhaps the most important difference between our model and that of Bernheim is that he assumes that people are judged by the type they are conceived to really be, while we assume that people are judged by their actions regardless of their true type. Plausibly, there is merit to both approaches, and the distinction turns out to have a great impact on the results. A further difference is that Bernheim assumes that both the social pressure function and the inner preference function are concave. As we will show, a number of other distributions arise if we only deviate from concave cases, and the purpose of this paper is to present this complete set of possible distributions and link them to societal traits. To this end, we will be less general in the paper by choosing a specific type of function – namely a power function – and let it take on both concave and convex shapes. The benefit of this functional form is that it is tractable and, more importantly, that it provides us with a set of parameters and results to which we can give reasonable interpretations in real world terms. Most results also hold with general functional forms.

The next section outlines the model with general function forms and derives some general results. Section 3 presents and interprets the model with power functions Then, the results of the model are presented in sections 4-8) according to the different possible subcases. Note that we assume that all individuals feel pressure from a common source (the social norm). However, since the resultant distributions are qualitatively independent of the exact location of the social norm, we present the results in a way which is agnostic about how the social norm is formed[2].

[2] You can think of an endogenous source for the norm, like the mean or the median

To characterize the general equilibrium, the subsequent section 9 analyzes which locations of the social norm can be achieved in equilibrium if the social norm is endogenously determined by the average stance in society. Section 10 suggests some casual observations that the model may explain. Next, in sections 11 and 12, we extend the model to a case where societal pressure is formed by the aggregation of individuals' pressure on each other and show how that affects stances in orthodox societies. Finally, we devote part of the concluding section 13 to discussing how an empirical test of the model may be approached. To keep the paper readable, the more elaborate analytical derivations and proofs are covered in the appendix.

2 The model and general results

An individual is represented by a type $t \in (t_l, t_h)$. The inner (hidden) preference of a type t is

$$D\left(t-s\right),\ \frac{dD}{d\left(\left|t-s\right|\right)}>0,$$

where s is the (openly declared) stance of an individual and thus a choice variable. If a person minimizes D only, it is immediate that $s\left(t\right)=t$. This way t represents the blisspoint of an individual in fulfilling her inner preferences and D can be interpreted as the cognitive dissonance or displeasure felt by taking a stance that is not in line with this bliss point. We can, for example, think of t as the position on a political scale.

Now, assume that an individual that takes s as a stance feels a social pressure $P\left(s-\bar{s}\right)$, where $\bar{s}$ can be understood as a social norm[3], with

$$\frac{dP}{d\left(\left|s-\bar{s}\right|\right)}>0.$$

The total disutility (or loss) of an individual is then the sum of the cognitive dissonance and the social pressure.

$$L\left(t,s\right)=D\left(t,s\right)+P\left(s,\bar{s}\right) \tag{1}$$

So, on the one hand, the individual feels an increasing inner displeasure (or cognitive dissonance) from taking a stance different than the bliss

of the distribution of stances in society or, alternatively, an exogenous source for the norm, such as rules of conduct determined by a religious authority

[3] The qualitative results from the model are independent of the location of $\bar{s}$. In section 9, a general equilibrium $\bar{s}$ is determined given that it is the average of all stances. Since there may be many other ways of determining $\bar{s}$, we prefer to remain agnostic about its location and origin as far as we can in the paper.

point. On the other hand, the more social pressure that is exerted the further the stance is from the norm. Then, it is immediate that each individual will take a stance somewhere in between (and including) its inner blisspoint and the social norm. That is

$$\forall t, s^* (t) \in \begin{cases} [\bar{s}, t], & \text{if } \bar{s} \leq t \\ [t, \bar{s}], & \text{if } t > \bar{s} \end{cases},$$

where $s^* (t)$ is the stance that minimizes the loss for type t. For the sake of tractability, we will restrict the analysis to cases where $s (t) \leq t$.[4] The analysis when $s (t) > t$ is similar. The first-order condition

$$L' = P' (s) - D' (t - s), \tag{2}$$

is equal to zero in inner extreme points while the second-order condition,

$$L'' = P'' (s) + D'' (t - s), \tag{3}$$

is positive in minimum points. Denoting the optimal stance by s^*, we then have the inner solutions

$$P' (s^*) = D' (t - s^*). \tag{4}$$

We now turn to look at $s^*(t)$, i.e. the function describing the inner solution (if it exists) for every t. More specifically, we concentrate on ranges of t for which the inner solution exists, and where $s^*(t)$ is continuous and twice differentiable[5]. Then, by way of the implicit function theorem, we can derive the following results.

$$\frac{ds^*}{dt} = \frac{D'' (t - s^*)}{P'' (s^*) + D'' (t - s^*)} \tag{5}$$

$$\frac{d^2 s^*}{dt^2} = \frac{\left[D''' (t - s^*) (P'' (s^*))^2 - P''' (s^*) (D'' (t - s^*))^2\right]}{(P'' (s^*) + D'' (t - s^*))^3} \tag{6}$$

A further assumption that simplifies the exposition of the results is that $t \sim U (t_l, t_h)$. This has no bearing on the stances at an individual level, i.e. $s^* (t)$, but will, of course, affect the aggregate distribution. As we want to present various qualitatively different distributions that emerge from the model, a uniform distribution of types implies that the effects from our model are analyzed in isolation from the reasons of the underlying distribution of hidden opinions.

To compare the extent of conformity to the norm and compromise by different individuals in society, we will use three measures.

[4] A sufficient requirement for the upcoming analysis to hold is that both P and D are three times continuously differentiable.

[5] This implies that we only look at ranges either where the solution is unique or where there are no discrete jumps between solutions.

Definition 1 *The conformity of t is $|s^*(t) - \bar{s}|$.*

This measure is of how close to the norm is an individual's stance. We will say that t conforms more than t' if $|s^*(t) - \bar{s}| \geq |s^*(t') - \bar{s}|$.

Definition 2 *The absolute concession of t is $|t - s^*(t)|$.*

This second measure essentially catches how far from its true opinion an individual's stance is, i.e. how large a step towards the norm an individual is taking. So, t is said to concede absolutely more than t' if $|t - s^*(t)| \geq |t' - s^*(t')|$.

Definition 3 *The relative concession of t is $|t - s^*(t)| / |t - \bar{s}|$.*

This final measure is meant to portray how much an individual is giving up on her beliefs compared to how much she could, maximally, if she completely conformed to the norm. We say that t concedes relatively more than t' if $|t - s^*(t)| / |t - \bar{s}| \geq |t - s^*(t')| / |t' - \bar{s}|$. Following these definitions, a useful lemma applies for types above the social norm[6].

Lemma 1 *For $t \geq \bar{s}$:*

1. *Conformity is locally weakly decreasing in t iff $\frac{ds^*}{dt} \geq 0$.*

2. *Absolute concession is locally weakly increasing in t iff $\frac{ds^*}{dt} \leq 1$.*

3. *In corner solutions, relative concession is locally constant. In inner solutions, relative concession is locally weakly increasing in t iff $(s^* - \bar{s}) P''(s^* - \bar{s}) \geq (t - s^*) D''(t - s^*)$.*

Proof. *1) and 2) trivially follow from definitions 1 and 2. 3) In corner solutions $s(t) \in \{\bar{s}, t\}$ which implies that, locally, relative concession is either equal to 1 or 0. For inner solutions: By differentiating the expression for relative concession w.r.t. t, performing a few algebraic steps making use of equations 4-6, it can be verified that the derivative is proportional to $\frac{(s^*-\bar{s})P''(s^*-\bar{s})-(t-s^*)D''(t-s^*)}{P''(s^*-\bar{s})+D''(t-s^*)}$. In min points the denominator is positive and the inequality then follows.* ■

For the upcoming analysis, we also need an expression for the distribution of stances. In our case of a uniform distribution of types, the partial probability density function, $pPDF$, of stances is as follows[7].

$$PDF(s^*) = \frac{1}{t_h - t_l}\frac{dt}{ds^*} \text{ when } \frac{ds^*}{dt} \neq 0$$

$$PDF(s^*) = \frac{\tilde{t} - t_{\min}}{t_h - t_l} \text{ when } \frac{ds^*}{dt} = 0 \text{ where } \forall t \in \left[t_{\min}, \tilde{t}\right],\ s^*(t) = s^*$$

[6] Equivalent statements apply to types below the social norm.

[7] For derivations see the appendix section 14.1.

The first expression characterizes the PDF for inner solutions while the second expression characterizes the PDF for corner solutions[8]. In inner solutions, the following results apply.

Lemma 2 *In inner solutions, the $pPDF$ is locally strictly increasing at s^* if $\frac{d^2s^*}{dt^2}$ is negative, and strictly decreasing at s^* if $\frac{d^2s^*}{dt^2}$ is positive.*
Proof. *See the appendix.* ■

Together with equation 6, the lemma expresses under what conditions we should expect a larger mass of stated opinions as we move away from the social norm.

3 The model with power functions

Nearly all upcoming results can be generalized using the previous lemmas and equations. For tractability and to facilitate the interpretation, we will now assume that cognitive dissonance and social pressure are power functions.

$$\begin{aligned} D(t,s) &= |t-s|^{\alpha} \ , \ \alpha \geq 0 \\ P(s,\bar{s}) &= K\,|s-\bar{s}|^{\beta} \ , \ \beta \geq 0. \end{aligned}$$

These functions are symmetric around $t = s$ and $s = \bar{s}$, respectively. For conservation of space, we will therefore mainly only present the problem and solution for $t \geq \bar{s}$ where we get the following minimization problem.

$$\min_{s} \left\{ (t-s)^{\alpha} + K(s-\bar{s})^{\beta} \right\}$$

with a first-order condition

$$-\alpha (t-s)^{\alpha-1} + \beta K (s-\bar{s})^{\beta-1} = 0 \tag{7}$$

and a second-order condition for an internal local minimum point.

$$(\alpha-1)\,\alpha (t-s)^{\alpha-2} + (\beta-1)\,\beta K (s-\bar{s})^{\beta-2} > 0. \tag{8}$$

At this point, it may be useful for the intuition to provide a loose interpretation of the parameters. Obviously, all results go through also without these interpretations. Most immediate is that K represents the

[8] Note that these expressions catch the "local" contribution to the PDF. E.g. in cases where we have both inner and corner solutions, these may overlap and the total PDF is characterized by a combination of an inner solution $pPDF$ and a corner solution $pPDF$. Also note the we have normalized the total mass of individuals in society to 1. Thus, a large gap between t_h and t_l represents a society which is spread out in terms of true opinions.

weight of social pressure relative to the cognitive dissonance. If P are legal repercussions, then K represents how harsh the punishment system is in general. In comparison, β catches how different deviations from the norm are sanctioned in relation to each other. When $\beta \leq 1$, already small deviations from the established norm or rule are fairly heavily sanctioned but only a minor distinction is made between small and large deviations. We believe that this kind of punctiliousness represents many orthodox societies since they often emphasize being "true to the book" but do not distinguish so much between large and small wrongdoings. As an opposite label, liberal societies are not very meticulous about small non-normative expressions as long as they are not too fundamental or far from the consensus. Hence, it is represented by $\beta \geq 1$. As for α, it catches how particular individuals are about taking a stance which is different from what they feel inside. $\alpha \leq 1$ represents the rigid approach where, once you deviate even slightly from your bliss point, it makes little marginal difference to deviate a great deal. $\alpha \geq 1$ represents a more lax approach towards deviations from one's bliss point – as long as the deviation is not too large, it matters little. Naturally, the same society may exhibit different β and K depending on the issue and likewise α may vary between societies and topics but we will analyze them one case at a time[9]. The relative size of α and β will turn out to be decisive in forming the distribution of revealed preferences and will also determine which individuals will concede the most.

4 A liberal society with lax personal opinions

We start by examining the case when α and β are greater than 1. From the second-order condition (8), it is immediate that there is an internal solution for every type t in this case. The properties of the resultant distribution are summarized in the following proposition.

Proposition 1 *If $\alpha \geq 1$ and $\beta \geq 1$ then:*

1. *If $\alpha < \beta$ and $\bar{s} \notin \{t_l, t_h\}$, then $|s^*(t) - \bar{s}|$ is increasing and concave, the distribution is bimodal and the relative concession is increasing with $|t - \bar{s}|$.*

2. *If $\alpha > \beta$ and $\bar{s} \notin \{t_l, t_h\}$, then $|s^*(t) - \bar{s}|$ is increasing and convex, the distribution is unimodal and the relative concession is decreasing with $|t - \bar{s}|$.*

[9] We also abstract from the possibility of varying the parameters at an individual level.

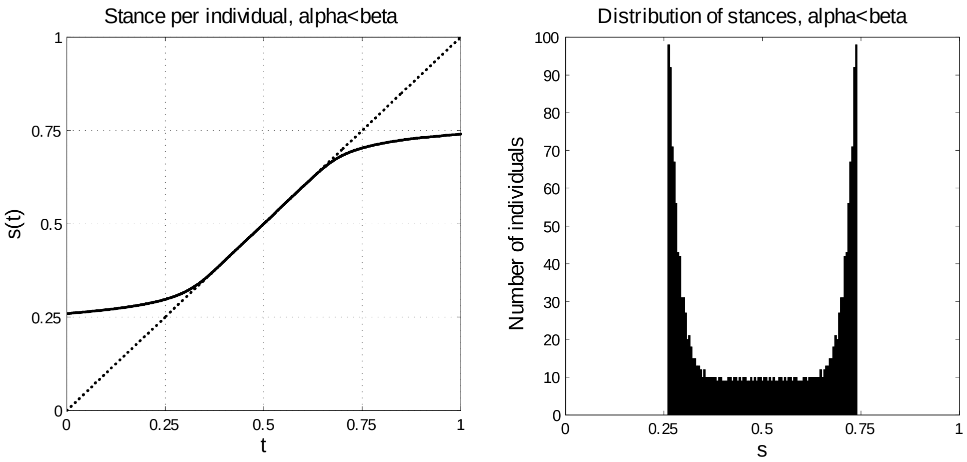

Figure 1: $1 < \alpha < \beta$ with $\bar{s} = .5$, $t \sim U(0,1)$. The left-hand schedule depicts $s^*(t)$ (full line) and in comparison to the $s = t$ (dashed line). The right-hand schedule depicts the probability density function.

3. *If $\alpha = \beta > 1$, then $|s^*(t) - \bar{s}|$ is increasing and linear, the distribution is uniform and the relative concession is constant.*

4. *Conformity is decreasing and the absolute concession is increasing in $|t - \bar{s}|$.*

Proof. *Since the functions are symmetric around $\bar{s}$, we settle by presenting the proof for the range of $t \geq \bar{s}$. That every t has a unique inner solution can easily be verified using 7 and 8. The statements on the convexity and concavity of $s^*(t)$ follow from applying the implicit function theorem to 7. The statements regarding relative concession follow from all types with an inner solution and by inserting the appropriate expressions into part 3 of Lemma 1 and noticing that $\beta > \alpha$ implies a positive sign and vice versa. By verifying that $\frac{d^2t}{ds^{*2}} > 0$ when $\beta > \alpha$ in equation 6 follows by Lemma 2 that the $pPDF$ is increasing with the distance to $\bar{s}$. As $s^*(t)$ is monotonic, the $pPDF$ represents total PDF. From the symmetry of the functions around $\bar{s}$, it then follows that iff $\bar{s} \in]t_l, t_h[$ the distribution is bimodal when $\beta > \alpha$. For unimodality when $\alpha > \beta$ and for uniformity when $\alpha = \beta$, a similar proof applies. The convexity of P and D implies that with $\forall t \geq \bar{s}$, we have $0 \leq \frac{ds^*}{dt} = \frac{D''(t-s^*)}{P''(s^*)+D''(t-s^*)} \leq 1$. Hence, by parts 1 and 2 of Lemma 1, it follows that conformity is decreasing and absolute concession is increasing $\forall t \geq \bar{s}$.* ■

The results are visualized in figures 1 and 2 where the left-hand schedule represents the resulting function $s^*(t)$ and the right-hand schedule represents the resultant distribution (the probability density function)

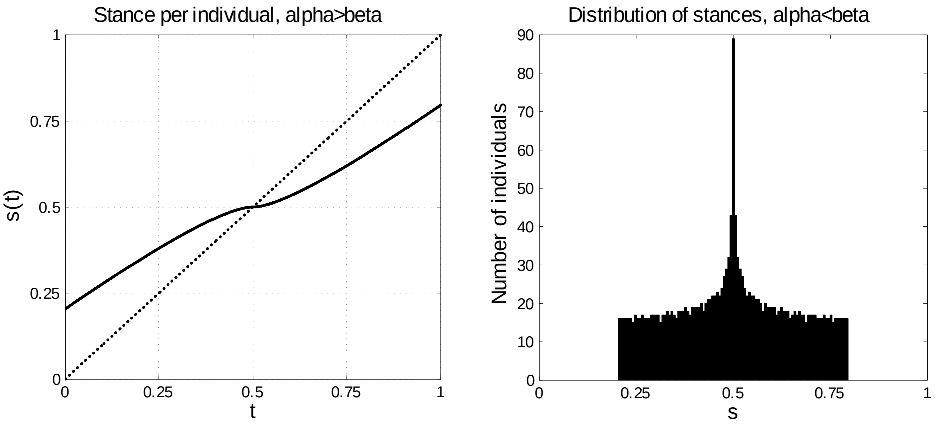

Figure 2: $1 < \beta < \alpha$ with $\bar{s} = .5$, $t \sim U(0,1)$. The left-hand schedule depicts $s^*(t)$ (full line) and in comparison to the $s = t$ (dashed line). The right-hand schedule depicts the probability density function.

given a uniform distribution of bliss points. The intuition is as follows. When $\beta > 1$ only extremists (t far from $\bar{s}$) feel any substantial pressure to comply with the norm. Then, as is the case here, when $a > 1$, an individual's inner preferences are also open for deviations from the bliss point, as long as the deviation is not too large. The important question is then which of the cognitive dissonance or the social pressure that is more open to large deviations, i.e. which of α and β that is the largest.

When $\beta > \alpha$ (the first result in the proposition), only extreme types will feel enough social pressure to actually take a large step from their blisspoint. Meanwhile moderates (t close to $\bar{s}$) will hardly be inclined to move from their blisspoint. There will then be a concentration of extreme types at a certain distance on each side of the norm. As α falls, the population becomes more polarized with a higher concentration at the peaks and less individuals taking intermediate stances (the "smile" in figure 1 becomes deeper).[10]

To get the intuition for the second part of the proposition ($\alpha > \beta$), it may be easiest to imagine a very large α. Then, an individual does hardly feel any dissonance from deviating a little from t. A moderate person may then just as well choose a stance very close to $\bar{s}$ in order to minimize the social pressure. An extreme type, however, will not be willing to move equally close to $\bar{s}$ since the inner discomfort will then be very large. Thus, in this scenario, moderates tend to concede relatively more to the norm.

[10] If $\bar{s}$ is biased towards one of the extremes, the peak on that side will be lower, but the distribution will be symmetrical in a neighborhood around $\bar{s}$.

The last part of the proposition essentially expresses that the two previous scenarios converge as the difference between β and α falls, implying that we get a uniform distribution of stances.

5 An orthodox society with rigid personal opinions

When $\beta \leq 1$, society is intolerant to small deviations from the consensus but does not distinguish to any large extent between moderate and large deviations. Likewise, when $\alpha \leq 1$, people are already sensitive to small deviations from their inner blisspoint but additional distance does not add much to their inner discomfort.

It is now immediate from the second-order condition (8) that any inner solution is a maximum implying that optimality will be found at either of the corners (at $s^*(t) = t$ or $s^*(t) = \bar{s}$). This is also intuitive since by taking a stance in between t and $\bar{s}$, one both feels great dissonance and is heavily pressured when the functions are concave.

Proposition 2 *If $\beta \leq 1$ and $\alpha \leq 1$.*

1. *If $\beta < \alpha$ then iff $|t - \bar{s}| \geq K^{\frac{1}{\alpha-\beta}}$, $s^*(t) = t$, and iff $|t - \bar{s}| < K^{\frac{1}{\alpha-\beta}}$, $s^*(t) = \bar{s}$. The distribution is unimodal and discontinuous with a peak at $\bar{s}$ (made of moderate types) and uniform tails at the extreme ends of the range (made of extreme types). Relative concession and conformity are weakly decreasing in $|t - \bar{s}|$. Absolute concession increases in $|t - \bar{s}|$ for $|t - \bar{s}| < K^{\frac{1}{\alpha-\beta}}$, then sharply decreases to zero (i.e. $|t - s^*(t)| = 0$) at $|t - \bar{s}| = K^{\frac{1}{\alpha-\beta}}$, and remains at a zero level for $|t - \bar{s}| > K^{\frac{1}{\alpha-\beta}}$.*

2. *If $\alpha < \beta$ then iff $|t - \bar{s}| \leq K^{\frac{1}{\alpha-\beta}}$, $s^*(t) = t$, and iff $|t - \bar{s}| > K^{\frac{1}{\alpha-\beta}}$, $s^*(t) = \bar{s}$. The distribution of stances is continuous and unimodal with a peak at $\bar{s}$ (made of extreme types) and uniform tails (made of moderate types). Relative and absolute concession are weakly increasing in $|t - \bar{s}|$. Conformity decreases in $|t - \bar{s}|$ for $|t - \bar{s}| < K^{\frac{1}{\alpha-\beta}}$, then sharply increases to full conformity at $|t - \bar{s}| = K^{\frac{1}{\alpha-\beta}}$, and remains at full conformity for $|t - \bar{s}| > K^{\frac{1}{\alpha-\beta}}$.*

3. *If $\alpha = \beta$ then $s^*(t) = t \forall t$ iff $1 > K$. Else (if $1 < K$) $s^*(t) = \bar{s} \forall t$. The distribution is then uniform or concentrated at $\bar{s}$, respectively. Relative concession is constant, absolute concession is either constant (if $1 > K$) or increasing (if $1 < K$) in $|t - \bar{s}|$ and conformity is either constant (if $1 < K$) or decreasing (if $1 > K$) in $|t - \bar{s}|$.*

Proof. *We prove part 1. The second-order condition (equation 8) is positive when $\alpha, \beta < 1$, which implies that any inner extreme point is a*

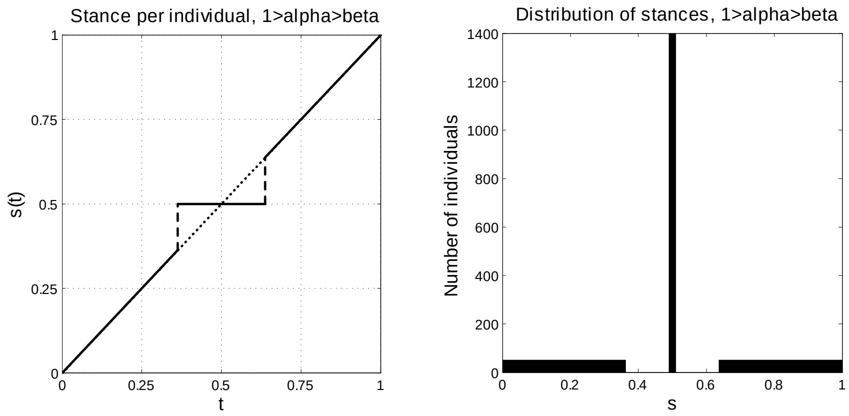

Figure 3: $\beta < \alpha \leq 1$ with $\bar{s} = .5$, $t \sim U(0,1)$. The left-hand schedule depicts $s^*(t)$ (full line) and in comparison, the $s = t$ (dashed line). The right-hand schedule depicts the probability distribution function.

maximum. The corner solutions are then either $L(s = \bar{s}) = |t - \bar{s}|^{\alpha}$ *or* $L(s = t) = K|t - \bar{s}|^{\beta}$*.* $L(s = \bar{s}) < L(s = t)$ *iff* $|t - \bar{s}| < K^{\frac{1}{\alpha-\beta}}$ *which implies that* t *close to* $\bar{s}$ *chooses* $s^*(t) = \bar{s}$ *while those far from* $\bar{s}$ *choose* $s^*(t) = t$*. The distribution then follows from this. In the segment of types choosing* $s^*(t) = \bar{s}$*, the relative concession is equal to* 1 *while in the segment of types choosing* $s^*(t) = t$*, the relative concession is* 0*. From this, it follows that the relative concession is weakly decreasing with distance to* $\bar{s}$*. Similarly, conformity is decreasing with distance to* $\bar{s}$*. Finally, for* t *near* $\bar{s}$*, absolute concession is equal to* $|t - \bar{s}|$ *which is increasing in* t *while for* t *far from* $\bar{s}$*, absolute concession is zero which is constant. For parts 2 and 3, similar proofs apply.* ■

In part 1 of the above proposition, society is more orthodox than personal opinions are rigid. Individuals with opinions close enough to the social norm ($t \in \left[\bar{s} - K^{\frac{1}{\alpha-\beta}}, \bar{s} + K^{\frac{1}{\alpha-\beta}}\right]$) will choose to fully comply while individuals with opinions far enough from the norm will simply cope with the full social pressure and choose the inner bliss point as their stance. The intuition is that these “extreme” people are not willing to take a stance that is close enough to the norm to alleviate the pressure. Then, since deviating even a little from their inner bliss point is very painful, they might as well take a stance that is in line with what they really feel inside. Altogether, this creates alienation in society where one either conforms fully with the norm or follows one’s heart. If a person feels that it is not possible to fulfill the norm, there is no point in trying to be a little bit accepted since this will hardly make a difference anyway. This way a society that is not tolerant to small deviations from the norm

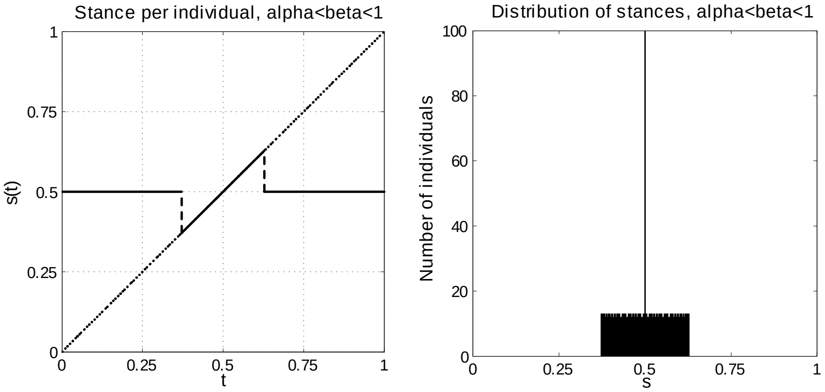

Figure 4: $\alpha < \beta \leq 1$ with $\bar{s} = .5$, $t \sim U\left(0,1\right)$. The left-hand schedule depicts $s^{*}\left(t\right)$ (full line) and in comparison the $s = t$ (dashed line). The right-hand schedule depicts the probability distribution function.

will tend not to succeed in moderating extreme people's stances.

We will now continue with the mirror image of the previous case – a social pressure which is less orthodox than personal opinions are rigid (part 2 of proposition 2). The observable outcome of this case is a distribution that looks close to a standard bell-shape. But there is an important twist. The concentration of stances at $\bar{s}$ consists of individuals with extreme inner blisspoints. The extreme types' declarations are more moderate than those of the moderates. This makes for a conformity which is increasing with the type's distance to the norm – an outcome that we will label "hypocrisy". The intuition is that moderates are now unwilling to conform since this would inflict too great displeasure when the dissonance is relatively more concave. For extremists, however, not conforming will imply too great social pressure since $P\left(t,\bar{s}\right)$ is increasing relative to $D\left(t,\bar{s}\right)$ with the distance to the norm $\left(\left|t-\bar{s}\right|\right)$.

The last part of proposition 2 expresses that as the dissonance and the social pressure become equally concave, the whole population either conforms completely or not at all.

6 An orthodox society with lax personal opinions

When $\beta \leq 1$ (small deviations from the norm matter more on the margin than large deviations) and $\alpha > 1$ (only large deviations from the blisspoint create dissonance), we get a mix of corner and inner solutions in line with the following proposition.

Proposition 3 *If $\beta \leq 1 < \alpha$:*

1. *$\exists \hat{t} > \bar{s}$ such that $s^{*}(t) = \bar{s}$ for every $t \in [\bar{s},\hat{t}[$ and such that $s^{*}\left(t\right) \in]\bar{s},t[$ for every $t \geq \hat{t}$. This image is mirrored at $\bar{s}$.*

2. *$s^*(t)$ is constant in the range $\left[\bar{s}, \hat{t}\right]$ and increasing in the range $\left[\hat{t}, t_h\right]$.*

3. *For a broad enough range of types, the distribution is discontinuously trimodal with a peak at $\bar{s}$ and a detached section on each side which rises towards the end of the range. For a sufficiently narrow range of types, the distribution is degenerate at $\bar{s}$.*

4. *Conformity and relative concession are weakly decreasing in $|t - \bar{s}|$. Absolute concession is increasing in $|t - \bar{s}|$ for t such that $|t - \bar{s}| < \left|\hat{t} - \bar{s}\right|$, then sharply decreases when $|t - \bar{s}| = \left|\hat{t} - \bar{s}\right|$, and gradually keeps decreasing as $|t - \bar{s}|$ grows further.*

Proof. *See the appendix.* ■

A visualization of the proposition can be seen in figure 5. The distribution has a peak at $\bar{s}$ and, for a sufficiently broad range of types and a sufficiently centered $\bar{s}$, tails with peaks which are increasing towards the edges[11]. What is shown by the proposition is that a group of individuals who are extreme enough will all choose an inner solution and that more extreme individuals conform relatively and absolutely less. Meanwhile, moderates will completely conform with the norm. This will create a concentration of individuals at the norm and two tails where the concentration of individuals is increasing towards the extreme ends.

The intuition is that since the social pressure is concave, small deviations from the norm draw relatively heavy pressure. Combining this with convex inner preferences – small deviations from the blisspoint are painless – implies that moderates will completely conform to the social norm. In comparison, extremists would feel great dissonance if they were to move close enough to the norm to have an effect on the pressure. However, since the dissonance is convex, the extremists do not mind making small concessions. Hence, they choose an inner solution in the range where it makes little difference what one chooses both from an inner preference and a social pressure point of view. As can be seen, this closely resembles the case where society is orthodox and opinions are rigid. Also here are extremists alienated from society but instead of completely "ignoring" the norm, they comply slightly. This is similar to the result in Bernheim's (1994) paper. What is interesting is that while Bernheim gets this distribution when both pressure and dissonance are convex functions (according to our way of defining them), we get it when

[11] Alternatively, this statement is also true for a sufficiently small K. I.e. we can get qualitatively similar societies by either adding heterogeneity (broadening the range of types) or by decreasing the weight of punishment (decreasing K).

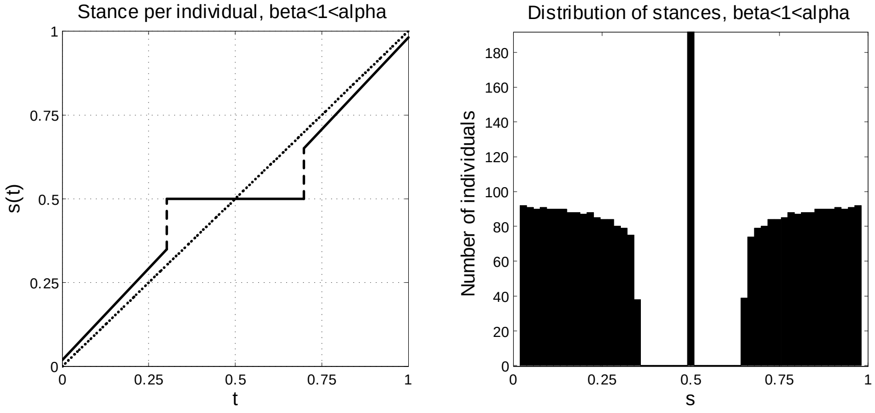

Figure 5: $\beta < 1 \leq \alpha$ with $\bar{s} = .5$, $t \sim U(0,1)$. The left-hand schedule depicts $s^*(t)$ (full line) and, in comparison, the $s = t$ (dashed line). The right-hand schedule depicts the probability distribution function. Note that the y-axis is truncated from above for visual purposes.

dissonance is convex but pressure is concave. This way, whether pressure is applied to actions (our model) or beliefs about types (Bernheim's model) makes an important difference.

It is generally hard to find a closed form solution for the cutoff between conformity and inner solutions. Likewise it is difficult to show that it exists in a specific range. However, the inner solution is increasing relative to the corner solution as t is distanced from the norm. Hence, for a broad enough range of bliss points, we know that the inner solution is preferred for extreme individuals. In contrast, and perhaps trivially, the cutoff is increasing in K in such a way that if the social pressure has enough weight, there can arise a case of everyone choosing the norm.

7 A liberal society with rigid personal opinions

When $\beta > 1$ and $\alpha \leq 1$, we once more get a combination of corner and inner solutions, but largely as a mirror image of the previous case.

Proposition 4 *If $\alpha < 1 \leq \beta$ then:*

1. *$\exists \hat{t} > \bar{s}$ such that $s^*(t) = t$ for every $\bar{s} \leq t < \hat{t}$ and such that $s^*(t) \in]\bar{s}, t[$ for every $t \geq \hat{t}$. This image is mirrored at $\bar{s}$.*

2. *$s^*(t)$ is increasing in the range $\left[\bar{s}, \hat{t}\right]$ and decreasing in the range $\left]\hat{t}, t_h\right]$.*

3. *For a broad enough range of types, the distribution is continuous and bimodal with a uniform section around $\bar{s}$ (consisting of mod-*

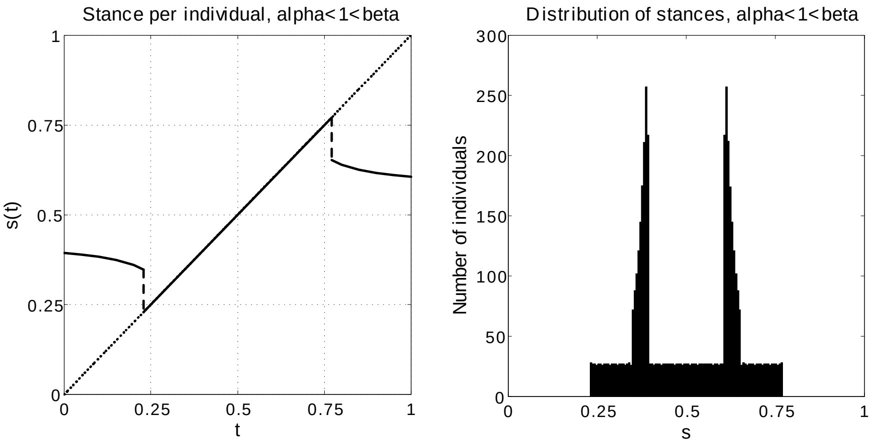

Figure 6: $\alpha < 1 \leq \beta$ with $\bar{s} = .5$, $t \sim U(0,1)$. The left-hand schedule depicts $s^*(t)$ (full line) and, in comparison, the $s = t$ (dashed line). The right-hand schedule depicts the probability distribution function.

erate types) overlapping a peak on each side of $\bar{s}$ (consisting of extreme types) peaking towards $\bar{s}$. For a sufficiently narrow range of types, the distribution is uniform.

4. *Absolute and relative concession are increasing in $|t - \bar{s}|$. Conformity is decreasing in $|t - \bar{s}|$ for t with $|t - \bar{s}| < \left|\hat{t} - \bar{s}\right|$, then sharply increases when $|t - \bar{s}| = \left|\hat{t} - \bar{s}\right|$, and gradually keeps increasing as $|t - \bar{s}|$ grows further.*

Proof. *See the appendix.* ■

Since social pressure is convex, it hardly affects moderates who can now freely choose their inner blisspoint. Extremists, on the other hand, will feel too much pressure by not conforming and since personal opinions are rigid, once they deviate from their blisspoint, they might as well conform a great deal.

As illustrated in figure 6 (left-hand schedule), the proposition implies that the extremists are more conform than some moderates and that, within the group of those conforming, the more extreme individuals are the most conformed. Thus, as in the case of $\alpha < \beta \leq 1$, we get hypocrisy. But now we get it at two levels – both between extremists and moderates and within the group of extremists. All in all, this will create a bimodal distribution (figure 6, right-hand schedule) where extremists form the peaks and there is a uniform distribution of moderates around the peaks. As α increases towards 1, these peaks will move outwards and as α passes 1, the hypocrisy ceases to exist and we are left with the same bimodally distributed liberal/lax society. Likewise, as we decrease β towards 1, the peaks move inwards and as it passes 1 the peaks are centered at $\bar{s}$ and we are left with a unimodal orthodox/rigid society where extremists conform completely. In that way, this case bridges the gap between the societies of $\alpha < \beta < 1$ and $1 < \alpha < \beta$.

8 Relative concession

Having gone through all possible cases, we are now ready to state a general result that spans through all parameter combinations.

Corollary 5 *Iff $\beta < \alpha$, then the relative concession is decreasing in $|t - \bar{s}|$.*
Proof. *Follows from propositions 1-4.* ■

What this proposition establishes is that when the social pressure is more concave (or less convex) than the cognitive dissonance, it mainly affects moderates[12]. This is intuitive since, roughly speaking, when the pressure is relatively concave, then small deviations from the norm matter more than large deviations. Likewise, if the social pressure is relatively more convex, it mainly induces the extremists to conform since it makes a significant difference what one does when being far from the norm. This way, the relative curvature of the social pressure can be said to determine who is affected by it and how the resultant distributions are formed.

9 Endogenizing the social norm

Up until now, nothing has been said about how $\bar{s}$ is determined. The previous analysis can therefore be viewed as a partial equilibrium. In order to establish which social norms are feasible, we will assume that this is determined by the average stance in society. Naturally, there may be other forces shaping the equilibrium position of a social norm, but the average stance seems like a reasonable first case to investigate.

$$\bar{s} = \frac{1}{t_h - t_l} \int_{t_l}^{t_h} s^* (\tau)\, d\tau$$

First, we can note that since the distribution of bliss points is uniform, the distribution of stances (S) is symmetric around $\bar{s}$. Thus, for a certain

[12] Note that this is the only result that does not generalize easily. With general functional forms, the condition $\gamma_P(x) \equiv \frac{xP''(x)}{P'(x)} < \gamma_D(x) \equiv \frac{xD''(x)}{D'(x)}$ is what makes a decreasing relative concession. $\gamma(x)$ is a measure of relative convexity (cf. Brander & Spencer, 1984) similar to the Arrow-Pratt measure of relative risk aversion.

social norm to constitute an equilibrium, S has to be symmetric around $\bar{s}$ over the whole range of types. For distributions of stances where some types have inner solutions, this cannot occur unless $\bar{s} = \frac{t_h+t_l}{2}$, i.e. the average stance is also the average bliss point. For distributions consisting of corner solutions, there may be multiple equilibria. This is expressed in the following proposition which describes all possible equilibria in the different cases.

Proposition 6 *If $\bar{s}$ is the average stance in society:*

1. *If $\beta > 1$ then $\bar{s} = \frac{t_h+t_l}{2}$ is a unique feasible equilibrium.*
2. *If $\beta < \alpha \leq 1$ then $\bar{s}$ is a feasible equilibrium iff $\bar{s} \in \left\{\frac{t_h+t_l}{2}\right\} \cup \left[t_h - K^{\frac{1}{\alpha-\beta}}, t_l + K^{\frac{1}{\alpha-\beta}}\right]$.*
3. *If $\beta \leq 1 < \alpha$ then $\bar{s} \in [t_l, t_h]$ is a feasible equilibrium iff it fulfills $(t-\bar{s})^{\alpha} \leq |t - s^*(t)|^{\alpha} + K|s^*(t) - \bar{s}|^{\beta}$ $\forall t \in [t_l, t_h]$, where $s^*(t)$ is given by $|t - s^*(t)|^{\alpha-1}(|s^*(t) - \bar{s}|)^{1-\beta} = K\beta/\alpha$.*
4. *If $\alpha < \beta \leq 1$ then $\bar{s}$ is a feasible equilibrium iff $\bar{s} \in \left\{\frac{t_h+t_l}{2}\right\} \cup \left[t_l + K^{\frac{1}{\alpha-\beta}}, t_h - K^{\frac{1}{\alpha-\beta}}\right]$ are feasible equilibria.*
5. *If $\beta = \alpha \leq 1$ and $K > 1$ then $\bar{s}$ is a feasible equilibrium iff $\bar{s} \in [t_l, t_h]$. If $K \leq 1$ then $\bar{s} = \frac{t_h+t_l}{2}$ is the unique feasible equilibrium.*

Proof. *See the appendix.* ■

Following the proposition, the only feasible equilibrium in a liberal society is where the social norm is equal to the average bliss point. This is due to the fact that in liberal societies, there is always a portion of individuals (sometimes all) who choose a compromise stance[13]. If the social norm is positioned somewhere else than in the middle, these compromising stances become more influential (a larger total weight) and thus, such a social norm is unsustainable.

In orthodox societies, however, there is a range of equilibria $\bar{s}$. What parts 2 and 3 of the proposition essentially express is that in an orthodox society with relatively more concave social pressure, the social norm must be positioned such that it makes all individuals totally conform as types far from the norm will otherwise be alienated. This is simplified if the weight on social pressure is large. As the weight of pressure falls, this leaves a more narrow range of feasible equilibria and, eventually, the only

[13] In a liberal society with rigid opinions, this reasoning also applies to the moderates who do not concede at all.

remaining equilibrium is when the social norm is equal to the average blisspoint. This implies that in very orthodox societies, the only way of upholding a skewed social norm is by either having severe social pressure or by having individuals with a tight range of bliss points.

The fourth part of the proposition states that in an orthodox society with relatively rigid inner preferences, the social norm has to be positioned such that it covers all moderate types that choose their own inner bliss point as a stance. Such a society, albeit being orthodox, only has to allow the freedom of expression of those close to the norm, since the inner preferences are very rigid. The extreme types do not play any role here since they choose to totally conform, thus giving up their effect in determining the norm.

A pattern emerges from the above proposition showing that liberal societies are bound to eventually have norms representing the average inner opinions in society – this is the only equilibrium. Only orthodox societies can sustain social norms which are not representative of the true opinions of the people. In that way, orthodox societies are history dependent since the initial set of common rules will also determine the long-run equilibrium outcome. This may also explain why orthodox societies (but less often liberal ones) have rules which are not, even on average, in people's interest . Furthermore, it rationalizes why orthodox societies with extremist rules more often resort to harsh punishments than liberal societies – only in the former is it possible to sustain skewed norms with the help of pressure. Therefore, we should observe a correlation between orthodox societies and harsh punishments.

10 Some casual observations

This section will present some examples that we believe the basic model does well in explaining. They are of an informal nature in the sense that we do not attempt to prove causality or that no other mechanisms can explain the observed distributions.

Figure 7 presents how people in Brazil and Sweden, respectively, respond to a question of government versus individual responsibility. The Brazilian distribution looks bimodal with one large group believing that individuals are completely responsible for themselves while another large group believes that the government is responsible for providing for the citizens. In comparison, the Swedish distribution is bell-shaped with the mass centered in between the two extreme stances. How can we explain this difference? Both these societies must be considered liberal in the sense that there is freedom of expression (within boundaries), i.e. $\beta > 1$. On the other hand, they differ substantially in economic inequality where Brazil ranks as one of the most economically unequal

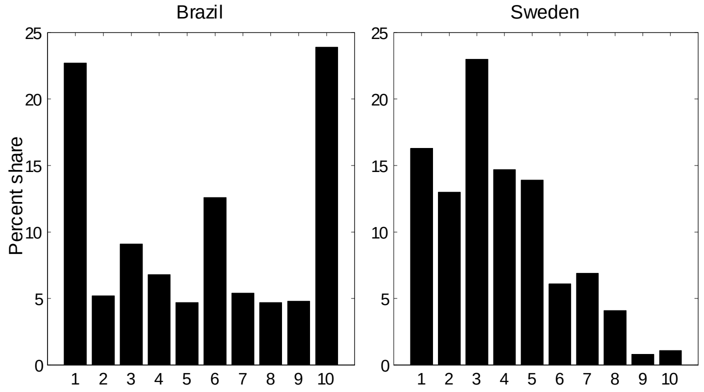

Figure 7: Distribution of respondents' answers to whether people themselves (1) or the government (10) should take more responsibility to ensure that everyone is provided for. Source: World Value Survey third wave.

countries while Sweden ranks as one of the most economically equal countries. The very poor people in Brazil may, as a matter of principle, feel very uncomfortable in stating that they are themselves responsible for their own situation. Likewise, the rich people may be uncomfortable in stating that they have a shared responsibility (by way of the government) for the poor people which may also imply that they are not themselves to praise for their material wealth. Hence, it is conceivable that α in Brazil is low or at least lower than β. In Sweden, on the other hand, since the stakes of being at the economic top or bottom are not as high, both the relatively rich and the relatively poor may well feel that it is fine to state that there is both an individual and a governmental responsibility, as long as they do not have to reverse their opinions. Hence, it is conceivable that α in Sweden is high and possibly higher than β. If this is indeed the case, the model predicts that we should see a bimodal distribution in Brazil and a unimodal distribution in Sweden (see section 4). If the difference between Sweden and Brazil is in the weight of social pressure, i.e. K instead of α, then we should not observe any qualitative difference in the shape of the distribution[14].

An example of an orthodox society is Afghanistan under the Taliban rule. Any deviation, large or small, from the right path was punished severely implying that β was very small and K was large. Consequently, as predicted by the theory, very few deviated from the norm established

[14] These results remain also when looking at related questions from the World Value Survey or when looking at these answers for example for Mexico and India – other countries that rank high in terms of economic inequality. However, this story has not dealt with potential reversed causality.

by the Taliban. In comparison, the ultra-orthodox Jewish society is also intolerant to deviations from the behavior it deems to be right, implying that β is small there. However, its punishments are not as harsh as those in the Taliban society. According to theory, this should lead people with personal opinions close to the established norm to completely conform, while those with personal views far from the orthodox norm will be alienated and essentially ignore the social pressure exerted.

The nature of hypocrisy and reversed stances as described in propositions 2 and 4 implies that it is hard to single out such cases in practice. However, an example may be when non-whites are part of, and sometimes even take on the leadership in, gangs of Nazis and white racists. The explanation for this may be that a non-white person growing up in a neighborhood where there is a group of skinheads will be harassed, implying that K is large. Since skin color or religion cannot easily be changed, even a small conformity towards the racist group will create a significant cognitive dissonance and hence, α must be very small and likely smaller than β. In this situation, a black person may actually choose to join the group, since staying close to one's inner blisspoint implies too harsh a punishment and once the non-white person has denounced his skin color, he might as well minimize the punishment totally. A documented such example is the convicted Swedish police-killer Jackie Arklöv[15], a dark skinned foreign adoptee by Swedish parents, who joined a group of Nazi felons. He was arguably the most violent person within that group. Even though such examples at an individual level only give limited information on the complete distribution and many other circumstances play a role in shaping such an individual, the theory presented here may shed some light on one mechanism influencing people far from the social norm. Moreover, in the example given here, the question remains why the Nazi gang would accept that a black person joins their group.

11 Aggregating individual pressure

This section and the next consider an extension of the basic model. Instead of assuming one social norm, we will now look at a situation where each individual puts pressure on each other individual. We call this pluralistic social pressure. In this section, we start by analyzing the specific issue of how pressure stemming from individuals is aggregated into a social pressure function. The next section then analyzes the actual stances individuals will take and the resultant distribution of stances in a specific case. Namely, when pressure comes from the individual's type,

[15] For a biography, see Sandelin (2010).

each individual pressure is orthodox and dissonance is rigid.

Consider an individual with a revealed stance s who meets people randomly. Each person x that the individual meets will punish her based on the distance from s to x. We will refer to x as the individual source of pressure.

$$\begin{aligned} p(s,x) &= p(|s-x|) \\ p(\cdot)' &> 0 \end{aligned}$$

We can think of x as being either the other individual's type (i.e. her inner bliss point) or her stance (i.e. her revealed stance). For the sake of generality, in analyzing aggregate pressure in this section, we do not explicitly need to determine the source of pressure x.

11.1 Pluralistic and uniform pressure sources

Now, if x is uniformly distributed from x_l to x_h, then an individual with stance s will expect to perceive the following pressure.

$$\begin{aligned} P_{all}(s) \equiv E[p(|s-x|)] &= \frac{1}{x_h - x_l} \int_{x_l}^{x_h} p(|s-x|)\,dx \\ &= \frac{1}{x_h - x_l}\left[P(x_h - s) + P(s - x_l) - 2P(0)\right],\ s \in [x_l, x_h] \end{aligned}$$

where, by convention, $P' \equiv p$.[16] This is the pressure that the individual with stance s can expect to feel when meeting people randomly (which, of course, is equivalent to the normalized aggregate pressure perceived when meeting all other individuals simultaneously). What are the properties of this aggregated pressure function? By differentiating P_{all}, we get

$$\begin{aligned} P'_{all}(s) &= p(s - x_l) - p(x_h - s) \\ P''_{all}(s) &= p'(s - x_l) + p'(x_h - s)\,. \end{aligned}$$

Now, define $\bar{s} \equiv \frac{x_l + x_h}{2}$. The following proposition then follows.

Lemma 3 *If x is uniformly distributed, then the aggregated pressure function $P_{all}(s)$:*

1. *Is strictly increasing if $s > \bar{s}$ and strictly decreasing if $s < \bar{s}$.*

2. *Is strictly convex if $s \neq \bar{s}$ and $s \in]x_l, x_h[$.*

[16] We assume that p is integrable.

3. Has a zero derivative at $\bar{s}$.

4. Is symmetric around $\bar{s}$.

Proof. *1), 2) and 3) follow trivially from the first and second derivatives of P_{all}, from $p(\cdot)' > 0$ and from inserting $\bar{s}$ into the first derivative. 4) To see the symmetry, let $\tilde{s}$ be the mirror image of s, i.e. $(s+\tilde{s})/2 = \bar{s}$, hence $\tilde{s} = 2\bar{s} - s = x_h + x_l - s$. Then we get $P_{all}(\tilde{s}) = P(\tilde{s} - x_l) + P(x_h - \tilde{s}) = P(x_h + x_l - s - x_l) + P(x_h - (x_h + x_l - s))$ $= P(x_h - s) + P(s - x_l) = P_{all}(s)$.* ■

That social pressure is increasing with the distance to the average x is perhaps not very surprising – the more extreme is one's stance, the more pressure will one feel. But that the aggregated pressure function is convex may be less obvious, given that we have not even specified whether the pressure stemming from each person is convex or concave. This means that under a uniform distribution of sources of pressure, social pressure will be convex even if the one-on-one pressure is concave. So a society, made up of "orthodox" individuals with uniform tastes, will, in fact, be "liberal" on the aggregate.

Moreover, $P_{all}(s)$ has a unique minimum point at $\bar{s} \equiv \frac{x_l + x_h}{2}$ around which it is symmetric. This suggests that qualitatively, the aggregation of punishment from a uniform distribution of sources of pressure is similar to having a "virtual" social norm at $\frac{x_l+x_h}{2}$, where the pressure is increasing with the absolute size of the deviation from this norm.

11.2 Combining pluralism with an authority

From the analysis of the basic model, we know that the previous result does not extend to the case where there is one source of pressure, i.e. an authority. The question is then under what distribution of sources of pressure it does extend. It turns out that this issue is hard to analyze in a general way. So, we will instead analyze a few specific cases.

Given the discrepancy between a uniform distribution of pressure sources and a single source, we will now look at a combination of the two. This would represent a society with two sources of pressure. First, an institutionalized norm which is the average of all opinions and, second, the aggregate pressure of individuals who pressure each other. We will look at the case where both individual and institutional pressures have the same functional form, p, which is concave. The total pressure function is then a weighted average of the two.

$$P_{combi} = P(x_h - s) + P(s - x_l) - 2P(0) + Ap\left(\left|s - \frac{x_h + x_l}{2}\right|\right)$$

where A is the relative weight of the institutional pressure. We here only analyze the case of $s \geq \frac{x_h - x_l}{2}$. By symmetry, the other case has the same properties. Differentiating, we get

$$P'_{combi} = p\left(s - x_l\right) - p\left(x_h - s\right) + Ap'\left(s - \frac{x_h + x_l}{2}\right) \tag{9}$$

$$P''_{combi} = p'\left(s - x_l\right) + p'\left(x_h - s\right) + Ap''\left(s - \frac{x_h + x_l}{2}\right). \tag{10}$$

The first two elements in each expression represent the individual pressure, while the third entity represents the authoritarian pressure. From equation (9), it is clear that at the point $s = \frac{x_h + x_l}{2}$, $P'_{combi} = Ap'\left(s - \frac{x_h + x_l}{2}\right)$. The marginal pressure is completely determined by the authority at this point. But whether P_{combi} is concave depends on the sign of equation (10). Assuming that $\lim_{y \to 0} p'(y) = \infty$ and $\lim_{y \to 0} p''(y) = -\infty$, and letting s approach $\frac{x_h + x_l}{2}$, it is clear that P''_{combi} is negative and thus concave around the virtual norm. In a similar fashion, it can be shown that as s approaches either of the extreme stances, x_h or x_l, P''_{combi} is positive and thus convex.

This means that P''_{combi} should change signs an uneven number of times in the interval $s \in \left[\frac{x_h + x_l}{2}, x_h\right]$. Now, with the assumption that $\lim_{y \to 0} p''(y) = -\infty$, it is necessary that $\lim_{y \to 0} p'''(y) > 0$. For simplicity, we will then assume that $p'''(y) > 0$ for all y.[17] This implies that $P''_{combi}(s)$ is monotonically increasing in the interval $s \in \left[\frac{x_h + x_l}{2}, x_h\right]$ and therefore that P''_{combi} changes signs exactly once in the interval $s \in \left[\frac{x_h + x_l}{2}, x_h\right]$.[18]

The interpretation of these results is that a society with mixed authoritarian and individual pressure – both being orthodox – will, on aggregate, tend to be orthodox towards stances around the norm, but liberal towards stances far from it.

11.3 Exponential and Gaussian distribution of pressure sources

Now, an interesting complement to this analysis is looking at other distributions of pressure sources but continuing with individual pressure being concave. It turns out that it is hard to say anything in general about this, so we will analyze two specific types of distributions, the Exponential and the Gaussian, with individual pressure being a power

[17] One set of functions that have all these assumed properties are power functions with an exponent less than 1.

[18] To see this, note that $P'''_{combi} = p''\left(s - x_l\right) - p''\left(x_h - s\right) + Ap'''\left(s - \frac{x_h + x_l}{2}\right)$. The last part is positive by $p''' > 0$. This also implies that $p''\left(s - x_l\right) > p''\left(x_h - s\right)$ since $s - x_l > x_h - s$ when $s \geq \frac{x_h + x_l}{2}$. So the total expression is positive.

function. These two distributions both have a clear peak and sharply declining tails. We are interested in seeing whether they can produce orthodox aggregate pressure. The derivations of the upcoming results can be found in the appendix in section 14.4.

Posit a distribution of pressure sources $f(x)$ which symmetrically has an exponential shape peaking towards some point from each side. Assume further that this point is at $s = 0$, i.e. $E(x) = 0$, and that the minimum and maximum pressure source in society are at $\pm\infty$.

$$\begin{aligned} P_{\exp}(s) &= \int_{t_l}^{t_h} p(|x-s|) f(x)\, dx \\ &= \frac{\lambda}{2} \int_{-\infty}^{\infty} |x-s|^{\alpha} e^{-\lambda|x|} dx = \frac{\lambda}{2} \int_{-\infty}^{\infty} |x|^{\alpha} e^{-\lambda|x+s|} dx \end{aligned}$$

where $0 < \alpha < 1$. It can be shown that $P''_{\exp}(0) > 0$ and that $\lim_{s\to\infty} P''_{\exp} \approx \frac{1}{2}\alpha(\alpha-1)s^{\alpha-2}$ which converges to 0 from below. This means that total pressure is convex near the norm and concave for extreme stances (at least when the extreme stances are sufficiently extreme for the limit case to be relevant). Unfortunately, it is hard to say anything about when and how many times it switches from convex to concave.

Let us now in a similar fashion analyze the case where the pressure sources follow a Gaussian distribution, so that $f(x) = \sqrt{\frac{\lambda}{\pi}} e^{-\lambda x^2}$, $t_l = -\infty$, $t_h = \infty$. As before, we continue with a concave power function for the pressure.

$$\begin{aligned} P_{gauss}(s) &= \int_{t_l}^{t_h} p(|x-s|) f(x)\, dx \\ &= \sqrt{\frac{\lambda}{\pi}} \int_{-\infty}^{\infty} |x-s|^{\alpha} e^{-\lambda x^2} dx = \sqrt{\frac{\lambda}{\pi}} \int_{-\infty}^{\infty} |x|^{\alpha} e^{-\lambda(x+s)^2} dx \end{aligned}$$

It can be shown that $P''_{\exp}(0) > 0$ and that $\lim_{s\to\infty} P''_{\exp} \approx \frac{1}{2}\alpha(\alpha-1)s^{\alpha-2}$ which converges to 0 from below. So the total pressure is convex around the norm and concave towards the extremes. In this case, it is once more hard to say anything about where and how many times the shift between convex and concave forms takes place.

Under Gaussian and Exponential distributions, it seems that the switch of pressure from liberal to orthodox towards the extremes is dependent on the pressure sources virtually vanishing. Then, from the point of view of someone taking an extreme stance, the perception is that there is just a mass of punishing individuals located at the norm. What truly is a distribution of pressure sources then looks like one authority for someone standing sufficiently far away.

11.4 Interpretation

The two previous examples illustrate societies whose sources of pressure are concentrated around the social norm, but with a slowly vanishing tail of extreme sources of pressure. An example of this could be a situation where the pressure sources represent the true opinions of people and these opinions are very but not completely concentrated. Here, although there is a clear peak around the norm and each individual is orthodox, stances close to the norm will be pressured in a liberal manner, while stances far from the norm will be pressured in an orthodox way. This is in contrast to the case with a combination of authoritarian and uniformly distributed individual pressure, where slight deviations from the norm are punished in an orthodox way, while extreme stances will feel a liberal pressure.

Although it is hard to make any general statements about this, it seems that we need a *single* authority for the pressure close to the norm to be orthodox. Otherwise, the accumulation of individual pressure gives a liberal aggregate. This suggests that upholding an orthodox society is not possible in a pluralistic society where heterogenous individuals pressure each other. Moreover, it predicts that orthodox societies will be authoritarian.

12 True opinions as a source of pressure in orthodox societies

In this section, we assume that the entire pressure comes from individuals in society (i.e. no authority), and that the sources of pressure are the individuals' types, i.e. $x = t$. We already analyzed the shape of aggregate social pressure in section 11, but now we further analyze what stances people will actually take given this social pressure. We will concentrate on a society with rigid and orthodox individuals as these results are the most interesting. Nested in this is the specific case where individuals punish each other in the same way that they punish the self for not telling the truth, i.e. $p = D$.

If types are uniformly distributed from t_l to t_h, then an individual with stance s will expect to perceive the following pressure,

$$\begin{aligned}P_{all}(s) \equiv E[p(|s-t|)] &= \frac{K}{t_h - t_l}\int\limits_{t_l}^{t_h} p(|s-t|)\,dt \\ &= \frac{K}{t_h - t_l}\left[P(t_h - s) + P(s - t_l) - 2P(0)\right],\ s \in [t_l, t_h],\end{aligned}$$

where $P' \equiv p$ and K is the weight of punishment from one individual. The optimization problem of the single individual of type t is then

$$\min_s L = P_{all}(s) + D(t-s).$$

The results of combining a concave p and D are summarized in the following proposition where $\bar{s} \equiv \frac{t_h + t_l}{2}$.[19]

Proposition 7 *In a pluralistic society, if true opinions are uniformly distributed and are the source of pressure and both p and D are concave then:*

1. *The aggregate pressure, $P_{all}(s)$, has a unique minimum point, a "virtual" norm, at $s = \bar{s}$.*
2. *$P_{all}(s)$ is convex with the distance to the virtual norm so that society is liberal.*
3. *$\exists \hat{t} > \bar{s}$ such that $s^*(t) = t$ for every $\bar{s} \leq t < \hat{t}$ and such that $s^*(t) \in]\bar{s}, t[$ for every $t > \hat{t}$. This image is mirrored at $\bar{s}$.*
4. *$s^*(t)$ is increasing in the range $\left[\bar{s}, \hat{t}\right]$ and is decreasing in the range $\left]\hat{t}, t_h\right]$. The image is mirrored at $\bar{s}$.*
5. *The distribution of stances is continuous and bimodal with a uniform section around $\bar{s}$ (made of moderate types) overlapping a peak on each side of $\bar{s}$ (made of extreme types).*
6. *Conformity is non-monotonic – it is increasing for moderates and decreasing for extremists.*

[19] Sufficient assumptions for the upcoming proposition is that $\lim_{x\to 0} D'(x) = \infty$. This is fulfilled by power functions. We will also assume here that the functions are such that if an inner local minimum point exists for a type then it is unique. A sufficient condition for this is that $L'''(s,t) < 0$. This holds for e.g. p and D being power functions. If the inner min points are not unique, we may get several peaks on each side of $\bar{s}$ in the distribution of stances.

7. *The absolute concession is weakly increasing in the distance from $\bar{s}$.*

Proof. *See the appendix section 14.2.2.* ■

The proposition expresses that in pluralistic and orthodox societies, a single "virtual" norm will be established and although each individual pressure is orthodox, society as a whole will be liberal. Furthermore, moderate individuals will tend to speak their mind truthfully while hypocrisy will arise among those who are extreme enough. This way, the virtual norm becomes like an unspoken consensus in society. If you are close enough to the consensus, you do what you want but if you are far from the consensus you make concessions to seem to be a moderate. The more extreme you are the more you will concede.

The connection found here between hypocrisy and orthodox societies does not arise by chance. If we interpret rigidness as orthodoxy at an individual level then already by equation 5 we will see that when there is compromise, orthodox individuals will behave in a hypocritical way. This seems to be true also for corner solutions as suggested by the case where personal opinions are more rigid than society is orthodox (see Proposition 2 part 2). Anecdotal support is not lacking here. There are plenty of stories where politicians from conservative parties turn out to be homosexual after having made a career of strongly opposing it. Likewise, there have been sex scandals among priests and revelations of doping among athletes who have earlier denounced it vehemently. These loose observations do, of course, have a rather complex background so it is hard to tell the direction of causality. But the model and the results shown here may provide one explanation for why such cases are observed.

13 Concluding remarks

This paper has presented a simple theory on how social pressure affects the distribution of stated opinions and visible actions in a society or a group. The core message is that even if the individual faces a fairly trivial problem, the results at an aggregate level are non-trivial, diverse and to an extent even surprising. For example, the paper shows that a liberal society will display compromise and either a unimodal or a bimodal distribution depending on how lax the personal opinions are relative to social pressure. In comparison, in orthodox societies, people will tend to either completely conform to the social norm or totally ignore it. If personal opinions are rigid relative to social pressure, people with inner opinions far from the norm will choose to comply, while those with inner opinions close to the norm will declare their inner opinions openly.

If personal opinions are lax relative to social pressure in an orthodox society, those close to the norm will conform completely while those far from it will essentially ignore social pressure.

Admittedly, we have only provided anecdotal evidence for the validity of the model. Like any other theory, it needs to go through the scrutiny of empirical testing. Since it distinguishes between hidden and revealed preferences and since most issues where one would expect social pressure to be at force are very complex, we believe that possibly the best way of testing the model is by experiments. The natural starting point would perhaps be to take the curvature of personal opinions as an unknown and exogenous parameter and vary the curvature and weight of social pressure through the experiment. Alternatively, by taking an issue where one is certain of the curvature of inner preferences, one can test the model predictions by varying the weight and curvature of punishment.

Since the true distribution of opinions is probably hard to observe in reality, our clearest empirical prediction relates to how stances change as a function of true opinions. An experimental approach could then be to first solicit the true opinions (types) of individuals in a setting without social pressure, for example where individuals do not observe each other. The second step would then be to redo the experiment with social pressure, for example by individuals observing each other, and see how the stances change for each type. The final step would be to fit a function to the mapping from type to stance. Using our theoretical predictions about how stances change as a function of type, one could then possibly back out the curvature of the individual dissonance in relation to the curvature of the social pressure function.

Another model prediction is that only orthodox societies can sustain a skewed social norm over time, whereas the social norm in liberal societies is bound to better represent the true distribution of inner preferences. This may also be connected to the loose observation that a liberal atmosphere is often connected with democracy, while orthodox societies are more often authoritarian in order to uphold a skewed social norm. We saw that this discrepancy also translates into pluralistic societies, where it will be hard to maintain orthodox pressure unless an authority is part of sanctioning wrongful behavior.

Extensions and applications to the model can be made along several dimensions. What comes first to our mind is to extend the pluralistic pressure model into a case where the sources of pressure are the individual stances. At the face of it, this makes for a significantly more complicated model since the pressure function changes as a function of itself.

References

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

[2] Brander, J., Spencer, B., (1984) “Tariff Protection and imperfect competition”, in *Monopolistic Competition and Product Differentiation and International Trade*, Henryk Kierzkowski ed., Oxford Economic Press, New York, 194-206.

[3] Brock, W.A., Durlauf, S.N., (2001), “Discrete Choice with Social Interactions”, *Review of Economic Studies* Vol. 68, pp. 235–260.

[4] Granovetter, M., (1976), “Threshold Models of Collective Behavior”, *The American Journal of Sociology,* Vol. 83, No. 6, pp. 1420-1443.

[5] Lindbeck, A., Nyberg, S. and Weibull, J. W. (2003), “Social norms and Welfare State Dynamics”, *Journal of the European Economic Association*, Vol 1, Iss 2-3, pp. 533–542.

[6] López-Pintado, D., Watts, D.J., (2008), "Social Influence, Binary Decisions and Collective Dynamics", *Rationality and Society*, Vol. 20, no. 4, pp. 399-443.

[7] Manski, C.F., (1993), “Identification of Endogenous Social Effects: The Reflection Problem”, *The Review of Economic Studies,* Vol. 60, No. 3, pp. 531-542

[8] Sandelin, M (2010), *Den svarte nazisten* Forum förlag, Stockholm, Sweden.

[9] Schelling, T., (1971), “Dynamic Models of Segregation”, *Journal of Mathematical Sociology*, Vol. 1, Iss. 2, pp.143–186.

14 Appendix - Proofs and derivations

14.1 Transformation from function to density

We now analyze the PDF of chosen stances in society. We restrict ourselves to cases where the optimal stance of each type is uniquely determined, i.e. whenever t has multiple solutions he chooses a single one of them[20]. We divide the range of types into $n+1$ subranges

$$T_0 = [t_{low}, t_1], T_1 = [t_1, t_2], ...T_n = [t_n, t_{high}]$$

such that:

1. In each subrange, the function $s^*(t)$ only consists of corner solutions or only inner solutions.

2. In case of inner solutions, $s^*(t)$ is continuous and strictly monotonic in a subrange[21].

[20] Otherwise we have no way of determining the chosen stance of some types.

[21] This means that D' cannot contain non-monotonic parts within the subrange.

We now investigate separately the contribution of each such subrange to the resultant PDF. The contribution of each such part is called a *partial* PDF, to be denoted $pPDF_{T_i}$, which fulfills

$$PDF = \sum_i pPDF_{T_i}.$$

14.1.1 Inner solutions

Now we investigate the properties of the $pPDF_{T_i}$ (shortly $pPDF$) in subranges with inner solutions. Denote by $s^*_{\min}$ the lowest stance taken by a type in the subrange (strict monotonicity ensures that this type is unique). Let $M_i(\tilde{s}^*)$ be the mass of types in T_i with stances in the range $(s^*_{\min}, \tilde{s}^*]$ for some $\tilde{s}^*$.

$$M_i(\tilde{s}^*) \equiv \int_{s^*_{\min}}^{\tilde{s}^*} pPDF|_s ds = \begin{cases} \int_{t_i}^{t(\tilde{s}^*)} f(\tau)\, d\tau \text{ if } s^*(t) \text{ is rising in the subrange } T_i \\ \int_{t(\tilde{s}^*)}^{t_{i+1}} f(\tau)\, d\tau \text{ if } s^*(t) \text{ is falling in the subrange } T_i \end{cases}$$

where $t(\tilde{s}^*) \equiv \{t \text{ s.t. } s^*(t) = \tilde{s}^*\}$.

Remembering that the distribution of types is uniform we get:

$$M_i(\tilde{s}^*) = \begin{cases} \frac{t(\tilde{s})-t(s^*(t_i))}{t_h - t_l} \text{ if } s^*(t) \text{ is rising in the subrange } T_i \\ \frac{t(s^*(t_{i+1}))-t(\tilde{s}^*)}{t_h-t_l} \text{ if } s^*(t) \text{ is falling in the subrange } T_i \end{cases} \tag{11}$$

$$pPDF|_{\tilde{s}^*} = \frac{dM_i(\tilde{s}^*)}{d\tilde{s}^*} = \frac{1}{t_h - t_l}\left|\frac{dt}{ds^*}|_{\tilde{s}^*}\right| \tag{12}$$

Note that the last derivation is valid only if $\frac{ds^*}{dt}|_{\tilde{s}^*} \neq 0$ as otherwise $\frac{dt}{ds^*}$ is not defined. This is ensured under the strict monotonicity of $s^*(t)$. We then have:

$$\frac{d\left(pPDF\left(\tilde{s}^*\right)\right)}{ds^*} = \begin{cases} \frac{1}{t_h-t_l}\frac{d^2t}{ds^{*2}}|_{\tilde{s}^*} \text{ if } \frac{dt}{ds^*}|_{\tilde{s}^*} > 0 \\ -\frac{1}{t_h-t_l}\frac{d^2t}{ds^{*2}}|_{\tilde{s}^*}\text{if } \frac{dt}{ds^*}|_{\tilde{s}^*} < 0 \end{cases}. \tag{13}$$

Proof of lemma 2

From equation 13, it follows that the $pPDF$ is increasing if $\frac{dt}{ds^*}$ and $\frac{d^2t}{ds^{*2}}$ have the same sign and decreasing if $\frac{dt}{ds^*}$ and $\frac{d^2t}{ds^{*2}}$ have opposite signs. We then use the fact that $\frac{d^2s^*}{dt^2} < 0$ if $\frac{dt}{ds^*}$ and $\frac{d^2t}{ds^{*2}}$ have the same sign, and $\frac{d^2s^*}{dt^2} > 0$ if $\frac{dt}{ds^*}$ and $\frac{d^2t}{ds^{*2}}$ have opposite signs.■

14.1.2 Corner solutions.

There are two candidate corner solutions. The first is $s(t) = t$, i.e. when type t chooses t as a stance, and then $\arg\min L(s) = \arg\min D(t-s)$. In a subrange of these corner solutions, the $pPDF$ is simply a uniform distribution with the trivial properties

$$pPDF|_{\tilde{s}^*} = \frac{1}{t_h - t_l}\frac{dt}{ds^*}|_{\tilde{s}^*} = \begin{cases} \frac{1}{t_h - t_l} \text{ if } \tilde{s}^*(t) = t \\ 0 \text{ otherwise} \end{cases}$$
$$\frac{d(pPDF)}{ds} = 0.$$

The other candidate corner solution corresponds to the solution of $\arg\min L(s) = \arg\min P(s)$. The solution of this equation is independent of t, so in a subrange of these corner solutions, the $pPDF$ is a degenerate single peak with a mass equalling the mass of types within that subrange.

$$pPDF|_{s^*} = \begin{cases} \frac{t_{i+1} - t_i}{t_h - t_l} \text{if } s^* = \arg\min P(s) \\ 0 \text{ otherwise} \end{cases}$$

Finally, in a subrange of mixed corner solutions, the $pPDF$ is constructed of both types of corner solutions.

14.2 General functions

The purpose of this section is to lay the ground for the upcoming proofs where one of P and D is concave and the other is convex. We will analyze the case where P has exactly one min point $\bar{s} \equiv \arg\min_s P(s)$. This naturally covers the case with an exogenous social norm but also the case where there is one endogenous "virtual" norm, like in the case of P arising from the aggregation of individuals pressuring each other. Sufficient conditions for the upcoming results to hold are:

1. The signs of both P'' and D'' are constant, i.e. each of the functions P and D is either linear or strictly convex or strictly concave at the whole range.

2. Furthermore, to avoid dealing with the special behavior of types near $\bar{s}$, we restrict our attention to functions P such that $P'(\bar{s}) = 0$ if $P(\cdot)$ is convex and $\lim_{x \to +\bar{s}} P'(x) = \infty$ if $P(\cdot)$ is concave. Equivalent requirements apply to D.

3. $\lim_{x \to \infty} P'(x) = \infty$ if $P(\cdot)$ is convex. Likewise for D.

We will perform the analysis for $t \geq \bar{s}$ (the analysis for $t < \bar{s}$ is equivalent).

14.2.1 Concave $P(s)$, Convex $D(t-s)$

When P is concave and D is convex we have that in the corners $L'(s(t)=\bar{s}) = P'(0) - D'(t-\bar{s}) \to \infty$ and $L'(s(t)=t) = P'(t-\bar{s}) - D'(0) = P'(t-\bar{s}) > 0$. This implies that potential corner solutions must be at $s=\bar{s}$. It also implies that we have either zero or an even number of inner extreme points.

We will now show that inner extreme points exist for a sufficiently broad range of types. For the following, $\bar{s}$ is assumed to be weakly positive. For sufficiently negative $\bar{s}$, some parameters need to be redefined but the result will hold nevertheless. In inner extreme points, the FOC needs to be zero. Since $L'(s=t) = P'(t-\bar{s}) > 0$, it is then sufficient to show that $L' < 0$ for some t and $s \in [\bar{s}, t]$.

Define $\dot{t}$ implicitly by $D'(\dot{t}-\bar{s}) = P'(\dot{t}-\bar{s})$. I.e. $\dot{t}$ is the type whose minimal marginal pressure (when choosing $s=\dot{t}$) is exactly equal to the maximal marginal dissonance (when choosing $s=\bar{s}$). We know by the previous results that $\dot{t} > \bar{s}$. We also know that this type exists in a broad enough but finite range since $\lim_{t\to\infty} P'(t-\bar{s}) = 0$ and $\lim_{t\to\infty} D'(t-\bar{s}) = \infty$. However, $\dot{t}$ will not have an inner extremum since $P'|_{s=\dot{t}} = D'|_{s=\bar{s}}$ is the only way to equal D' and P' and since $\dot{t} \neq \bar{s}$.

Let us now look at the type $\ddot{t} = \bar{s} + 2\left(\dot{t}-\bar{s}\right) + \varepsilon$ where $\varepsilon \geq 0$. This is the type which is just beyond twice as far from the norm as $\dot{t}$. If $\ddot{t}$ chooses $s=\dot{t}$, we have

$$\begin{aligned} L'\left(\bar{s} + 2\left(\dot{t}-\bar{s}\right) + \varepsilon, \dot{t}\right) &= P'(\dot{t}-\bar{s}) - D'(\bar{s} + 2\left(\dot{t}-\bar{s}\right) + \varepsilon - \dot{t}) = \\ P'(\dot{t}-\bar{s}) - D'(\dot{t}-\bar{s}+\varepsilon) &= D'(\dot{t}-\bar{s}) - D'(\dot{t}-\bar{s}+\varepsilon). \end{aligned}$$

Since D is convex $\varepsilon > 0$ gives a strictly negative L' which proves the existence of inner extreme points for broad enough but finite ranges.

We will now show that for a broad enough range of types, some will have an inner global min point. Suppose that an inner local min point exists. Let us now compare the losses of the inner and the corner min points. In the corner $s=\bar{s}$ and the inner min point we denote by $\hat{s}$.

$$\begin{aligned} Diff &\equiv L\left(s\left(t\right)=\bar{s}\right) - L\left(s\left(t\right)=\hat{s}\right) \\ &= D\left(t-\bar{s}\right) - \left[P\left(\hat{s}-\bar{s}\right) + D\left(t-\hat{s}\right)\right] \end{aligned}$$

Thus, for $t=\bar{s}$ the corner solution is preferred since then $Diff < 0$.

$$\begin{aligned} \frac{dDiff}{dt} &= D'\left(t-\bar{s}\right) - P'\left(\hat{s}-\bar{s}\right)\frac{d\hat{s}}{dt} - D'\left(t-\hat{s}\right)\left(1-\frac{d\hat{s}}{dt}\right) \\ &= \{\text{use } D'=P'\} = D'\left(t-\bar{s}\right) - D'\left(t-\hat{s}\right) > 0 \text{ with convex } D. \end{aligned}$$

This implies that there is one cutoff at most where s^* changes from being a corner to an inner solution. So, for t sufficiently close to $\bar{s}$, $s^*(t) = \bar{s}$. To see whether the cutoff between corner and inner solutions exists, note that

$$\frac{d^2 Diff}{dt^2} = D''(t - \bar{s}) - D''(t - \hat{s}) \left(1 - \frac{d\hat{s}}{dt}\right) > 0.$$

We know that the last parenthesis is negative by inserting a concave P and a convex D in equation 5, implying that $d\hat{s}/dt > 1$. Thus, as $Diff$ is increasing and convex, we know that types sufficiently far from $\bar{s}$ have an inner solution. Let $\hat{t}$ be the cutoff type. Showing that this $\hat{t}$ exists in a certain range is hard. For a sufficiently narrow range of types, the result is a distribution of stances that is Dirac Delta at $\bar{s}$. However, if there is a sufficiently broad range of types on each side of $\bar{s}$, then the resultant distribution contains a peak at $\bar{s}$, with tails at each side of it (and a gap between each tail and the peak at $\bar{s}$).

Using $ds^*/dt > 1$ in Lemma 1, we get that conformity and absolute concession are decreasing with t in inner solutions. In corner solutions, conformity is constant while absolute concession is increasing since $ds^*/dt = 0$.

14.2.2 Convex $P(s)$, Concave $D(|t - s|)$

Proof of proposition 7

1) and 2) follow directly from Lemma 3. Since 2) implies that $P_{all}(s - \bar{s})$ is convex, we will analyze the general case of a convex P and a concave D. Note also that $\lim_{s \to \bar{s}} P'_{all}(s - \bar{s}) = 0$ and that $\lim_{t_h \to \infty, s \to t_h} P'_{all}(s - \bar{s}) = \infty$.

3): When D is concave and P is convex, we have that in the corners $L'(s = t) = P'(t - \bar{s}) - \lim_{x \to 0} D'(x) = -\infty$ while $L'(s = \bar{s}) = P'(0) - D'(t - \bar{s}) = -D'(t - \bar{s}) < 0$. This implies that potential corner solutions must be at $s = t$. It also implies that we either have zero or an even number of inner extreme points, e.g. if there are two extreme points, one is a min and the other is a max.

We will now show that inner extreme points exist for a sufficiently broad range of types. In the following, $\bar{s}$ is assumed to be weakly positive. For sufficiently negative $\bar{s}$, some parameters need to be redefined but the result will hold nevertheless. In inner extreme points, the FOC needs to be zero. Since $L'(s = \bar{s}) = -D'(t - \bar{s}) < 0$, it is then sufficient to show that $L' > 0$ for some t and $s \in [\bar{s}, t]$.

Define implicitly $\dot{t}$ by $D'(\dot{t} - \bar{s}) = P'(\dot{t} - \bar{s})$. I.e. $\dot{t}$ is the type whose maximal marginal pressure (when choosing $s = \dot{t}$) is exactly equal to the minimal marginal dissonance (when choosing $s = \bar{s}$). We know by the previous results that $\dot{t} > \bar{s}$. We also know this type exists in a broad enough but finite range since $\lim_{t\to\infty} D'\left(t-\bar{s}\right) < \infty$ when D is concave and $\lim_{t\to\infty} P'\left(t-\bar{s}\right) = \infty$. However, $\dot{t}$ will not have an inner extremum since $P'|_{s=\dot{t}} = D'|_{s=\bar{s}}$ is the only way to equal D' and P' and since $\dot{t} \neq \bar{s}$.

Let us now look at the type $\ddot{t} = \bar{s} + 2\left(\dot{t}-\bar{s}\right) + \varepsilon$ where $\varepsilon \geq 0$. This is the type which is just beyond twice as far from the norm as $\dot{t}$. If $\ddot{t}$ chooses $s = \dot{t}$, we have

$$L'\left(\bar{s}+2\left(\dot{t}-\bar{s}\right)+\varepsilon,\dot{t}\right) = P'(\dot{t}-\bar{s}) - D'(\bar{s}+2\left(\dot{t}-\bar{s}\right)+\varepsilon-\dot{t}) =$$
$$P'(\dot{t}-\bar{s}) - D'(\dot{t}-\bar{s}+\varepsilon) = D'(\dot{t}-\bar{s}) - D'(\dot{t}-\bar{s}+\varepsilon).$$

Since D is concave $\varepsilon > 0$ gives a strictly positive L' which proves the existence of inner extreme points for a broad enough but finite range.

We will now show that for a broad enough range, the local min will be the global min. Suppose now that an inner local min point exists. Let us now compare the losses of the inner and the corner solutions. In the corner $s = t$ and the inner extreme point, we denote by $\hat{s}$.

$$\begin{aligned} Diff &\equiv L\left(s\left(t\right)=t\right) - L\left(s\left(t\right)=\hat{s}\right) \\ &= P\left(t-\bar{s}\right) - \left[P\left(\hat{s}-\bar{s}\right) + D\left(t-\hat{s}\right)\right] \end{aligned}$$

Thus, for $t = \bar{s}$ the corner solution is preferred since then $Diff < 0$.

$$\begin{aligned} \frac{dDiff}{dt} &= P'\left(t-\bar{s}\right) - P'\left(\hat{s}-\bar{s}\right)\frac{d\hat{s}}{dt} - D'\left(t-\hat{s}\right)\left(1-\frac{d\hat{s}}{dt}\right) \\ &= \{\text{using } D' = P'\} = P'\left(t-\bar{s}\right) - P'\left(\hat{s}-\bar{s}\right) > 0 \text{ with a convex } P \end{aligned}$$

This implies that there is at most one cutoff where s^* changes from being a corner to an inner solution. So, for t sufficiently close to $\bar{s}$, $s^*\left(t\right) = t$. To see whether the cutoff between the corner and inner solutions exists, note that

$$\frac{d^2Diff}{dt^2} = P''\left(t-\bar{s}\right) - P''\left(\hat{s}-\bar{s}\right)\frac{d\hat{s}}{dt} > 0.$$

We know that $d\hat{s}/dt$ is negative in inner solutions by inserting a convex P and a concave D in equation 5 (this proves 4). Thus, as $Diff$ is increasing and convex we know that types sufficiently far from $\bar{s}$ have an inner solution which concludes 3).

5): Let $\hat{t}$ be the cutoff type. Showing that the inner min points are unique is hard. Assuming this to be the case, if there is a sufficiently broad range of types on both sides of $\bar{s}$, the resultant distribution contains a uniform part at the range surrounding $\bar{s}$ and, on top of this, two formations, one on each side of $\bar{s}$ (these "formations" must be on top of the uniform part because the cutoff type $\hat{t}$ chooses an inner solution $s^*(\hat{t}) < \hat{t}$, and every $t > \hat{t}$ chooses $\bar{s} < s^*(t) < s^*(\hat{t})$ because $ds^*/dt < 0$).

6) and 7): Using $ds^*/dt < 0$ in Lemma 1, we get that both conformity and absolute concession are increasing with t in inner solutions. In corner solutions, conformity is decreasing while absolute concession is constant since $ds^*/dt = 1$.■

14.3 Power functions and punishment from one norm

14.3.1 Case: $\beta < 1 < \alpha$

Proof of proposition 3

In this case, we have a concave $P(s)$ with a unique min point $\bar{s}$, and a convex $D(t-s)$. We also have $P'(\bar{s}) \to \infty$. As we saw in the general analysis (section 14.2.1), types near $\bar{s}$ choose $s = \bar{s}$. We can also check and see that $\forall t > \bar{s}$ there is at most one candidate inner solution (i.e. one local minimum). The FOC gives

$$\alpha (t-s)^{\alpha-1} = \beta K (s-\bar{s})^{\beta-1} \Rightarrow \beta K/\alpha = (t-s)^{\alpha-1} (s-\bar{s})^{1-\beta} \equiv f(s).$$

Note that $f(s)$ is strictly positive in $]\bar{s}, t[$, and that $f(s) = 0$ at both edges of the range (i.e. at $s = \bar{s}$ and at $s = t$). This means that $f(s)$ has at least one local maximum in $]\bar{s}, t[$. We need this maximum to be larger than $\beta K/\alpha$ for the FOC to hold at some point.

We now proceed to check whether this local maximum of $f(s)$ is unique:

$$f'(s) = (t-s)^{\alpha-2} (s-\bar{s})^{-\beta} [(1-\beta)(t-s) - (\alpha-1)(s-\bar{s})]$$

Since $(t-s)^{\alpha-2} (s-\bar{s})^{-\beta}$ is strictly positive in $]\bar{s}, t[$, and $[(1-\beta)(t-s) - (\alpha-1)(s-\bar{s})]$ is linear in s, positive at $s = \bar{s}$ and negative at $s = t$, $f'(s) = 0$ at exactly one point at this range (i.e. a unique local maximum of $f(s)$ in $]\bar{s}, t[$). From the continuity of $f(s)$, we get that if the value of $f(s)$ at this local maximum is greater than $\beta K/\alpha$, then $L(t,s)$ has exactly two extrema in the range $]\bar{s}, t[$. From the positive values of $L'(t,s)$ at the edges of this range, we finally conclude that the first extremum (where $f(s)$ is rising) is a maximum point of $L(t,s)$, and the second extremum (where $f(s)$ is falling) is a minimum point of $L(t,s)$, i.e. $L(t,s)$ has a unique local minimum. Conversely, if the value of $f(s)$ at its local maximum point is smaller than $\beta K/\alpha$, there is no local extremum to $L(t,s)$ in the range $]\bar{s}, t[$ and therefore $s(t) = \bar{s}$.

Once we know that there is one candidate inner solution at most, we know from the general analysis (section 14.2.1) that there exists a

$\hat{t}$ such that every type with $\bar{s} \leq t < \hat{t}$ chooses $\bar{s}$ and every type with $t > \hat{t}$ chooses his (unique) inner solution. If $\hat{t} > t_h$, this results in a distribution of stances that is Dirac Delta at $\bar{s}$. However if $\hat{t} < t_h$, then due to symmetry, the resultant distribution contains a peak at $\bar{s}$, with identical tails at each side of it (and a gap between each tail and the peak at $\bar{s}$).

We can further analyze the shape of the tails. Applying the functional form into Lemma 2, we get that the right-hand tail of the distribution has an increasing $pPDF$. In addition, since $s^*(t)$ is continuous and strictly monotonic ($\frac{ds^*}{dt} = \frac{D''(t-s^*)}{P''(s^*-\bar{s})+D''(t-s^*)} > 1$), the $pPDF$ gives us the PDF of the whole right-hand tail of the distribution of stance (the PDF of the left tail is strictly decreasing because it is the mirror image of the right tail). So the whole distribution of stances is (for $\tilde{t} < t_h$) a trimodal distribution with tails rising toward the edges of the distribution (with a gap between the peak and the tails on each side).

To conclude the proof, we notice that applying this specific power function case to part 3 of Lemma 1 implies that relative concession is decreasing at the range $\left[\hat{t}, t_h\right]$. Since types with $\bar{s} \leq t < \tilde{t}$ choose $\bar{s}$, i.e. fully conform, we get that relative concession is decreasing for every t s.t. $\bar{s} \leq t$, and therefore that relative concession is decreasing in $|t - \bar{s}|$ $\forall t$ (since $\bar{s}$ is a reflection point). As shown in section 14.2.1, conformity and absolute concession are decreasing in inner solutions while in corner solutions, conformity is constant while absolute concession is increasing. Thus, conformity is weakly decreasing and absolute concession is nonmonotonic.■

14.3.2 Case: $\alpha < 1 < \beta$

Proof of proposition 4

In this case we have a convex $P(s)$ with a unique min point $\bar{s}$, and a convex $D(t-s)$. We also have $\lim_{x \to 0+} D'(x) = \infty$. So as we saw in the general analysis (section 14.2.2), types near $\bar{s}$ choose $s = t$. We can also check and see that $\forall t > \bar{s}$ there is at most one candidate inner solution (i.e. one local minimum), by following the same stages as in the proof of proposition 3 (while noticing that this time $[(1-\beta)(t-s) - (\alpha-1)(s-\bar{s})]$ is negative at $s = \bar{s}$ and positive at $s = t$), we once more get that $L(t,s)$ has at most one local minimum (this time, it will be the first of the two extremum points).

Once we know that there is one candidate inner solution at most, we know from the general analysis (section 14.2.2) that there exists a $\hat{t}$ such that every type with $\bar{s} \leq t < \hat{t}$ chooses t and every type with $t > \hat{t}$ chooses his (unique) inner solution. If $\hat{t} > t_h$, this results in a uniform distribution of stances. However, if $\hat{t} < t_h$, then due to symmetry (and since $s^*(t)$ is continuous), the resultant distribution is continuous and bimodal with a uniform section around $\bar{s}$ and one peak on each side of $\bar{s}$ (on top of the uniform section).

We can further analyze the shape of the tails. Applying the functional form into Lemma 2, we get that the right-hand tail of the distribution has a decreasing $pPDF$. This means that the peak on the right-hand side of $\bar{s}$ has a falling slope to its right-hand side and, by symmetry, the peak on the left-hand side of $\bar{s}$ has a falling slope to its left-hand side (see figure 6).

To conclude the proof, we notice that applying this specific power function case to part 3 of Lemma 1 implies that relative concession is increasing at the range $\left[\hat{t}, t_h\right]$. Since types with $\bar{s} \leq t < \hat{t}$ choose t, i.e. do not conform at all, we get that the relative concession is increasing for every t s.t. $\bar{s} \leq t$, and therefore that the relative concession is increasing in $|t - \bar{s}|$ $\forall t$ (because $\bar{s}$ is a reflection point). As shown in section 14.2.2, conformity and absolute concession are increasing in inner solutions while in corner solutions, conformity is decreasing while absolute concession is constant. Putting the inner and the corner solutions together, conformity is non-monotonic while absolute concession is increasing.■

14.3.3 Equilibrium $\bar{s}$

Proof of proposition 6

We know that the distribution is symmetric in a neighborhood of $\bar{s}$. For $\bar{s}$ to be the average of all stances, the distribution of stances must be symmetric in the whole range. By symmetry of P around $\bar{s}$, it then follows that $\bar{s} = \frac{t_h+t_l}{2}$ is a feasible equilibrium in all cases.

1.) Assume that $\bar{s} \neq \frac{t_h+t_l}{2}$. For $\beta > \alpha \geq 1$, all types have unique inner solutions which implies $\int_{t_l}^{\bar{s}} \left(\bar{s} - s\left(\tau\right)\right) d\tau \neq \int_{\bar{s}}^{t_h} \left(s\left(\tau\right) - \bar{s}\right) d\tau$ which clearly violates symmetry. By the same reasoning, it must hold also for $\alpha < 1$ when some types have inner solutions. When $\alpha < 1$ and all types have corner solutions then $s^*\left(t\right) = t \; \forall t$ which is symmetric iff $\bar{s} = \frac{t_h+t_l}{2}$.

2.) Assume that $\bar{s} > t_l + K^{\frac{1}{\alpha-\beta}}$. Proposition 2 then implies that $\int_{t_l}^{\bar{s}} s\left(\tau\right) d\tau = \int_{t_l}^{\bar{s}} dt = \left(\bar{s} - K^{\frac{1}{\alpha-\beta}}\right) - t_l$ and $\int_{\bar{s}}^{t_h} s\left(\tau\right) d\tau = \int_{\bar{s}}^{t_h} tdt = t_l - \left(\bar{s} + K^{\frac{1}{\alpha-\beta}}\right)$ which are then not equal which violates symmetry. This proves necessity. For sufficiency, note that all $\bar{s}$ in the range imply that $s^*\left(t\right) = \bar{s}\forall t$ which is clearly symmetric.

3.) Assume that the condition is not fulfilled for some $\bar{s} > \frac{t_h+t_l}{2}$, then there is a strictly larger mass of inner solutions in the range $\left[t_l, \bar{s}\right]$ than in the range $\left[\bar{s}, t_h\right]$, which violates symmetry. A corresponding argument applies to $\bar{s} < \frac{t_h+t_l}{2}$. This concludes necessity. For sufficiency, when the requirement is fulfilled then $s^*(t) = \bar{s}\forall t$ which is clearly symmetric.

4) Proposition 2 implies that an $\bar{s}$ outside of the range has uniform tails around $\bar{s}$ which are not symmetric. Since types such that $|t - \bar{s}| > K^{\frac{1}{\alpha-\beta}}$, $s^*(t) = \bar{s}$ this also implies sufficiency.

5.) $K > 1$ implies that $s^*(t) = \bar{s}\forall t$ which is symmetric while $K \leq 1$ implies that $s^*(t) = t\forall t$ which is not symmetric.■

14.4 Aggregating individual pressure

14.4.1 Exponential distribution of pressure sources

Posit a distribution of pressure sources $f(x)$ which symmetrically has an exponential shape peaking towards the social norm from each side. W.l.o.g. let the social norm be at $s = 0$, i.e. $E(x) = 0$. The minimum and maximum pressure source in society are at $\pm\infty$.

$$\begin{aligned} P_{\exp}(s) &= \int_{t_l}^{t_h} p(|x-s|) f(x)\, dx \\ &= \frac{\lambda}{2}\int_{-\infty}^{\infty} |x-s|^{\alpha} e^{-\lambda|x|} dx = \frac{\lambda}{2}\int_{-\infty}^{\infty} |x|^{\alpha} e^{-\lambda|x+s|} dx \end{aligned}$$

where $0 < \alpha < 1$. Differentiating we get

$$\begin{aligned} P'_{\exp}(s) &= -\frac{\lambda^2}{2}\int_{-\infty}^{\infty} |x|^{\alpha} e^{-\lambda|x+s|} sgn(x+s)\, dx \\ P''_{\exp}(s) &= \frac{\lambda^3}{2}\int_{-\infty}^{\infty} |x|^{\alpha} e^{-\lambda|x+s|} dx. \end{aligned}$$

To see the behavior of this function around the social norm, we now look at

$$\begin{aligned} P''_{\exp}(0) &= \frac{\lambda^3}{2}\int_{-\infty}^{\infty} |x|^{\alpha} e^{-\lambda|x|} dx = \lambda^3\int_{0}^{\infty} x^{\alpha} e^{-\lambda x} dx \\ &= \lambda^{2-\alpha}\Gamma(\alpha+1) > 0 \end{aligned}$$

where $\Gamma(\alpha+1)$ is an incomplete Gamma function. This implies that total pressure is convex near the norm. Let us now investigate the asymptotic behavior of $P_{\exp}(s)$ for $s \to \infty$. To this end, let us use the "dimensionless" integration variable $z = x/s$, so that $P_{\exp}(s) =$

$\frac{\lambda}{2}s^{\alpha+1}\int\limits_{-\infty}^{\infty}|z-1|^{\alpha}e^{-K|t|}dt$, where $K=\lambda s$. The integral can be written as

$$\begin{aligned}P_{\exp}(s)&=\frac{\lambda}{2}s^{\alpha+1}\left[\int\limits_{-\infty}^{0}(1-z)^{\alpha}e^{Kt}dt+\int\limits_{0}^{1}(1-z)^{\alpha}e^{-Kt}dt+\int\limits_{0}^{1}(z-1)^{\alpha}e^{-Kt}dt\right]\\&=\frac{\lambda}{2}s^{\alpha+1}\left[K^{-\alpha-1}e^{K}\Gamma(\alpha+1,K)+K^{-\alpha-1}e^{-K}\gamma(\alpha+1,-K)+K^{-\alpha-1}e^{-K}\Gamma(\alpha+1)\right]\end{aligned}$$

where $\Gamma(a,b)$ are incomplete Gamma functions. For large K, $\Gamma(\alpha+1,K)\approx K^{\alpha}e^{-K}$, so the second and third terms of the sum, which contain the rapidly decreasing exponent e^{-K}, can be neglected and we finally obtain for $s\to\infty$:

$$P(s)\approx\frac{\lambda}{2}s^{\alpha+1}K^{-1}=\frac{s^{\alpha}}{2},$$

whence $P''\approx\frac{1}{2}\alpha(\alpha-1)s^{\alpha-2}\to-0$.

14.4.2 Gaussian distribution of pressure sources

We now analyze the case where the pressure sources follow a Gaussian distribution so that $f(x)=\sqrt{\frac{\lambda}{\pi}}e^{-\lambda x^{2}}$, $t_l=-\infty$, $t_l=-\infty$. The pressure is a concave power function

$$\begin{aligned}P_{gauss}(s)&=\int\limits_{t_l}^{t_h}p(|x-s|)f(x)dx\\&=\sqrt{\frac{\lambda}{\pi}}\int\limits_{-\infty}^{\infty}|x-s|^{\alpha}e^{-\lambda x^{2}}dx=\sqrt{\frac{\lambda}{\pi}}\int\limits_{-\infty}^{\infty}|x|^{\alpha}e^{-\lambda(x+s)^{2}}dx\end{aligned}$$

$$\begin{aligned}P'_{gauss}(s)&=-2\lambda\sqrt{\frac{\lambda}{\pi}}\int\limits_{-\infty}^{\infty}|x|^{\alpha}(x+s)e^{-\lambda(x+s)^{2}}dx\\P''_{gauss}(s)&=-2\lambda\sqrt{\frac{\lambda}{\pi}}\int\limits_{-\infty}^{\infty}|x|^{\alpha}\left[1-2\lambda(x+s)^{2}\right]e^{-\lambda(x+s)^{2}}dx\\P''_{gauss}(0)&=-4\lambda\sqrt{\frac{\lambda}{\pi}}\int\limits_{0}^{\infty}x^{\alpha}\left[1-2\lambda x^{2}\right]e^{-\lambda x^{2}}dx\end{aligned}$$

Substituting the integration variable with $u = x^2$, we have

$$\begin{aligned}P''_{gauss}(0) &= -2\lambda\sqrt{\frac{\lambda}{\pi}}\int\limits_0^\infty u^{(\alpha-1)/2}\left[1-2\lambda u\right]e^{-\lambda u}du \\ &= -\frac{2}{\sqrt{\pi}}\lambda^{3/2}\lambda^{-(\alpha+1)/2}\left[\Gamma\left(\frac{\alpha+1}{2}\right)-2\Gamma\left(\frac{\alpha+3}{2}\right)\right] \\ &= -\frac{2}{\sqrt{\pi}}\lambda^{(2-\alpha)/2}\Gamma\left(\frac{\alpha+1}{2}\right)\left[1-2\frac{\alpha+1}{2}\right] = \frac{2\alpha}{\sqrt{\pi}}\lambda^{(2-\alpha)/2}\Gamma\left(\frac{\alpha+1}{2}\right) > 0.\end{aligned}$$

Here, we have used the property of the Gamma function, $\Gamma(z+1) = z\Gamma(z)$. Let us now investigate the asymptotic behavior of $P(s)$ for $s \to \infty$. To this end, let us use the "dimensionless" integration variable $z = x/s$, so that $P(s) = \sqrt{\frac{\lambda}{\pi}}s^{\alpha+1}\int\limits_{-\infty}^{\infty}|z-1|^\alpha e^{-\lambda z^2}dz$, where $K = \lambda s^2$. The integral above has a saddle point at $t = 0$, so for $K \to \infty$, $P(s) \approx \sqrt{\frac{\lambda}{\pi}}s^{\alpha+1}\int\limits_{-\infty}^{\infty}\left[1-\alpha z+O\left(z^2\right)\right]e^{-Kz^2}dz = \sqrt{\frac{\lambda}{\pi}}s^{\alpha+1}\sqrt{\frac{\pi}{K}}\left[1+O\left(\frac{1}{K}\right)\right] \approx s^\alpha$. From here, $P'' \approx \alpha(\alpha-1)s^{\alpha-2} \to -0$.